PhD THESIS DECLARATION
(use a pc to fill it out)

I, the undersigned

FAMILY NAME	HUFFAKER
NAME	CARRIE
Student ID no.	1759802

Thesis title:

The Impact of Broadband Internet on Immunizations

PhD in	Economics and Finance
Cycle	2013-14
Student's Advisor	Ubfal Diego
Calendar year of thesis defence	2019

DECLARE

under my responsibility:

1) that, according to Italian Republic Presidential Decree no. 445, 28th December 2000, mendacious declarations, falsifying records and the use of false records are punishable under the Italian penal code and related special laws. Should any of the above prove true, all benefits included in this declaration and those of the temporary "embargo" are automatically forfeited from the beginning;

2) that the University has the obligation, according to art. 6, par. 11, Ministerial Decree no. 224, 30th April 1999, to keep a copy of the thesis on deposit at the "Biblioteche Nazionali Centrali" (Italian National Libraries) in Rome and Florence, where consultation will be permitted, unless there is a temporary "embargo" protecting the rights of external bodies and the industrial/commercial exploitation of the thesis;

3) that the Bocconi Library will file the thesis in its "Archivio Istituzionale ad Accesso Aperto" (Institutional Registry) which permits online consultation of the complete text (except in cases of temporary "embargo");

4) that, in order to file the thesis at the Bocconi Library, the University requires that the thesis be submitted online by the student in unalterable format to Società NORMADEC (acting on behalf of the University), and that NORMADEC will indicate in each footnote the following information:

- PhD thesis "The Impact of Broadband Internet on Immunizations";
- by Huffaker Carrie ;
- defended at Università Commerciale "Luigi Bocconi" – Milano in the year 2019;
- the thesis is protected by the regulations governing copyright (Italian law no. 633, 22nd April 1941 and subsequent modifications). The exception is the right of Università Commerciale "Luigi Bocconi" to reproduce the same, quoting the source, for research and teaching purposes;

- **only when a separate "Embargo" Request has been undersigned**: the thesis is subject to "embargo" for (indicate duration of the "embargo") ..... months;

5) that the copy of the thesis submitted online to Normadec is identical to the copies handed in/sent to the members of the Thesis Board and to any other paper or digital copy deposited at the University offices, and, as a consequence, the University is absolved from any responsibility regarding errors, inaccuracy or omissions in the contents of the thesis;

6) that the contents and organization of the thesis is an original work carried out by the undersigned and does not in any way compromise the rights of third parties (Italian law, no. 633, 22nd April 1941 and subsequent integrations and modifications), including those regarding security of personal details; therefore the University is in any case absolved from any responsibility whatsoever, civil, administrative or penal, and shall be exempt from any requests or claims from third parties;

7) **choose 7a or 7b**: 7a

7a) that the thesis is not subject to "embargo", i.e. that it is not the result of work included in the regulations governing industrial property; it was not written as part of a project financed by public or private bodies with restrictions on the diffusion of the results; is not subject to patent or protection registrations.

or

7b) that the PhD thesis is subject to "embargo" as per the separate undersigned "PhD Thesis Temporary ″Embargo″ Request".

Date 16/04/19

Abstract

The first two chapters of this thesis are an investigation of the effect of broadband internet on local-level immunization rates in the U.S. from 2000 to 2008. In chapter 1, I analyze the compliance of kindergarteners in California's public school system with official state guidelines. I find that public schools located in communities with access to broadband undergo a decline in the percentage of fully-immunized students, while Personal Belief Exemptions rise. MMR coverage falls significantly in public schools due to this exposure.

Chapter 2 extends the analysis by examining immunization dynamics among children aged 19-35 months in 257 U.S. counties over the same time period. Although the diffusion of broadband access is exogenous at the most disaggregated level- as I demonstrate in chapter 1- this crucial assumption fails at the county level. I therefore adopt an IV strategy, instrumenting number of broadband providers per county with the Technet Index. This measure summarizes state laws governing access of Internet Service Providers to public rights-of way, and thence their ability to penetrate local markets. The results of the county-level analysis are far weaker than those obtained in the first chapter. I find that broadband access leads to modest declines in the percentage of toddlers immunized against Haemophilus influenzae (Hib) and in the full coverage rate. A heterogeneous effects analysis reveals that varicella uptake fell in those counties with higher family poverty and a greater share of foreign-born residents.

Chapter 3 is independent of the first two. I employ a spatial autocorrelation model to analyze spillovers in local-level conflict on the African continent from 1997 to 2015. Subnational districts are the basic unity of analysis. I de ne spatial proximity in two different ways: By geographic contiguity and by the contiguity of ethnic groups that are divided by national borders. I find that when violence breaks out in a given district, conflict increases in geographical neighbors. This increase is even larger in magnitude when neighborhoods are defined by ethnicity. Finally, the distance of districts from central power plays a mediating role in the spread of conflict.

Chapter 1:
Information Shocks and Immunization Behavior in California Schools

Carrie Huffaker
Bocconi University

Advisor: Dr. Diego Ubfal

April 16, 2019

1 Introduction

In late December 2014, a number of recent visitors to the Disneyland Resort in Anaheim, California fell ill with measles. Over the next four months, 159 people from 18 states and several different countries contracted the disease, including 110 cases in California itself. The vast majority (82%) of victims were either unvaccinated or of unknown vaccination status (Clemmons, 2015). The outbreak attracted urgent attention to a problem that had been developing in California and the rest of the country for years: Falling vaccination rates. Over the past two decades, the U.S. has undergone outbreaks of vaccine-preventable diseases (VPDs) of a severity not seen since the 1960s and 70s, when the federal government instituted the first systematic nationwide immunization program. Studies have linked the dramatic resurgence in measles, pertussis and other preventable diseases to parental refusal to comply with vaccine guidelines and, in particular, to the widespread phenomenon of personal exemptions (Feiken et al., 2000; Glanz et al., 2009; Majunder et al., 2015; Phadke et al., 2016; Clemmons et al., 2017).

One explanation for the decline in vaccination rates and rising exemptions is the growing reliance of parents on broadband internet as a source of health information. Expert medical information is easier to access than at any time in human history. The Centers for Disease Control and Prevention, state health departments, and physician-directed organizations such as Mayo Clinic and WebMD have maintained an online presence since the early 1990s. Yet the internet is also characterized by a plethora of unverified misinformation. In opposition to public health efforts, an online anti-vaccine community, popularly known as the "anti-vaxx" movement, has arisen. The movement has grown enormously influential over time, shaping the beliefs about vaccination of countless families (Dempsey et al., 2015).

The MMR vaccine, in particular, has been under attack by the anti-vaxx movement for twenty years, ever since *The Lancet*- Britain's leading medical journal- published a study in 1998 that linked the vaccine to the development of autism disorder. Although the author, Andrew Wakefield, falsified data to support his claims and had undisclosed conflicts of interest (Meikle and Boseley, 2010), the vaccine-autism link has since attained the status of fact in the dynamic, ever-growing population of blogs and websites dedicated to attacking government-set immu-

nization requirements. Wakefield and other anti-vaxx activists argue that the online movement provides an alternative to the "failings of mainstream media," informing popular opinion in a context in which "no one knows quite what to believe" (Boseley, 2018).

This paper tests whether the diffusion of broadband internet has an impact on immunization rates, and which source of information- medical expertise or anti-vaxx misinformation- prevails in the decision-making process of parents. I match zipcode-level data on broadband provision from the Federal Communications Commission (FCC) to public school immunization and exemption records provided by the California Department of Public Health for the period of 2000-2008. I build upon FCC Form 477 data on number of Internet Service Providers (ISPs) present in each California zipcode to construct a new measure that identifies whether a public school and its student population are located in an area served by a competitive broadband market. I estimate a panel data model with school-level fixed effects on the resulting dataset of 16,071 school-year observations. Because the decision of ISPs to enter into new zipcode markets is orthogonal to past immunization behavior of local families, we can identify the impact of competitive broadband from the difference between non-exposed and exposed school populations.

I find that the spread of competitive broadband caused the share of fully-immunized children to drop by 0.83 percentage points, equal to 70% of the decline that occurred over this time period. In particular, immunizations against DTP, MMR, Hepatitis B and varicella fell due to exposure. One-quarter of the increase in demand in Personal Belief Exemptions (PBEs) is attributable to broadband. The effect is heterogeneous over time: While initial exposure drives immunizations up and exemptions down, the prevailing long-term effect is to depress immunization behavior.

I examine Google Trends data to investigate how internet use shifts the beliefs of parents, and find suggestive evidence that interest in individual vaccine-preventable diseases increases demand for immunization, but interest in the vaccines that protect against them are correlated with a higher perception of risk and reduced vaccine intentions.

California requires children to be immunized against five diseases in order to gain admittance to childcare and primary school. The state's public education system represents an invaluable case study because prior to the Disneyland outbreak, exemption laws were among the most

liberal in the country. In addition to Personal Medical Exemptions (allowed everywhere), California permitted students to forgo shots for personal, philosophical and religious reasons, all falling under the rubric of Personal Belief Exemptions (PBEs) and not requiring any input from a licensed physician. The state legislature responded to the crisis in June 2015 by passing California Senate Bill 277, which eliminated PBEs.

This paper contributes to a large literature examining the impact of new media upon social and health outcomes. A good summary of the literature addressing the link between television and development goals is provided by La Ferrara (2015). La Ferrara et al. (2008) find that the popularity of soap operas on the new Globo television network lowered fertility rates in Brazil, by spreading new social norms about the desirability of small, more "modern" families. Kearney and Levin (2014) discover a similar dampening effect on teen pregnancy rates in the U.S. as a result of MTV's *16 and Pregnant* program. The diffusion of cable TV in India has altered social preferences about domestic violence and fertility, and led to an increase in female autonomy (Jensen and Oster, 2008). Other recent contributions have analyzed the information effect of sexual education programs on adolescent fertility and risk-taking, both traditionally-run (Dupas et al., 2018) and in an online format (Chong et al., 2013). Also related is the body of literature examining the impact of broadband internet upon political outcomes. Lelkes (2015) show that the spread of broadband access has increased partisan hostilities in the U.S.. Research focusing on voter turnout in Germany and the U.K. has found a negative impact of broadband (Falck et al., 2014; Gavazza et al., 2016). Larcinese and Miner (2017) show the opposite to hold true in the U.S, while Heblich (2016) also finds a positive effect upon German turnout in more recent years.

My contribution is to investigate how the diffusion of broadband has influenced a new type of outcome, immunization. There has been little empirical work done to address the question of how the ability to access health information online has transformed health outcomes in the real world.[1] To my knowledge, this paper is the first to empirically test whether broadband access changes immunization outcomes.

The organization of the paper is as follows. In the next section, I discuss how broadband

[1]Billari et al. (2019) find that expansion of broadband internet in Germany has increased fertility among highly-educated women, although in their case the mechanism is greater flexibility in work hours rather than an information shock.

has revolutionized information-seeking about personal health. I address the role of the online anti-vaccine movement in shifting perceptions about immunization. The dataset and empirical strategy are introduced in section 3. Public school estimation results are presented in section 4. I test for selection bias in the next section. Analysis of heterogeneous effects and robustness checks follow in sections 6 and 7. The paper closes with a discussion of mechanisms and concluding remarks.

2 The Internet and Immunizations

By the mid-1990s, the advent of the WorldWideWeb, Web browsers and efficient search engines had enabled the American public to access the internet at far higher rates than ever before. Privatization of the internet backbone led to the growth of many new Internet Service Providers (ISPs) to satisfy the soaring demand. This expansion in usership was eventually followed by an expansion in who was able to create and publish new content on the internet. Blog-publishing tools such as Blogger had appeared by the late 1990s, enabling the non-technical public to build their own websites for the first time. New forms of social media followed, including Facebook (2004), Youtube (2005) and Twitter (2006). The explosion of user-generated content, together with the near-universal online presence of real-world organizations and institutions, has produced a democratization of information without precedent.

Studies carried out by the Pew Research Center reveal the increasing dependence of Americans upon the internet as a source of information. By 2000, 55% of Americans with internet access had looked up health information online; within two years, this figure had risen to 80% (Fox and Rainie, 2000; Fox, 2005). By 2006, a large proportion of the population had also come to rely upon the internet as their primary source of science news and information (Horrigan, 2006). The Horrigan study is interesting for two reasons. First, it reveals that individuals with home broadband connections use the internet differently. They are far more likely than dial-up users to obtain most of their science information from the internet (34% vs. 22%). This result also holds in the context of health: 88% of broadband users researched health information online in 2009, versus 72% of dial-up users (Fox and Jones, 2009). Broadband access is associated with a substitution away from other sources of information like television,

magazines and newspapers.

Second, young adults age 18-29 (followed closely by adults age 30-49), are more likely than any other demographic to look up health and science topics on the internet, and to use it as their primary source of information. This reliance has increased over time. The same result holds true for parents of children under age 18. The implications are clear: Many young parents, as well as adults about to enter into their childbearing years, rely upon the internet for health and science information. In many cases- about four in ten- web resources are their principal learning reference. It is possible that these habits extend to exploiting online resources when deciding whether or not to vaccinate a child.

The main explanation for this dependence on the internet, given by 7 of 10 online seekers, is convenience (Horrigan, 2006). The internet has drastically lowered search costs, creating a world in which information is cheap and plentiful. Yet online research is characterized by one major drawback: a wealth of unsubstantiated and false "facts" compete with good sources in the online marketplace. Three-quarters of online seekers do not consistently check the date and source of online health information (Fox and Jones,2009). One in four feel overwhelmed by the sheer amount of available health information, while one in five feel confused by what they read online. Similarly, a Pew report on internet use during the 2008 electoral campaign found that it is "usually difficult" for 56% of online adults to distinguish between what is true and what is false on the internet (Smith, 2011).

The online "anti-vaxx" movement has exploited this ambiguity between true and false information in creating an audience. Content analyses of prominent anti-vaccine websites- most of which present themselves as objective explorations of truth- reveal a common narrative placing doctors and other authorities in opposition to concerned parents. The movement encourages parents to assume the role of self-taught experts on vaccinations and their child's health (Kata, 2010). It accuses the medical establishment of hiding information about the negative consequences of vaccines in order to maximize profits for themselves and for pharmaceutical companies (ECDC, 2012; Betsch et al., 2012). Anti-vaccine websites frequently reject statistics and scientific evidence- indeed, the scientific process as a whole- on the grounds that the true risks of vaccination are subjective, unknown, and ultimately *unknowable* (Hobson-West, 2007; Brownlie and Howson, 2005).

Experimental evidence suggests that anti-vaccine websites can influence real-world vaccine decisions through changes in the preference for vaccination. Betsch et al. (2010) randomly exposed subjects to either a vaccine-critical or vaccine-neutral website (run by the German Ministry of Health) for 30 minutes, in order to evaluate how risk perceptions and intention to immunize changed. They found that accessing the vaccine-critical website significantly increased perceived riskiness of immunizations, while decreasing the perceived risk of not immunizing. Intention to immunize was negatively impacted. In a second experiment (2011), the authors built a simulated online bulletin board with statistical and narrative information about vaccines. They again found that anti-vaccine narratives distorted risk perceptions of subjects, decreasing their trust in scientific evidence. The more emotional the anecdote, the greater its power to sway beliefs and intentions. Overall, there is a significant connection between exposure to anti-vaccine websites and reduced vaccine intentions (Kata, 2010).

Beliefs are fundamental to real-world vaccination behavior. A significant share of parents who choose to delay or decline one or more required vaccines believe that vaccines cause both short-term side effects and potentially longer-term, more severe side effects. For these parents, following the CDC-recommended schedule is *more dangerous* than the alternative of no immunizations at all, or delayed ones. Such beliefs are widespread even among parents who adhere to the schedule (Dempsey et al., 2015). Vaccines are perceived, in many cases, as riskier than the diseases they prevent (Saada et al., 2015).

3 Data and Empirical Strategy

3.1 Internet Data

Data on broadband access are drawn from Federal Communication Commission (FCC) Form 477 Additional Data webpage.[2] All suppliers of high-speed internet are required to report to the FCC on a biannual basis, in June and December. The ISPs report number of HS lines provided; their geographic location; and the type of technology used. The data is available beginning from December 1999.

[2] Available at: https://www.fcc.gov/general/fcc-form-477-additional-data.

For the December 1999 to June 2008 period, only supply-side data are available to the public at the local level. The FCC reports the number of providers present in each U.S. zipcode. Those zipcodes with 1-3 providers are grouped into a single class, and marked with an asterisk in the data. At the state level, the FCC reports total number of providers present on the territory, as well as number of providers and number of HS lines disaggregated by type of technology. Finally, within the larger class of total HS lines, the number of residential and small business subscribers is identified.

The identity of the specific ISPs serving each state is considered proprietary information and not available to the public. Nor is it possible to distinguish between zipcodes with one, two or three providers. Several Freedom of Information Act requests of this nature to the FCC were rejected.

In response to these two challenges- the lack of local-level demand data, and ambiguity about the actual number of providers in smaller markets- I employ three separate measures of broadband provision. The first is number of providers in a zipcode market, which has been exploited by other authors such as Larcinese and Miner. I assign the median value of "2" to the 1-3 class; results (available upon request) are quite similar if one instead uses "1" or "3." Next, I construct two novel measures. *Broadband* is equal to one if at least one provider is present in a zipcode, and zero if there are no providers. *Competitive broadband* equals one if four or more providers are present, and zero if less than four are present. *Broadband* therefore measures whether a market exists at all in a given zipcode, while *competitive broadband* distinguishes between a large or small (if any) number of providers, and constitutes a sort of proxy for the extent to which a market is competitive or oligarchic. I employ the June data for each year to construct these measures.

Figure 1 illustrates the expansion of broadband internet across California from 1999 to 2008. In June 2000, 22 ISPs provided broadband to the state of California. Eight years later, the market had nearly quadrupled in size, with 84 ISPs participating. The demand side expanded even more rapidly. Total HS internet subscriptions grew by 3,566%, from about 500,000 in December 1999 to 18.6 million eight years later. The residential market grew by 892%, from about one million subscriptions in December 2000 to 10.4 million in 2008. During this period, residential and small business subscribers varied between 56% (in 2008) to 94% (in 2004) of

the total market. This is due to differential trends in the growth of the large corporate and residential sectors. While the corporate sector experienced explosive growth from 2005, the residential sector grew most quickly in the first half of the decade.

Local-level markets are also characterized by steady growth in broadband provision (Figure 2). Each biannual observation represents the zipcode average for number of providers, and the proportion of zipcodes enjoying *broadband* and *competitive broadband*. Mean number of providers per zipcode grew by 289%, from 1.7 in December 1999 to 6.6 nine years later. As the access measure reveals, in 1999 California was already quite developed relative to other national and global markets. Nearly half of all zip codes had at least one provider by the beginning of the period, while 2/3 had access by the end. This is unsurprising, given the prominence of Silicon Valley and other high-tech enclaves in the development of the state economy.

Competitive broadband is the preferred measure of internet provision that I utilize throughout the paper. The measure starts at a much lower level than *broadband*, at 21.4 percent of all zipcodes in 1999, but reaches near parity by 2008. Throughout the decade, the proportion of zipcodes with *competitive broadband* closely tracks average number of providers. The growth of *competitive broadband* is not homogeneous across state territory, however. In 2000, zipcodes with a large number of providers tended to be clustered in the Bay Area, Los Angeles and San Diego metropolitan areas. *Competitive broadband* radiates slowly outward from these regions until, by 2008, the majority of zipcodes across state territory have access to 4 or more providers. Figure 3 provides year-by-year snapshots of this evolution.

Zipcodes with one or more provider are already scattered across the state in 2000, although richly-populated regions along the coast and in the center of the state are over-represented. Access expands rapidly across California, until- by June 2008- the statewide market is more or less saturated, with few areas lacking access (Appendix Figure 1). Far greater geographic heterogeneity characterizes the provider measure (Appendix Figure 2). Mirroring the dynamics of *high access*, most zipcodes, save for those in the Bay Area and southwest corner of the state, have a low number of providers in 2000. Provider numbers slowly increase state-wide, but the very highest number of providers are always concentrated in the zipcodes of high-tech markets.

Some explanation is needed for why chunks of California appear to be missing from the three sets of zipcode maps. In fact, what is being mapped are not zipcodes but Zipcode Tabulation Areas (ZCTAs). The U.S. Census Bureau introduced ZCTAs in 2000 for the purpose of tabulating area summary statistics. While zipcodes include (uninhabited) P.O. Boxes and large-volume customers, ZCTAs correspond to the populated areas of California. Due to the presence of P.O. Boxes, multiple zipcodes may by covered by the same ZCTA. Therefore, in order to verify whether the residents served by a local school inhabit an area with access to broadband, it was first necessary to map all school addresses to ZCTAs using a zipcode-to-ZCTA crosswalk dataset provided by the Census Bureau. Zipcode data provided by the FCC were likewise mapped to ZCTAs in order to ensure that no zipcode was recorded as lacking internet simply because it corresponded to a P.O. Box. Thus each school is uniquely identified by an inhabited ZCTA with internet data, although for simplicity I will continue to refer to these throughout the paper as zipcodes. As the topographical map of Appendix Figure 1 makes clear, the missing chunks in the maps correspond to mountain ranges, deserts and other uninhabited areas of the state, not covered by any ZCTA.

3.1.1 Significance of the Constructed Internet Measures

The utility of the three provider measures presented in this section rests upon two assumptions: First, that supply-side measures can capture the unobserved demand side of the market; and second, that the constructed measures *broadband* and *competitive broadband* provide useful information about the actual ability of and extent to which local residents consume broadband internet. I first show that a strong, positive relationship subsists between the supply and demand sides of the market, before turning to the second and more complicated question of what access means for consumer demand. This requires a deeper examination of the dynamics of U.S. broadband markets at the local level.

The supply-demand comparison is based on FCC data for all fifty U.S. states, plus the District of Columbia, from December 1999 to June 2008. As before, I aggregate zipcode measures up to state level to obtain the mean number of providers per zipcode and the proportion of zipcodes in each state with a value of one for *broadband* and *competitive broadband*. Coefficient estimates from the univariate regression of log residential high-speed lines on the

provider measures are presented in Table 1. All estimates are positive and significant at the 1% level when state fixed effects are included (columns 1 and 3), although the two provider variables lose significance once year fixed effects are added (columns 2 and 4). The relationship with the two constructed measures is stronger and more robust: A larger proportion of zipcodes with access to broadband internet is associated with a higher level of residential subscriptions. These findings support the use of *broadband* and *competitive broadband*, rather than number of providers, as the principal proxies for broadband demand dynamics.

To better understand the *economic meaning* of these two measures, it is important to consider how the U.S. broadband market has evolved over time. In the early years of the internet age, the majority of internet subscriptions were not high-speed, but dial-up services that utilized the existing infrastructure of telephone lines. Interexchange Carriers, also known as long-distance phone companies, controlled the basic transmission mechanism for internet backbone traffic. Even more important, however, were the Local Exchange Carriers (LECs)- local telephone companies- which possessed the *last mile* of copper wire supplying telephony and internet services to nearly every home and business in the U.S. This monopoly granted the incumbent LECs a considerable amount of market power in the burgeoning market.

Newer technologies enabled data transmission at much higher bandwidths. Digital subscriber lines (DSL) provided increased bandwidth from the existing copper loops of local telephone companies, and unlike dial-up service did not interfere with the carriage of voice service. By 1998, many incumbent LECs had begun to offer DSL service. At the same time, the innovation of broadband access via coaxial cable enabled a new participant- local cable companies- to enter the market. This entrance was eased by the fact that 2/3 of American households already subscribed to cable television. As with LECs, local cable companies enjoyed market power through the ownership of the last mile of coaxial cable. This presented a challenge to outside ISPs. While many LECs and cable companies themselves entered the market as internet providers, outside ISPs depended on the last mile facilities of others for access to retail customers (FCC, 1998).

Broadband markets may represent natural monopolies. The conversion of phone lines and cable networks to permit the transmission of broadband data requires an enormous upfront investment. Should the firm then be forced to leave the market at a future date due to losses,

broadband technology cannot be repurposed for other uses- which means that a large share of capital costs will be ultimately non-recoverable (Prieger and Connolly, 2013). In this case, as well, difficulty of exit may impede firm entrance into the market, and non-competitive monopoly or oligopoly conditions will persist.

Regulators are not unaware of the challenges facing competition in the telecommunications industry in general, and broadband internet in particular. Congress passed the Telecommunications Act of 1996 with the intention of establishing greater competition in local and long-distance telecom markets. Yet these reforms have proven a failure for a number of reasons.

First, they failed to account for firm incentives. A zipcode-level study of broadband penetration from 2005 to 2008 found a positive correlation between ISP entry and exit rates (Prieger and Connolly). The authors interpret this as evidence of an underlying correlation between entry and exit barriers, given that exit barriers are entry barriers to forward-looking firms. While regulators have pushed entry into less-profitable rural markets, ISPs may be hesitant to risk losses in a market that will then be difficult to exit (again, due to regulatory pressure).

Second, the absence of penalties for delay or noncompliance on the part of incumbent LECs charged with allowing entry to competitors ensured that a decade later, local loops remained bottlenecks, controlled by monopolies. Thus non-competitive conditions have persisted in the DSL market. Consolidation has also occurred in the coaxial cable market. In the late 1990s, AT&T adopted a new strategy of purchasing local cable companies in order to convert cable television wires into broadband infrastructure. The acquisitions of AT&T ensure that in many cases, local consumers are faced with few alternatives in the coaxial cable market as well (Economides).

Because American markets feature few suppliers, the maximum speed of connection offered is low, on average (Hussain et al., 2013). Slower service may have implications for usage by consumers, as even users with access to broadband subscriptions- frustrated by slower speeds- may exploit those subscriptions to a lesser extent. This would be coherent with the finding that broadband subscribers use the internet more- and more intensively- than do their dial-up counterparts.

There is evidence that lack of competition is pricing some American consumers out of the market. The U.S. National Telecommunications and Information Administration found in 2013 that 8% of U.S. households could not afford broadband. The Pew Research Center's "Home Broadband 2015" report likewise finds that high prices were the number one reason why households lacked a subscription; 59% of non-broadband users cited "monthly cost of home broadband subscription is too expensive" as the most important reason for their choice (Horrigan and Duggan, 2015).

Broadband and *competitive broadband* both enjoy a stronger and more robust relationship with number of residential high-speed connections than does number of providers. But as the above evidence suggests, merely considering the extensive margin of internet provision- as measured by *broadband*- is inadequate. Zipcodes with broadband access may not translate into the actual ability of residents to pay for and utilize that broadband. Moreover, noncompetitive conditions may result in the supply of overly slow connections. For these reasons, *competitive broadband*- which accounts not only for the availability of broadband, but whether a local market is more or less concentrated- is a better proxy for residential usage and is the principle measure I will focus upon in this paper. Access is key, but whether a market is monopolistic, oligopolistic or competitive is even more crucial to the consumption choices of local households.

3.2 School Immunization Data

In analyzing the dynamics of immunization rates among kindergarteners in California's public and private schools from 2000 to 2008, I rely upon data provided by the Health and Human Services Agency of the California Department of Public Health (CDPH). Each fall, schools are required to check the immunization records of all new K-12 student admissions, as well as of all students progressing to seventh grade. These data are compiled into summary reports by dedicated school officials and communicated to the CDPH.

The parents of each child entering kindergarten (age 5-6) must provide the school with the child's personal immunization record- usually filled out by the family's doctor or pediatrician- before she can be admitted. School personnel then complete a California School Immunization Record for each child, evaluating her shot history in light of state guidelines in order to deter-

mine whether she is up-to-date (UTD) on all immunizations. For admission to kindergarten, these requirements are 5 doses of Diphtheria, Tetanus and Pertussis (DTP) vaccine; 4 doses of Polio; 2 doses of Measles, Mumps and Rubella (MMR); 3 doses of Hepatitis B; and finally, one dose of Varicella.

Four categories of students were enrolled during this period: Those who were UTD or conditionally exempt, and those who obtained Personal Medical or Personal Belief Exemptions. *Up-to-date* students were free to attend school with no further action. *Conditionally admitted* students had received all needed immunizations for their age (requirements are tied to specific birthdays), but required more shots in the future. School personnel reviewed the records of these students each month, and contacted parents when future shots fall due; children whose parents refused to comply were expelled.

Personal Medical Exemptions (PMEs) were intended for children with autoimmune deficiencies or other health problems, to whom immunizations pose real danger. PMEs required a written statement by a licensed physician, explaining which vaccines are exempt and whether the exemption is permanent or temporary. Finally, *Personal Belief Exemptions* (PBEs) were available to children whose parents opposed one or more immunizations on religious or philosophical grounds. Such an exemption was easier to obtain than a PME, as the child's parent or guardian needed only submit a signed affidavit.[3]

Once school staff reviewed the personal immunization records and filled out the CSIR forms, they compiled an annual immunization report for the entire school. This consisted of summary statistics on the immunization status of enrolled students. These statistics, which were then submitted to the CDPH, form the basis of my analysis.[4] My data consists of 16,071 public school-year observations, and 11,063 private school-year observations. Data are not available for all schools in all years (also because schools enter and exit the market for education with some frequency), so the panel is unbalanced. Schools that appear only once in the sample are

[3]The 2015 abolition of PBEs by the state of California applies only to new admissions. Students enjoying a PBE who are already enrolled at a school continue to be exempt under a grandfather clause. However, the exemption is no longer valid for admission to seventh grade; all junior high students (save for PMEs) must have all immunizations. For more details on how California's immunization requirements are applied, see the "California Immunization Handbook For Child Care Programs and Schools" (2016), published by the CPDH.

[4]The CDPH exercises a degree of oversight in the form of visits each spring to a random sample of schools by state and local health department staff. The staff assess compliance with immunization requirements and the effectiveness of follow-up with conditionally-admitted students.

excluded from the analysis.

3.3 Empirical Model and Estimation

The model for immunization behavior in California's public schools is given by:

$$v_{idt} = \gamma Net_{idt} + \varphi Ct + \beta X_{idt-1} + \alpha_i + \tau_t + dt + \varepsilon_{zt}$$

where school *i* is nested within district *d* and observed in year *t*. *Net* is the *competitive broadband* indicator. The charter school dummy C is interacted with a linear time trend. The vector of (lagged) school-specific control variables is given by *X*. All specifications include school fixed effects α_i and year fixed effects τ_t, while some also include the district-specific time trend *dt*. Finally, *v* is the immunization outcome of interest, measured in percentages. Standard errors are clustered at the zipcode level because it is at this level that the broadband indicators vary.

If variation in broadband provision is not exogenous, then a causal effect cannot be identified. In section 5, I run a number of validity checks. These include an analysis of the impact of future levels of competitive broadband on past immunization outcomes and a check for pre-existing trends in the relationship between immunization and internet variables.

3.3.1 School and District Control Variables

I include two classes of time-varying school-level control variables. While the first class is specific to the set of kindergarten-grade students, the second pertains to the entire school and is obtained from Academic Performance Index (API) records. All data are obtained from the California Department of Education.

The API was a system for evaluating K-12 public education in the state of California, centered around biannual standardized testing (at the beginning and end of each academic year) and in use until 2013. Students were first tested in 2nd grade. API statistics are relative to the body of students who took the test; for most elementary schools, this would be the set of 2nd, 3rd, 4th and 5th graders. The variables I take from the API are the school's beginning-of-school-year API score, which reflects historical school performance; growth in the API over the

academic year, which reflects improvements in childrens' educational experience; the School Characteristic Index, a composite measure of overall socioeconomic status of the student body; a school mobility measure that captures student turnover; average kindergarten through third grade class size; and finally, the proportion of teachers with emergency credentials, who are not fully certified to teach. I also include proportion of students eligible for free and reduced meals, which directly correlates to family income, and parents' average educational level.[5]

A second class of demographic variables is measured for kindergarten students alone. I include percentage of students who are either learning English, or have been reclassified as (non-native speaking) English proficient. I also control for kindergarten population share of Native American, Asian, Pacific Islander, Filipino, Hispanic and black students. Non-Hispanic white students are the excluded category.

The choice of these explanatory variables is informed by research on immunization and broadband preferences. A 2015 Pew Research poll found important demographic differences in support of mandatory vaccination laws. Income and education are positively correlated with support. Larger shares of Hispanic than black or white Americans support the laws, although these differences are not statistically significant. Finally, the 18-29 demographic is least likely to support mandatory vaccination, as are parents of children under 18. These levels of support are strongly correlated with the perception of whether vaccines such as MMR are generally safe (Pew Research Center, 2015). Although the findings pertain to a later period, it is probable that these variables also shaped vaccine preferences from 2000 to 2008.

A 2005 survey finds many of the same observable characteristics to be associated with broadband adoption by American households. Early broadband adopters tended to earn higher incomes; they were white, highly educated, and young. The 18-29 age group has the highest adoption rate (Horrigan, 2006).

Considering these two surveys together, it is clear that broadband uptake, vaccine beliefs and vaccine intentions are all correlated. Conditional on the observable characteristics, we can exclude the possibility that immunization rate dynamics spring from modifications in school quality or in the sociodemographic composition of the student body. Identification of the

[5]There may be some measurement error in this variable, as parental education levels are reported by the students themselves.

treatment effect then relies upon the assumption that there are no time-varying unobservable factors correlated with both internet provision and kindergarten immunization behavior.

4 Results: Broadband Internet and Immunization Outcomes

4.1 Percent Fully Immunized and Exemption Results

The results of those specifications employing the *competitive broadband* indicator are presented in the main body of the paper, for both public and private schools. Corresponding results using *broadband* and *number of providers* are presented in the appendix. Table 2 displays the coefficient estimates for the impact of high access upon percent up-to-date and the share of Personal Belief Exemptions. When the number of broadband providers in a zipcode increases from the 1-3 class to four or more, the proportion of public school kindergarteners who are up-to-date on all immunizations falls by 0.83 points, on average (Column 1). The effect is significant at the 1% level. Once a school district-specific trend is added (Column 2), the coefficient estimate halves in magnitude and loses statistical significance.

The increase in Personal Belief Exemptions resulting from the transition from a low- to high-provider regime is far greater in magnitude (Columns 3 and 4). The estimated effect of 0.22 increases to 0.311 with the inclusion of the time trend. The estimate is significant at the 1% level.

The estimation results presented in Table 3 confirm the predicted orthogonality between broadband provision and Personal Medical Exemptions; though PME rates did rise slightly throughout the decade, this change does not derive from broadband use. Conditional Exemptions, like PBEs, respond positively to high access, though the estimate loses significance when district trends are added. Based upon these results, the decline in % UTD resulting from broadband use appears to largely reflect a marked increase in PBEs.

Tables 4 and 5 present coefficient estimates of the impact of *competitive broadband* upon the five individual vaccine series required by state law, for the subset of public schools. The tran-

sition to four or more providers causes all vaccination rates to drop; the DTP, MMR, Hepatitis B and varicella estimates are statistically significant, though only MMR retains significance (at the 5% level) once district trends are included.

Estimation results utilizing the *broadband* and number of provider indicators, as well as the private school regressions, are found in the appendix.

4.2 Magnitude of the Estimation Results

Compared to the sample means of each immunization and exemption variable, the estimated effects appear quite small. What must be taken into consideration, however, is the limited magnitude of the changes in behavior that occurred over this time period. Among public schools, the full coverage rate fell from 92.51% to 91.34% from 2000 to 2008; the estimated treatment effect of -0.83 amounts to 70.82% of this decrease. Similarly, Personal Belief Exemptions rose from 0.80 to 2.07. The treatment effect is one-quarter of this increase. The positive Conditional Exemptions estimate is striking, as demand for those exemptions actually fell by .19 points. The estimated effect is 3.13 times larger in magnitude. With the exception of varicella (which rose by 1.03 points), all of the individual vaccine series declined during the period of study, with an average shift of 1.66 points. The MMR coefficient of -.25 constitutes 15.4% of the fall in uptake for that vaccine.

5 Selection

In the this section, I present the results of the various validity checks that address the question of selection and provide evidence of the exogeneity of the competitive broadband indicator. The fulfillment of this key assumption is what allows us to identify the public school coefficient estimates presented in the previous section as the true treatment effect of broadband internet upon immunization outcomes. I perform the same set of checks for private schools; results are presented in the appendix.

The first and most crucial of the validity checks is graphical evaluation of the parallel trends assumption. If this assumption holds, then the estimated broadband coefficient $\gamma = 0$ for all

time periods prior to the introduction of broadband, with $\gamma \neq 0$ only once the technological change has occurred. Broadband must not exercise a significant impact upon immunization outcomes prior to its introduction. Evaluation of this hypothesis is complicated by the fact that different zipcodes receive internet in different years. My broadband indicator refers to June of each year, while the immunization variables are measured in September of the same year. Let *t* denote the June in which *competitive broadband* or *broadband* equal one for the first time. The coefficients on $June_{t-1}$, $June_{t-2}$, $June_{t-3}$, ... should all equal zero, while $June_{t+1}$, $June_{t+2}$, $June_{t+3}$, ... may be positive or negative. The specification is otherwise identical to that introduced above.

I plot the coefficient estimates and confidence intervals for the public school regressions of % UTD and the three exemption rates on *competitive broadband* in Figures 4 and 5. The reference category is $June_t$. The plots show clear confirmation of the parallel trends assumption. Prior to time t+1, none of the *competitive broadband* coefficients are significantly different from zero. In the three years following the switch to a competitive number of providers, the impact becomes negative and significant for % UTD, positive and significant for the PBE rate. Competitive broadband has a large, positive effect on Conditional Exemptions at time t+2. All of these findings reflect the regression results discussed above. The t-2 effect on PME is negative and significant at the 10% level. However, the overall coefficient estimate for PME is not statistically significant, suggesting that the finding is not of great concern.

Next, I perform a placebo test that pairs the broadband indicator with its one-year lead. Future internet measures should have no impact upon present immunization rates; otherwise, the measure cannot be considered exogenous. I adopt the more permissive specification, excluding district trends. The results do not change when trends are included.

Coefficient estimates for the full coverage rate and three types of exemptions, for the subset of public schools and utilizing the competitive broadband indicator, are presented in Table 6. While the % UTD, PBE and conditional exemption coefficients are statistically significant- as before- their one-year leads are, reassuringly, never significant.

The two validity tests presented here provide strong evidence that the key identifying assumption of the analysis- the orthogonality of broadband provision to time-varying unobservable factors that might be correlated with immunization behavior- is satisfied. The assumption holds

for both public and private schools, and for all three broadband indicators (results available upon request). Though local broadband provision is hardly random, selection into treatment does not pose a threat to identification and we can interpret the coefficient estimates as causal effects.

6 Heterogeneous Effects

In this section, I evaluate how different population groups respond to the information shock posed by competitive broadband by subdividing the dataset and testing for heterogeneous effects. I concentrate on the set of public schools. First, I divide the sample based upon beginning-of-period immunization rates. This test allows us to evaluate how pre-existing immunization behavior shapes the response of communities, and whether free riding- falling coverage as a result of herd immunity- is occurring. Next, I differentiate between charter and non-charter public schools. Finally, I subdivide the sample of public schools based upon two characteristics of the student body: Eligibility for free and reduced meals and API test scores.

6.1 Initial Vaccination Rates

One of the most popular explanations for falling vaccination rates is free riding. Immunization behavior has clear positive externalities. While health consumers equate marginal private benefit with marginal cost in determining the optimal personal level of immunization, in doing so they ignore the marginal external benefit of their immunization decision. In the aggregate this leads to the systematic underprovision of immunizations (Folland et al., 2013). Bauch and Earn (2004) perform a game-theoretic analysis of vaccination and find that when coverage rates are high, even a slight risk associated with vaccines will outweigh the perceived risk from infection for the individual. They conclude that it is "impossible to eradicate a disease through voluntary vaccinations when individuals act according to their own interests." There are numerous documented cases of vaccine scares that caused coverage levels to fall below the herd immunity rate necessary to protect the entire population from illness, leading to the reintroduction of vaccine-preventable diseases (Poland and Jacobson, 2011). A key mediating

factor variable in the feedback loop between disease outbreaks and the strategic behavior of individuals is prevalence elasticity of demand for immunization: How changes in prevalence of a disease move immunization coverage (Folland). As discussed in section 2, the broadband information shock can shape prevalence elasticity of demand by acting upon perceptions of both vaccine risk and infection risk.

I divide the set of public schools into upper and lower halves based upon the median value in 2000 of percent up-to-date (equal to 90%).[6] Schools not observed in 2000 were dropped from the dataset. Regression results for the full coverage and three exemption rates are presented in Tables 7 and 8. There is no significant impact of *competitive broadband* upon % UTD or Conditional Exemptions. The PBE coefficients are both positive and similar in magnitude to the full dataset estimate. This is offset, in the bottom half of the dataset, by a decline in Personal Medical Exemptions..

The individual vaccine rates demonstrate greater heterogeneity, with decreases in every series except polio, but for different halves of the dataset (Tables 9 and 10). MMR alone fell for both sets of schools. Among those schools with initially low coverage rates, the Hepatitis B and varicella rates also fell, although none of the estimates survive the inclusion of district trends. A more robust result is the decrease in DTP, MMR and Hepatitis B coverage rates for those schools that began the period with higher immunization rates.

Taken as a whole, the evidence is weakly supportive of the hypothesis that schools with initial higher coverage rates- and therefore, with a greater incentive to free ride- responded to broadband by immunizing less. Schools enjoying both stronger and weaker levels of herd immunity responded to the information shock, but the response of the the latter group is more ambiguous.

6.2 Charter School Status

Traditional public schools receive government funding and are overseen by local school districts. Charter schools also operate using public money, in the form of vouchers, but are run independently by private individuals and organizations, with relatively less government

[6] Similar results (available upon request) obtain when the top and bottom terciles are instead compared.

oversight. Only 5% of public schools are charters.

Coefficients partitioned by charter school status are Tables 11-14. In this case, results are diametrically opposed. While charter school immunization rates rose as a result of the broadband treatment, those of non-charter schools fell. Non-charter schools also saw an increase in Belief and Conditional Exemptions, while the latter type of exemption declined among charter schools. The individual coverage rates of charter schools rose for every series except varicella. Non-charter coverage rates instead declined for all five series. Selection into charter schooling is therefore associated with a different processing of online information: Whereas charter parents become more inclined to adhere to official vaccination schedules as a result of exposure to broadband, non-charter parents grow less so. Families seeking an alternative to traditional public education appear to respond more strongly to online expert knowledge than to the overload of misinformation.

6.3 Eligibility for Free and Reduced Meals

The top and bottom halves of the public school dataset are divided by a median value of students eligible for free and reduced meals equal to 54.5%. This measure, which is directly determined by household income, does not distinguish between schools in a definite way. Schools with low- versus high-income student bodies react to expanded broadband access in much the same way (Tables 15-18). For both types of schools, access results in a greater number of Personal Belief Exemptions (while failing to influence % UTD). MMR coverage rates fall. However, wealthier schools also witness a decline in the Hepatitis B and varicella coverage rates, though estimates are only significant at the 10% level and not robust to the inclusion of district trends. Thus both low- and high-income schools respond to competitive broadband by reducing immunizations, but the response of high-income schools is more dramatic, with three of five individual series rates decreasing.

6.4 API Scores

Estimation results for the partitioning of schools by API test score are presented in Tables 19-24. In low-performance schools, the full coverage rate fell while Conditional Exemptions

rose. In particular, DTP, MMR and Hepatitis B immunization rates declined. Among high-performance schools, PBEs rose, apparently in response to a large decrease in demand for the MMR vaccine.

We can conclude that broadband exposure acted upon low-performance schools by depressing demand for most of the required immunizations, and instead impacted high-performance schools by increasing demand for PBEs requested by parents opposed to the MMR vaccine.

7 Robustness

In this section, I perform two types of robustness checks, concentrating on the set of public schools. First, I substitute the *competitive broadband* indicator with length of exposure to competitive broadband, measured in months. This measure exploits a second type of intensive margin variation: Rather than number of providers serving a zipcode, the length of time they have been present and the long-term effects resulting from that presence.

The relationship between months of competitive broadband and immunization is non-linear (Tables A10-A13). Initial exposure drives up demand for immunizations, but longer periods depress it. Loss of precision of the initial exposure estimate, once district trends are added, suggests that the long-run effect is the crucial one; in fact the treatment effect upon % UTD in Table 2 is negative. Personal Belief and Conditional Exemptions decline and then rise again, mirroring the positive impact of *competitive broadband* upon these variables. There is no relationship between length of exposure and Personal Medical Exemptions.

The individual vaccine series present greater variation. Early exposure drives down DTP rates, in parallel with the treatment effect in Table 4, while the long-term effect of broadband is null. Polio uptake falls, only to rise again in the long run. The two effects appear to cancel each other out, as the overall impact of broadband upon polio uptake is zero, but the positive coefficient on the quadratic term may reflect the positive treatment effect for charter schools found in Table 13. Though the early- and late-exposure MMR coefficients are both negative, they are imprecisely estimated.

Initial exposure to broadband increases Hepatitis B uptake, with no long-run effect. Varicella

uptake increases and then declines, though the long-run effect loses significance when district trends are added. The two sets of estimates run in the opposite direction of those found for the competitive broadband indicator in Table 5. However, the positive treatment effects are coherent with the real-world dynamics of demand for the Hepatitis B and varicella vaccines from 2000 to 2008. Hepatitis B coverage fell by a smaller amount than any other series, with a decline of 0.30 points; varicella coverage did not fall at all, but increased by 1.03 points.

In investigating the short- and long-term impact of exposure we uncover a more complicated pattern of causality than analysis of number of ISPs can provide. The impact of broadband varies over time. Initial exposure drives up the full coverage rate, driven by Hepatitis B and varicella. It decreases demand for Personal Belief and Conditional Exemptions. With longer periods of exposure these effects are reversed. Vaccine uptake declines and exemptions increase. DTP and polio are two important exceptions, with an initial negative impact and (in the case of polio) a long-run positive one. The results suggest that in general, the early impact of broadband is to render consumers more conscientious about health care decisions and increase vaccine intentions. Information about official vaccine guidelines, and where vaccination services are offered, becomes far more accessible. Over time, however, exposure causes consumers to feel overloaded with information and exposes them to misinformation that shifts vaccine perceptions. Despite the large body of online scientific evidence that demonstrates the contrary, broadband users come to view vaccines as inherently riskier and exemptions as a safer strategy for their children.

In previous sections I have captured broadband exposure using two binary (*broadband* and *competitive broadband*) and one continuous (number of providers) variables. The second robustness check is to substitute *competitive broadband* with several new indicators that allow for ever-finer categorization of provision, thus bridging the gap between the *competitive broadband* and number of providers variables. The indicators are:

Competitive Broadband, 3 Categories = 0, if 0-3 providers; 1, if 4-6; 2, if 7+.
Competitive Broadband, 4 Categories = 0, if 0-3 providers; 1, if 4-6; 2, if 7-9; 3, if 10+.
Competitive Broadband, 5 Categories = 0, if 0-3 providers; 1, if 4-6; 2, if 7-9; 3, if 10-14; 4, if 15+.

Figure A5 demonstrates how California's public schools were distributed across the final five categories of broadband provision throughout the 2000-2008 period. In the year 2000, the majority of schools were clustered in the 0-3 and 4-6 provider categories. By 2008, the first category had all but disappeared and the second was dwarved by the 10-14 and 15+ categories, which experienced explosive growth. The share of observations falling in the intermediate 7-9 category slowly declined before rebounding in 2006 and 2007, only to decline again in 2008. By 2008, 84% of all schools were located in a zipcode with 10 or more broadband providers.

Coefficient estimates for the analysis of heterogeneous effects by number of providers are presented in Tables A14-A17. The coefficients of each category are reported separately; the 0-3 category is excluded. Examining the results, it is immediately clear that the estimates are similar in magnitude and significance across different categories, and similar to the *competitive broadband* estimates for public schools. As a consequence of broadband exposure, the full coverage rate falls (though this estimate tends not to survive the inclusion of district trends), while Personal Belief Exemptions rise. Conditional Exemptions also increase, but there is no significant impact on Personal Medical Exemptions. Coverage rates for the DTP, MMR, Hepatitis B and varicella series all decline. Only a single polio coefficient is statistically significant.

In examining the UTD and exemption rates, we find that coefficients tend to increase in magnitude for higher categories of broadband provision. This result does not hold for any of the individual series, however; instead estimates grow smaller or display non-monotonic behavior. The categorical variables analysis supports the assumption that a simple binary measure distinguishing between "many" and "few or no" providers can successfully capture the heterogeneity of broadband provision. Indeed, when the binary *competitive broadband* indicator is replaced with competitive broadband, 3, 4 and 5 categories, the results (available upon request) are extremely similar.

8 Mechanisms

In this section I focus on interpreting the coefficient estimates and investigate the underlying causal mechanisms. I discussed the attributes associated with support for mandatory vaccina-

tion laws and with broadband uptake in Section 3.3.1. Unfortunately I cannot directly observe what types of parents use web resources to learn about immunization, how they undertake research online, nor how they use the new information to update their beliefs. The Google Trends (GT) platform does however allow us to access the relative frequency with which localities search for specific terms in a given time period; it thus represents an index of online interest in that topic. In this section I demonstrate how a specific online behavior- Google searches- is correlated with kindergarten vaccination rates in California. I introduce infectious disease data into the analysis to investigate how real-world disease dynamics interact with online search of vaccine-related terms to drive the ultimate outcome of immunization.

Google Trends communicates the relative popularity of each search term across different localities by producing a ranked hierarchy of scores, with a value of "100" assigned to the locality in which the search term was most popular. In the year 2007, for example, the search topic "Vaccine" was most popular in the metropolitan area of Watertown, New York.[7] The Yuma Arizona-El Centro California MSA had the 24th highest frequency of searches and receives a score of 47, meaning that its search frequency amounted to 47% of that of Watertown. The lowest-ranked MSAs, Alpena Michigan and Glendive Montana, had a frequency of 0% relative to Maryland, and so receive a score of 0.

I translated Google Trends MSA data to the county level using the University of Missouri's MABLE Geocorr correspondence tool,[8] and subset the resulting dataset to the state of California. County-level averages of kindergarten immunization rates were obtained from the California Department of Public Health. Because kindergarten averages are only available from 2007 on, my analysis is limited to the 2007-08 period. Infectious disease rates by county were also obtained for the CDPH. Finally, I obtained control variable data from the U.S. Census American FactFinder website. The county-level indicators included in the analysis are natural log of kindergarten enrollment, population density, employment rate, family poverty rate, share population with college degree of higher, share hispanic, share black, share foreign-born, share population aged 0-4 and share aged 5-9.

[7] Google trends allows researchers to view the popularity of words and phrases defined as either "search terms" or "topics." Search terms have specific spelling and capitalization; "vaccine" returns different (if similar) results to "vaccines." Topics instead incorporate different spellings and formulations of the same concept: "varicella," "varcella," "chicken pox" and "chickenpox" all fall under the same topic of "varicella." I therefore use the topic rather than search term results for a given concept whenever possible.

[8] Available at http://mcdc.missouri.edu/applications/geocorr2000.html.

California experienced outbreaks of three vaccine-preventable diseases in 2006 and 2007: Measles, tetanus and Hepatitis B. I introduce each individual rate into the regressions for MMR, DTP and Hepatitis B. For other regressions, I average the three rates to obtain a single vaccine-preventable disease rate, measured as number of cases per 100,000 people. I interact the VPD rate with each Google search term to investigate how the actual prevalence of infection in the population mediates the relationship between online activity and immunization outcomes.

8.1 Google Trends Results

Table 23 presents the estimated coefficients for the regressions of % UTD, PBE, PME and Conditional Exemptions on the generic search term "vaccine." The term is negatively correlated with Personal Belief Exemptions; a one-standard deviation increase in a geography's GT score is associated with a 0.18-point drop in PBEs. In the absence of a disease outbreak, searches for vaccine have no effect on % UTD. Average VPD rate is negatively correlated with the full coverage rate and positively correlated with each of the exemptions. If both disease prevalence and vaccine searches increase by one s.d., the full coverage rate drops by half a point; the negative effect of the VPD rate offsets the positive one of the search.

Results for DPT-related search terms are presented in Tables 24-27. Searches for "diptheria" are associated with rising DTP and full coverage rates, and falling PBE and Conditional Exemption rates. Increased prevalence of tetanus is instead linked to the decline of DTP and full coverage rates, as well as PBE and PME. Only Conditional Exemptions rise with tetanus rates. At higher rates of disease prevalence, the beneficial impact of diphtheria searches is erased; they are associated with falling immunization coverage and more PBEs.

"Whooping cough" searches are associated with an increase in the full coverage rate and decreasing PBEs. DTP coverage itself is not impacted. The tetanus rate is only significant in conjunction with whooping cough searches: Once more, higher prevalence triggers a negative response in the immunization rates, and a positive one in PBE. No coefficient estimates result significant for the search term "tetanus."

"DTP vaccine" search term estimates run in the opposite direction of those pertaining to the

individual diseases. Enhanced interest in the vaccine is correlated with lower immunization rates. Conditional exemptions increase. PMEs fall, though the estimated effect is quite small and significant at the 10% level. In this case, the interaction between searches and prevalence causes exemption rates to fall (without impacting coverage).

With no infectious disease outbreaks, "polio" searches influence Conditional Exemptions negatively, while "polio vaccine" searches are positively correlated (Tables 28 and 29). The full coverage rate declines with an increase in VPD prevalence. The impact on exemptions is ambiguous. The joint effect of prevalence and polio-related searches upon % UTD is positive, a result largely unique to the polio results; the joint effect on polio coverage itself is negative.

Turning to the MMR results (Tables 30-33), the "measles" search term estimates stand out. Searches for "measles" are correlated with rising MMR and full coverage rates and declines in every type of exemption. Interest in "rubella" is likewise associated with increased MMR coverage and decreased PME. The contrast with "MMR vaccine" searches is stark: While largely orthogonal to exemption rates, interest in the MMR vaccine is associated with drops in both MMR and total immunization coverage. "Mumps" searches are not correlated with anything except PME (in conjunction with the disease rate; the estimated effect is negative). The interaction between prevalence and GT searches is almost never significant, but does cause total immunization coverage to drop even further in the case of "MMR vaccine."

The MMR results largely repeat the pattern set by the DTP-related searches, with one important difference. Searching for individual diseases results in higher coverage rates and fewer exemptions, as before. Searching for the vaccine that protects against those diseases results in lower coverage, as before. The novelty is the impact of measles prevalence. For the search term "measles," it is negatively correlated with MMR coverage, as occurred for DTP; but in the case of "rubella," it is positively associated with % UTD. This does not occur for any other vaccine series or search term: The negative relationship between prevalence and immunization behavior is, in fact, one of the most consistent findings of the GT analysis.

Interest in "Hepatitis B" is orthogonal to immunization behavior (Table 34). The only significant result in this table is the familiar negative relationship between disease prevalence and vaccine uptake, a result repeated (for MMR uptake) for "hepatitis B vaccine." The relationship between interest in the Hepatitis B vaccine and immunization is complicated (Table 35). As

before, curiosity about the vaccine results in reduced demand for it and increased requests for PBEs. However, it is also associated with fewer PMEs and Conditional Exemptions, and higher demand for vaccines overall. Likewise, prevalence of Hepatitis B is associated with lower demand for the Hepatitis B vaccine and an increase in PBEs, but also fewer PMEs. The joint effect of prevalence and hepatitis B vaccine searches is to lower the full coverage rate.

Searches for "chickenpox" are uncorrelated with vaccine demand (Table 36). Interest in "varicella vaccine" is associated with fewer PBEs and more Conditional Exemptions (Table 37). In both cases, prevalence is correlated negatively with uptake and positively with exemptions. Prevalence and "varicella vaccine" searches interact to increase PBEs and reduce Conditional Exemptions.

Several clear patterns emerge from the GT analysis. Curiosity about individual diseases is tied to an increase in vaccine uptake and falling demand for exemptions; the generic term "vaccine" is also associated with a dccline in PBEs. In contrast, interest in the vaccines that protect against these diseases is tied to decreased uptake and rising exemption demand. The specific websites visited by health consumers after searching for these terms are unobservable. It seems probable that disease-specific searches led to information that communicated the danger of vaccine-preventable diseases, and so heightened the perceived risk of not vaccinating. Vaccine-specific searches may have instead guided consumers to the anti-vaxx web, exposing them to websites that minimized or ignored the dangers of VPDs while focusing upon the dangers posed by vaccines and the vaccine industry. This would shift risk perceptions in the other direction, and reduce vaccine intentions. In several cases, the interaction between Google searches and VPD prevalence is associated with lower uptake and higher exemptions, suggesting that real-world disease dynamics can generate curiosity about diseases and vaccines that actually results in lower intention to vaccinate. These findings are especially marked for DTP and MMR, two of the best-known and most controversial vaccines, frequently linked in popular wisdom to adverse outcomes.

The Hepatitis B, varicella and polio results are more ambiguous, if coherent with the actual dynamics of each series from 2000-2008 and with the coefficient estimates presented earlier. Across the different specifications discussed above, the Hepatitis B and varicella coefficients tend to be negative, but are not robust to the inclusion of district trends. GT polio searches

are linked in a clear and consistent way only to Conditional Exemptions. In the main analysis, polio estimates have differed from zero only in a few scattered cases. This combination of results is unsurprising, as the polio vaccine is less controversial than others and the disease is well-known.

One final pattern, consistent across all series, is the inverse relationship between VPD prevalence and vaccine uptake. This relationship is at odds with the prevalence elasticity of demand estimated for various types of preventative care (Folland). When disease rates are lower, parents have the incentive to allow their children to free ride on the herd immunity provided by others; with increased prevalence, the opposite should be true. The most plausible explanation for why the opposite pattern emerges in my dataset is the relative infrequency of VPD outbreaks from 2000-2008. Varicella, mumps and pertussis were not yet present in the population. Measles cases were still rare. As a non-communicable disease, tetanus provoked little concern. Although vaccine coverage rates were declining and VPD rates gradually rising, the public was relatively unaware of it as a serious public health issue in the years leading up to the Disneyland measles outbreak. For many parents, vaccines presented the greater danger. Thus vaccine coverage continued to fall in the very localities where VPDs grew more pervasive.

This explanation finds support in the quadratic relationship between months of internet exposure and vaccine uptake presented in Section 7. During this time period, the initial impact of broadband was to increase vaccine uptake and reduce demand for exemptions. As length of exposure increases, however, the role of broadband transforms. Online activity causes parents to become more wary of most vaccines and to request exemptions at a greater rate. This long-run effect seems to have prevailed. Since 2008, broadband growth has slowed as the market reaches saturation, yet immunization rates have continued to decline: Resulting in new, ever-more widespread outbreaks of vaccine-preventable diseases in California.

9 Conclusions

In this paper, I show that access to a competitive broadband market represents a good proxy for broadband demand and use. I find that expansion of broadband across California's zipcodes from 2000 to 2008 caused immunization uptake to fall, coupled with an increase in demand for

exemptions. In particular, the proportion of fully-immunized pupils in public kindergartens fell by 0.83 percentage points as a result of broadband exposure: Equal to 71% of the total decline in uptake. Four out of five vaccine series- DTP, MMR, Hepatitis B and varicella- fell. Personal Belief Exemptions rose by .31 points, or one-quarter of the increase in that variable. There is some evidence that uptake increased and demand for exemptions declined in California's private kindergartens, though estimation results are less robust and subject to data concerns.

Broadband played a heterogeneous role over time. Initial exposure caused parents to increase demand for most vaccines and request fewer exemptions. The full coverage rate rose. For longer periods of time, however, the effect is reversed. The role of internet thus shifts over the course of the decade. If access initially enabled consumers to become more informed about healthcare decisions and assume greater responsibility for their childrens' medical care, such empowerment proved a net negative when anti-vaxx content online overpowered public health messaging, leaving parents confused and mistrustful.

The analysis of how Google searches and disease prevalence fed into immunization behavior- though correlative, and not evidence of causation- provides some clue as to how broadband access may have evolved into a threat to coverage rates. Searches for individual vaccine-preventable diseases, as well as the generic term "vaccine," led parents to vaccinate their children at higher rates, presumably by increasing awareness of the danger posed by VPDs. In contrast, searches for specific vaccines appear to have lowered vaccine intentions and rates of uptake, perhaps by guiding parents to anti-vaxx websites that shifted their beliefs about the risks of vaccination. Disease rates are negatively correlated with demand for vaccines during this period, highlighting the lack of awareness on the part of many parents of how their choices were slowly eroding the herd immunity enjoyed by their communities.

The long-run negative impact of broadband access upon most immunizations is ominous, and can only be counteracted by more online discussion of the risks and benefits of vaccines that is grounded in scientific evidence. One promising avenue for future research is to experiment with employing the same tactics as anti-vaxx websites: First, by engaging with the same audience of parents deeply concerned about safeguarding their childrens' health; second, by presenting medical information and statistics in a more engaging and interactive way, so as to increase their salience; and third, by publishing emotionally-rich narratives drawn from the experiences

of other parents which bring to light the dangers of vaccine-preventable diseases and the power of vaccines to save lives.

References

[1] Abbate, J. 1999. *Inventing the Internet*. Cambridge: MIT Press.

[2] Andersen, T.B., J. Bentzen, C. Dalgaard and P. Selaya. 2011. "Does the Internet Reduce Corruption? Evidence From U.S. States and Across Countries." The World Bank Economic Review, 25(3): 387-417.

[3] Andersen, T.B., J. Bentzen, C. Dalgaard and P. Selaya. 2012. "Lightning, IT Diffusion, and Economic Growth Across U.S. States." *The Review of Economics and Statistics*, 94(4): 903-924.

[4] Bauch, C.T. and D.J.D. Earn. 2004. "Vaccination and the Theory of Games." *PNAS*, 101(36): 13391:13394.

[5] Betsch, C., N.T. Brewer, P. Brocard, P. Davies, W. Gaissmaier, N. Haase, J. Leask, F. Renkewitz, B. Renner, V.F. Reyna, C. Rossmann, K. Sachse, A. Schachinger, M. Siegrist and M. Stryk. 2012. "Opportunities and Challenges of Web 2.0 for Vaccination Decisions." *Vaccine*, 30(25): 3727-3733.

[6] Betsch, C., F. Renkewitz, T. Betsch and C. UlshÃ¶fer. 2010. "The Influence of Vaccine-Critical Websites on Perceiving Vaccination Risks." *Journal of Health Psychology*, 15(3): 446-55.

[7] Betsch C., C. UlshÃ¶fer, F. Renkewitz and T. Betsch. 2011. "The Influence of Narrative v. Statistical Information on Perceiving Vaccination Risks." *Medical Decision Making*, 31(5): 742-53.

[8] Billari, F.C., O. Giuntella and L. Stella. 2019. "Does Broadband Internet Affect Fertility?" *Population Studies*.

[9] Boseley, S. 2018. "How Disgraced Anti-Vaxxer Andrew Wakefield Was Embraced By Trump's America." *The Guardian.*

[10] Brownlie, J. and A. Howson. 2005. "Leaps of Faith and MMR: An Empirical Study." *Sociology*, 39(2): 221-239.

[11] Campante, F., R. Durante and F. Sobbrio. 2013. "Politics 2.0: The Multifaceted Effect of Broadband Internet and Political Participation." NBER Working Paper 19029.

[12] Canadian Radio-television and Telecommunications Commission. 2013. "Price Comparisons of Wireline, Wireless and Internet Services in Canada and with Foreign Jurisdictions."

[13] Centers for Disease Control and Prevention- National Center for Immunization and Respiratory Diseases. 2015. "Vaccines and Immunizations: Glossary." https://www.cdc.gov/vaccines/terms/glossary.html

[14] Chong, A., M. Gonzalo-Navarro, D. Karlan and M. Valdivia. 2013. "Effectiveness and Spillovers of Online Sex Education: Evidence From a Randomized Evaluation in Columbian Public Schools." NBER Working Paper 18776.

[15] Clemmons, N.S., P.A. Gastanaduy, A.P. Fiebeikorn, S.B. Redd, and G.S. Wallace. 2015. "Measles- United States, January 4-April 2, 2015." Centers for Disease Control and Prevention Morbidity and Mortality Weekly Report.

[16] Clemmons, N.S., G.S. Wallace, M. Patel and P.A. Gastanaduy. 2017. "Incidence of Measles in the United States, 2001-2015." *Journal of the American Medical Association*, 318(3): 1279-1281.

[17] Delamater, P.D., T.F. Leslie and Y.T. Yang. 2017. "Change in Medical Exemptions for Immunization in California After Elimination of Personal Belief Exemptions." *Journal of the American Medical Association*, 318(9): 863-864.

[18] DellaVigna, S. and E. Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 879–924.

[19] Dempsey, A.F., S. Schaffer, D. Singer, A. Butchart, M. Davis and G.L. Freed. 2011. "Alternative Vaccination Schedule Preferences Among Parents of Young Children." *Pediatrics*, 128(5): 848-856.

[20] Dupas, P., E. Huillery and J. Seban. 2018. "Risk Information, Risk Salience, and Adolescent Sexual Behavior: Experimental Evidence from Cameroon." *Journal of Economic Behavior & Organization*, 145: 151-175.

[21] Economides, N. 2004. "Telecommunications Regulation: An Introduction."

[22] European Center for Disease Prevention and Control. 2012. "Communication on Immunisation – Building Trust." Stockholm: ECDC.

[23] Fangjun, Z., A. Shefer, J. Wenger, M. Messonnier, L.Y. Wang, A. Lopez, M. Moore, T.V. Murphy, M. Cortese and L. Rodewald. 2014. "Economic Evaluation of the Routine Childhood Immunization Program in the United States, 2009." *Pediatrics*, 133(4): 577-585.

[24] Falck, O., R. Gold and S. Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review*, 104(7): 2238-2265.

[25] Federal Communications Commission. 1998. "First Broadband Progress Report."

[26] Federal Communications Commission. 2000a. "Federal Communications Commission Releases Data on High-Speed Services for Internet Access."

[27] Federal Communications Commission. 2000b. "Deployment of Advanced Telecommunications Capability: Second Report."

[28] Federal Communications Commission. 2002. "Third Broadband Progress Report."

[29] Federal Communications Commission. 2008. "Fifth Broadband Progress Report."

[30] Feiken, D.R., D. C. Lezotte, R.F. Hamman, D.A. Salmon, R.T. Chen and R.E. Hoffman. 2000. "Individual and Community Risks of Measles and Pertussis Associated with Personal Exemptions to Immunizations." *JAMA*, 284(24): 3145-3150.

[31] Fine, P., K. Eames and D.L. Heymann. 2011. "Immunizations: A Rough Guide." *Clinical Infectious Diseases*, 52(7): 911-916.

[32] Folland, S., A.C. Goodman and M. Stano. 2012. *The Economics of Health and Health Care*. Upper Saddle River, NJ: Prentice Hall.

[33] Fox, S. 2005. "Health Information Online." Pew Internet & American Life Project.

[34] Fox, S. 2006. "Online Health Search 2006." Pew Internet & American Life Project.

[35] Fox, S. and S. Jones. 2009. "The Social Life of Health Information." Pew Internet & American Life Project.

[36] Fox, S. and L. Rainie. 2000. "The Online Health Care Revolution." Pew Internet & American Life Project.

[37] Funk, C. 2015. "Public and Scientists' Views on Science and Society." Pew Research Center.

[38] Funk, C., B. Kennedy and M. Hefferon. 2017. "Vast Majority of Americans Say Benefits of Childhood Vaccines Outweigh Risks." Pew Research Center.

[39] Gavazza, A., M. Nardotto and T. Valletti. 2016. "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies."

[40] Gentzkow, M. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics*, 121(3): 931-972.

[41] Glanz, J.M., D.L. McClure, D.J. Magid, M.F. Daley, E.K. France., D.A. Salmon and S.J. Hambidge. 2009. "Parental Refusal of Pertussis Vaccination is Associated With an Increased Risk of Pertussis Infection in Children." *Pediatrics*, 123(6): 1446-51.

[42] Greenstein, S. 2015. *How the Internet Became Commercial*. Princeton: Princeton University Press.

[43] Heblich, S. 2016. "The Effect of Internet on Voting Behavior." *IZA World Of Labor*, 294.

[44] Hinman, A.R., W. A. Orenstein and A. Schuchat. 2011. "Vaccine-Preventable Diseases, Immunizations, and MMWR — 1961–2011." *Morbidity and Mortality Weekly Report Supplements*, 60(04): 49-57.

[45] Hobson-West, P. 2007. "'Trusting Blindly Can Be the Biggest Risk of all': Organised Resistance to Childhood Vaccination in the UK." *Sociology of Health and Illness*, 29(2):198-215.

[46] Holmes, Allan and C. Zubak-Skees. 2015. "U.S. Internet Users Pay More and Have Fewer Choices Than Europeans." The Center for Public Integrity.

[47] Horrigan, J.B. 2006. "Home Broadband Adoption 2006." Pew Internet & American Life Project.

[48] Horrigan, J.B. 2006. "The Internet as a Resource for News and Information about Science." Pew Internet & American Life Project.

[49] Horrigan, J.B. 2007. "Broadband: What's All the Fuss About?" Pew Internet & American Life Project.

[50] Horrigan, J.B. and L. Rainie. 2002. "The Broadband Difference: How Online Behavior Changes With High-speed Internet Connections." Pew Internet & American Life Project.

[51] Horrigan, J.B. and M. Duggan. 2015. "Home Broadband 2015." Pew Research Center.

[52] Howard, J. 2017. "Anti-vaccine Groups Blamed in Minnesota Measles Outbreak." *CNN*, http://edition.cnn.com/2017/05/08/health/measles-minnesota-somali-anti-vaccine-bn/index.html.

[53] Hussain, H., D. Kehl, B. Lennett, C. Li and P. Lucey. 2012. "The Cost of Connectivity: A Comparison of High-Speed Internet Prices in 22 Cities Worldwide. New America Foundation.

[54] Jensen, R. and E. Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *The Quarterly Journal of Economics*, 124(3): 1057–1094.

[55] Jolley, D. and K.M. Douglas. 2014. "The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions." *Plos One*, 9(2).

[56] Kata, A. 2010. "A Postmodern Pandora's Box: Anti-Vaccination Misinformation on the Internet." *Vaccine*, 28: 1709-1716.

[57] Kearney, M.S. and P.B. Levine. 2014. "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." NBER Working Paper 19795.

[58] La Ferrara, E. 2016. "Mass Media and Social Change: Can We Use Television to Fight Poverty?" *Journal of the European Economic Association*, 14(4): 791-827.

[59] La Ferrara, E., A. Chong and S. Duryea. 2008. "Soap Operas and Fertility: Evidence from Brazil." Inter-American Development Bank Working Paper 633.

[60] Lahart, J. 2009. "There Go the Servers: Lightning's New Perils." *The Wall Street Journal*, March 22.

[61] Larcinese, V. and L. Miner. 2017. "The Political Impact of the Internet on U.S. Presidential Elections."

[62] Lelkes, Y., G. Sood and S. Iyenger. 2015. "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Effect." *American Journal of Political Science*, 61(1): 5-20.

[63] Lewandowsky, S., G.E. Gignac and K. Oberauer. 2013. "The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science." *Plos One*, 8(10).

[64] Majunder, M.S., E.L. Cohn, S.R. Mekaru, J.E. Huston and J.S. Brownstein. 2015. "Substandard Vaccine Compliance and the 2015 Measles Outbreak." *JAMA Pediatrics*, 169(5): 494-495.

[65] Meckel, R.A. 2004. "Levels and Trends of Death and Disease in Childhood, 1620 to the Present." *Children and Youth in Sickness and Health: A Handbook and Guide*, ed. J. Golden, R.A. Meckel and H.M. Prescott. Westport, CT: Greenwood Press.

[66] Meer, J. and J. West. 2016. "Effects of the Minimum Wage on Employment Dynamics." *Journal of Human Resources*, 51(2): 500-522.

[67] Meikle, J. and Sarah Boseley. 2010. "MMR row doctor Andrew Wakefield struck off register." *The Guardian*.

[68] Mesch, G.S. and K.P. Schwirian. 2014. "Confidence in Government and Vaccination Willingness in the U.S." *Health Promotion Internation*, 30(2): 213-221.

[69] National Conference of State Legislatures. 2016. "States with Religious and Philosophical Exemptions from State Immunization Requirements." http://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx

[70] National Telecommunications and Information Administration. 1998. "Falling Through the Net II: New Data on the Digital Divide."

[71] National Telecommunications and Information Administration. 2013. "Exploring the Digital Nation: America's Emerging Online Experience."

[72] Nichols, T. 2017. *The Death of Expertise*. Oxford: Oxford University Press.

[73] Oster, E. 2017. "Does Disease Cause Vaccination? Disease Outbreaks and Vaccination Response."

[74] Otto, S. 2016. *The War on Science.* Minneapolis: Milkweed Editions.

[75] Pew Research Center. 2015. "Americans, Politics and Science Issues."

[76] Phadke, V.K., R.A. Bednarczyk, D.A. Salmon and S.B. Omer. 2016. "Association Between Vaccine Refusal and Vaccine-Preventable Diseases in the United States." *JAMA*, 315(11): 1149-1158.

[77] Poland, G.A. and R.M. Jacobson. 2011. "The Age-Old Struggle Against the Antivaccinationists." *The New England Journal of Medicine*, 364(2): 97-99.

[78] Prieger, J. and M. Connolly. 2013. "A Basic Analysis of Entry and Exit in the U.S. Broadband Market, 2005-2008." Pepperdine University School of Public Policy Working Papers.

[79] Saada, A., T.A. Lieu, S.R. Morain, B.J. Zikmund-Fisher and E. Wittenberg. 2015. "Parents' Choices and Rationales for Alternative Vaccination Schedules: A Qualitative Study." *Clinical Pediatrics*, 54(3): 236-243.

[80] Siegried, J.J. and L.B. Evans. 1994. "Empirical Studies of Entry and Exits: A Survey of the Evidence." *Review of Industrial Organization*, 9(2): 121-155.

[81] Smith, A. 2011. "The Internet and Campaign 2010." Pew Internet & American Life Project.

[82] Snyder, J.M. and D. Stromberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy*, 118(2): 355-408.

[83] Stromberg, D. "Radio's Impact on Public Spending." 2004. *The Quarterly Journal of Economics*, 119(1): 189–221.

Appendix

Evidence From Alternative Measures of Broadband Provision

In this section, I present coefficient estimates for the full coverage rate and four exemption series, substituting *broadband* and number of providers for the *competitive broadband* indicator. The *broadband* variable influences belief exemptions positively, with an estimated effect equal to 23% of the sample mean, yet has no significant impact upon the other outcomes (Tables A1 and A2). Turning to the provider indicator results (Tables A3 and A4), we find that one standard-deviation increase in number of providers causes a 0.88-point drop in the coverage rate, similar in magnitude to the *competitive broadband* estimate. Conditional exemptions rise by 0.79 points. A change in the number of providers does not influence PBEs or PMEs, however.

Private School Estimation Results

The private school model parallels that estimated for the set of public schools, presented in Section 3.3, save that the charter dummy is substituted by a dummy for religious schools, R, and control variables are no longer measured at the school level:

$$v_{idt} = \gamma Net_{idt} + \phi Rt + \beta X_{dt-1} + \alpha_i + \tau_t + dt + \varepsilon_{zt}$$

where school *i* is nested within district *d* and observed in year *t*. *Net* is the *competitive broadband* indicator. The immunization outcome of interest is *v*, measured in percentages.

School-level control variables are available only for public schools and a very small number of private ones. I account for this problem by aggregating control variable observations up to the district levels. Thus while the public school analysis features school-level indicators, specific to those particular schools, the private school analysis includes district-level indicators that essentially represent the mean value for public schools in the same district.

The private school data present a challenge in two respects. First, public school children must attend the kindergarten they are zoned for, given their home address. Thus in most cases the school address zipcode is likely to coincide with the zipcode of the neighborhood in which a pupil resides. Yet families select into private schooling. Parents may choose to enroll their child in a private school that is located several zipcodes away. The correspondence between the zipcode given by the school address, and the zipcode in which families reside, is imperfect. This mismatch will result in attenuation bias; the coefficient estimates obtained on the private school sample may be far smaller in magnitude than the true population relationship. While the public school results can be analyzed with a fair amount of confidence, in the case of private schools we should not accept a coefficient indistinguishable from zero as concrete evidence of the absence of a relationship. In statistical terms, we merely cannot reject the null hypothesis of no relationship, and the estimates do no reflect causal effects.

Second, because the California Department of Education does not collect data on private school students we lack information on how the public- and private-school populations differ, and how families select into private schooling. We can observe number of public and private schools per district per year, however. The average school district has 22 private schools, with a minimum of 0 and a maximum of 351. In order to examine how districts with high and low demand for public schooling differ, I divided the dataset into quartiles based upon proportion of schools that are private. In top-quartile districts, public schools tend to be wealthier, with fewer students eligible for free and reduced meals, and parents are slightly more educated on average. A smaller share of teachers have emergency credentials, indicating lower turnover and overall higher quality of staff. Class size and school test scores are essentially identical for the top and bottom quartiles, however. This evidence suggests that private schools are not a substitute for failing public schools; instead, there is greater demand for private schooling in more affluent, educated districts where teaching quality is somewhat higher. In demographic terms, top-quartile school districts tend to be more diverse, with a larger share of Asian and black students, a lower share of white students, and the same number of Hispanics. These districts also have more English learners.

Turning to the estimation results, we find that private schools demonstrate very different immunization behavior from public schools. Moving from a non-competitive to competitive

broadband regime is associated with an increase in private school PBEs, but no concordant fall in coverage rates; indeed the UTD coefficient, though insignificant, is positive (Table A5). The significant decrease in Conditional Exemptions may explain this result (Table A6). The most unexpected result is the significant negative relationship between religious school status and belief exemptions, which survives no matter which broadband indicator is used. This is particularly surprising given that PBEs encompass religious objections to immunization, and suggests interesting implications for the small set of secular private kindergartens.

The competitive broadband indicator performs poorly in explaining variation in individual series rates among the set of private schools (Tables A7 and A8). The MMR estimate is positive and significant at the 10% level, but the district trend causes it to lose significance. Religious schools continue to have significantly higher coverage rates, specifically in the cases of MMR and Hepatitis B.

The coefficient estimates suggest that private and public school families respond in a diametrically opposed way to broadband treatment, though the private school findings- as discussed above- are merely indicative, not causal. As public school households enjoy greater exposure to high-speed internet, fewer parents choose to immunize their children. These children are admitted to school with Personal Belief and Conditional Exemptions. Charter school families, in particular, immunize less and exempt more: An indication of their underlying dissatisfaction with traditional public education and, by extension, perhaps government regulations in general. In contrast, a greater share of private school families choose to comply with state immunization recommendations. Personal Belief Exemptions fall. These behaviors are particularly marked among families that select into religious schooling.

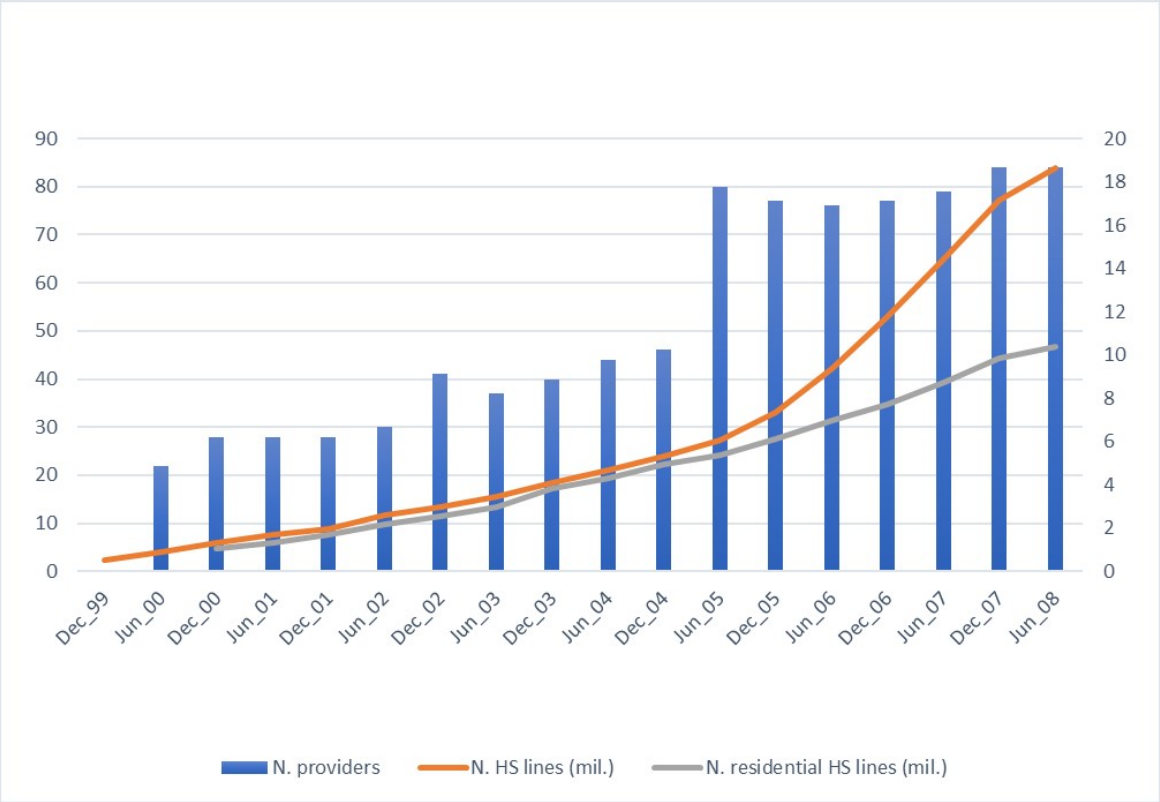
Selection

The two validity checks presented in section 5 are here reproduced for the set of private schools. The graphical analysis of the parallel trends assumption is presented in Figures A3 and A4. The private school estimates for competitive broadband are all equal to zero, with two exceptions. The Conditional Exemption coefficient is negative in year t+3. This is coherent with the negative coefficient estimate presented in Table A6. The t-1 effect on PME is positive

and significant at the 10% level; as before, however, the overall coefficient estimate for PME is equal to zero.

In Table A9, the results for the regression of current immunization rates upon future competitive broadband are presented. Leads of the competitive broadband indicator are never significant.

Figure 1: Broadband diffusion in California, 1999-2008



Source: Elaboration of FCC Form 477 data.

Figure 2: Broadband diffusion in California by zipcode, 1999-2008

Notes: "N. providers" refers to the number of ISPs serving a zipcode. "BB access" is the proportion of zipcodes with access to at least one ISP. "High BB access" is the proportion of zipcodes with access to four or more ISPs.

Figure 3: Diffusion of competitive BB across California ZCTAs, June 2000-June 2008.

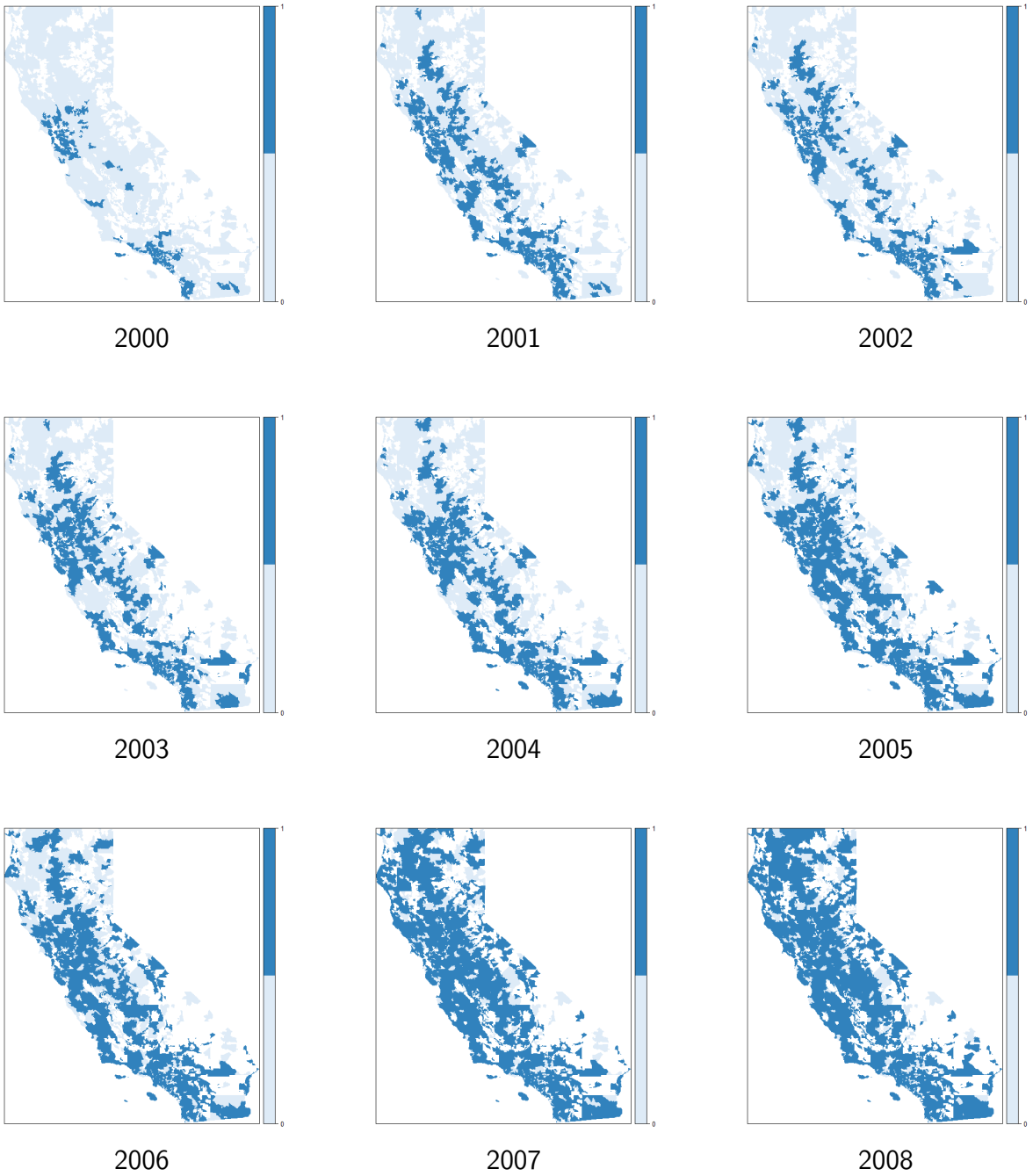
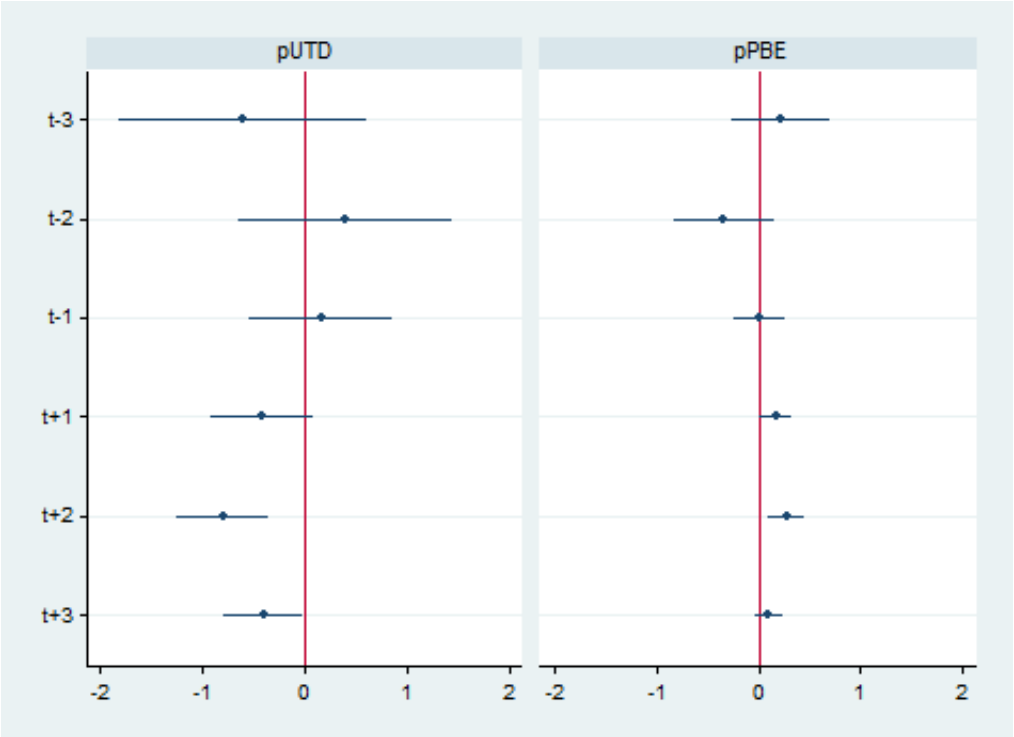
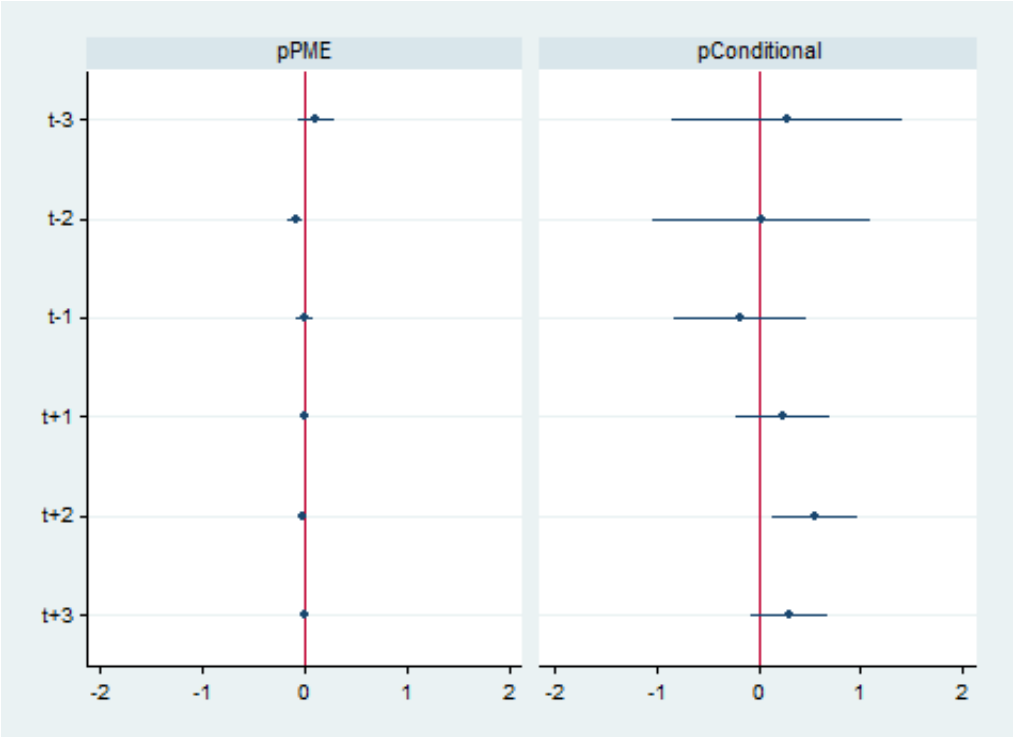


Figure 4: Public schools
Effect of competitive broadband on immunization outcomes over time



Notes: The left figure plots coefficient estimates for the impact of "competitive broadband" time dummies on percent Up-To-Date. The right figure plots estimates for the impact of the dummies on Personal Belief Exemptions. Time 't' is defined as the first year in which a zipcode obtained access to four or more providers. See Table 2 for notes on the specification.

Figure 5: Public schools
Effect of competitive broadband on immunization outcomes over time



Notes: The left figure plots coefficient estimates for the impact of "competitive broadband" time dummies on Personal Medical Exemptions. The right figure plots estimates for the impact of the dummies on Conditional Exemptions. Time 't' is defined as the first year in which a zipcode obtained access to four or more providers.

Table 1: Correlation Between Provider Measures and Residential High-Speed Lines

	(1)	(2)	(3)	(4)
	Res. HS Lines	Res. HS Lines	Res. HS Lines	Res. HS Lines
BB	7.544***	1.009**		
	(.718)	(.391)		
Competitive BB			4.132***	.504**
			(.187)	(.215)
R2	.608	.944	.739	.941
Number of Providers	.569***	-.016		
	(.021)	(.033)		
State Providers			.028***	-.001
			(.005)	(.001)
R2	.745	.940	.520	.940
Year F.E.		X		X
N	803	803	803	803
Mean of dependent	12.583	12.583	12.583	12.583

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The observed sample is U.S. states from December 1999 to June 2008. All regressions include time and state fixed effects. "BB" is proportion of zipcodes in the state with access to at least one provider; "Competitive BB" is proportion of zipcodes with access to four or more providers; "Number of Providers" is mean number of providers per zipcode; "State Providers" is number of ISPs serving the entire state. Standard errors are clustered at the state level.

Table 2: Effect of Competitive BB on Public School Outcomes: UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Competitive BB	-.827***	-.387	.220***	.311***
	(.296)	(.312)	(.067)	(.081)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.475	.493	.489	.515
N	16071	16071	16071	16071
Mean of dependent	91.805	91.805	1.699	1.699

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. The school controls are a dummy for charter schools interacted with a linear trend, beginning-of-school-year API, growth in API, the School Characteristic Index, pupil turnover, average K-12 class size, proportion of teachers with emergency credentials, proportion of students eligible for free or reduced meals, average educational level of parents, share of English-learning or English-proficient students, and the kindergarten population shares for Native American, Asian, Pacific Islander, Filipino, Hispanic and black. Standard errors are clustered at the zipcode level.

Table 3: Effect of Competitive BB on Public School Outcomes: PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Competitive BB	.016	-.023	.592**	.099
	(.029)	(.035)	(.293)	(.307)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.102	.117	.476	.494
N	16071	16071	16071	16071
Mean of dependent	0.172	0.172	6.325	6.325

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table 2 for notes on the specification.

Table 4: Effect of Competitive BB on Public School DTP, Polio and MMR Coverage

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Competitive BB	-.395*	-.091	-.297	-.035	-.450***	-.248**
	(.217)	(.217)	(.198)	(.203)	(.136)	(.118)
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X
Adjusted R2	.409	.436	.419	.443	.283	.331
N	16071	16071	16071	16071	16071	16071
Mean of dependent	94.287	94.287	94.681	94.681	97.289	97.289

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table 2 for notes on the specification.

Table 5: Effect of Competitive BB on Public School Hepatitis B and Varicella Coverage

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Competitive BB	-.447***	-.214	-.529*	-.287
	(.154)	(.149)	(.304)	(.334)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.467	.515	.544	.556
N	16071	16071	13931	13931
Mean of dependent	96.286	96.286	96.581	96.581

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table 2 for notes on the specification.

Table 6: Effect of Future Competitive BB on Public School Outcomes

	(1) UTD	(2) PBE	(3) PME	(4) Conditional
Competitive BB	-.815***	.232***	.012	.571**
	(.291)	(.065)	(.028)	(.287)
Future competitive BB	-.085	-.081	.021	.144
	(.423)	(.132)	(.068)	(.409)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
Adjusted R2	.475	.489	.102	.476
N	16071	16071	16071	16071
Mean of dependent	91.805	1.699	0.172	6.325

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in column (1) is the full coverage rate; in (2), percent Personal Belief Exemptions; in (3), Personal Medical Exemptions; and in (4), Conditional Exemptions. See Table 2 for notes on the specification.

Table 7: Impact of Competitive BB by Initial Vaccine Coverage, UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Lower 50% - Coverage	-.885	.118	.223*	.304**
	(.606)	(.648)	(.120)	(.136)
Adjusted R2	.467	.482	.566	.587
N	7140	7140	7140	7140
Mean of dependent	89.596	89.596	1.512	1.512
Upper 50% - Coverage	-.266	-.446	.193**	.304***
	(.218)	(.274)	(.078)	(.098)
Adjusted R2	.278	.291	.293	.319
N	8225	8225	8225	8225
Mean of dependent	95.147	95.147	1.008	1.008
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table 2 for notes on the specification.

Table 8: Impact of Competitive BB by Initial Vaccine Coverage, PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Lower 50% - Coverage	-.005	-.121**	.667	-.302
	(.050)	(.061)	(.606)	(.655)
Adjusted R2	.114	.158	.472	.486
N	7140	7140	7140	7140
Mean of dependent	.163	.163	8.730	8.730
Upper 50% - Coverage	.032	.021	.041	.122
	(.030)	(.045)	(.208)	(.260)
Adjusted R2	.076	.082	.292	.305
N	8225	8225	8225	8225
Mean of dependent	.143	.143	3.702	3.702
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table 2 for notes on the specification.

Table 9: Impact of Competitive BB by Initial Vaccine Coverage, DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Lower 50% - Coverage	-.338	.528	-.358	.312	-.627**	-.142
	(.432)	(.442)	(.397)	(.411)	(.262)	(.224)
Adjusted R2	.387	.414	.404	.426	.262	.313
N	7140	7140	7140	7140	7140	7140
Mean of dependent	93.173	93.173	93.817	93.817	97.157	97.157
Upper 50% - Coverage	-.181	-.309*	-.054	-.110	-.192**	-.217**
	(.146)	(.171)	(.135)	(.157)	(.086)	(.107)
Adjusted R2	.240	.250	.233	.247	.223	.230
N	8225	8225	8225	8225	8225	8225
Mean of dependent	96.730	96.730	97.092	97.092	98.503	98.503
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table 2 for notes on the specification.

Table 10: Impact of Competitive BB by Initial Vaccine Coverage, Hepatitis B and Varicella

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Lower 50% - Coverage	-.529*	.084	-1.322**	-.895
	(.293)	(.276)	(.548)	(.593)
Adjusted R2	.298	.340	.394	.462
N	7140	7140	6126	6126
Mean of dependent	95.668	95.668	96.102	96.102
Upper 50% - Coverage	-.135	-.249*	.195	.356
	(.112)	(.144)	(.407)	(.450)
Adjusted R2	.147	.148	.212	.292
N	8225	8225	7098	7098
Mean of dependent	97.728	97.728	97.968	97.968
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table 2 for notes on the specification.

Table 11: Impact of Competitive BB by Charter Status, UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Charter	10.277**	13.383***	.114	-.110
	(4.614)	(4.759)	(.544)	(.454)
Adjusted R2	.563	.568	.629	.729
N	449	449	449	449
Mean of dependent	83.146	83.146	6.510	6.510
Non-charter	-.944***	-.528*	.221***	.310***
	(.290)	(.301)	(.068)	(.082)
Adjusted R2	.466	.486	.472	.495
N	15613	15613	15613	15613
Mean of dependent	92.508	92.508	1.289	1.289
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table 2 for notes on the specification.

Table 12: Impact of Competitive BB by Charter Status, PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Charter	-.046	.084	-10.347**	-13.357***
	(.161)	(.309)	(4.321)	(4.462)
Adjusted R2	.004	-0.027	.501	.498
N	449	449	449	449
Mean of dependent	.231	.231	10.114	10.144
Non-charter	.017	-.022	.705**	.239
	(.029)	(.036)	(.289)	(.297)
Adjusted R2	.103	.118	.473	.493
N	15613	15613	15613	15613
Mean of dependent	.162	.162	6.040	6.040
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table 2 for notes on the specification.

Table 13: Impact of Competitive BB by Charter Status, DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Charter	5.851*	8.037***	5.667**	7.630***	5.645*	7.558**
	(3.028)	(2.909)	(2.609)	(2.332)	(3.358)	(3.265)
Adjusted R2	.579	.596	.602	.626	.582	.583
N	449	449	449	449	449	449
Mean of dependent	87.544	87.544	87.826	87.826	91.243	91.243
Non-charter	-.466**	-.172	-.366*	-.120	-.484***	-.289**
	(.216)	(.214)	(.196)	(.199)	(.135)	(.115)
Adjusted R2	.397	.426	.404	.430	.260	.312
N	15613	15613	15613	15613	15613	15613
Mean of dependent	94.975	94.975	95.457	95.457	97.834	97.834
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table 2 for notes on the specification.

Table 14: Impact of Competitive BB by Charter Status, Hepatitis B and Varicella

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Charter	5.132*	6.459**	-2.824	-.624
	(2.662)	(2.608)	(2.079)	(2.896)
Adjusted R2	.473	.500	.486	.472
N	449	449	1677	1677
Mean of dependent	90.186	90.186	89.623	89.623
Non-charter	-.507***	-.284**	-.561*	-.300
	(.150)	(.139)	(.302)	(.335)
Adjusted R2	.284	.326	.356	.380
N	15613	15613	13520	13520
Mean of dependent	96.737	96.737	97.073	97.073
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table 2 for notes on the specification.

Table 15: Impact of Competitive BB by FRM Eligibility, UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Lower 50% - Meals	-.674	-.250	.370***	.503***
	(.430)	(.422)	(.101)	(.112)
Adjusted R2	.419	.460	.428	.454
N	8841	8841	8841	8841
Mean of dependent	92.806	92.806	1.982	1.982
Upper 50% - Meals	-.376	.069	.298***	.436***
	(.420)	(.458)	(.098)	(.113)
Adjusted R2	.515	.516	.527	.530
N	8353	8353	8353	8353
Mean of dependent	92.122	92.122	.529	.529
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table 2 for notes on the specification.

Table 16: Impact of Competitive BB by FRM Eligibility, PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Lower 50% - Meals	.027	-.015	.276	-.237
	(.036)	(.046)	(.424)	(.404)
Adjusted R2	.105	.116	.419	.460
N	8841	8841	8841	8841
Mean of dependent	.229	.229	4.983	4.983
Upper 50% - Meals	.010	-.029	.068	-.476
	(.038)	(.047)	(.416)	(.444)
Adjusted R2	.062	.078	.512	.509
N	8353	8353	8353	8353
Mean of dependent	.077	.077	7.272	7.272
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table 2 for notes on the specification.

Table 17: Impact of Competitive BB by FRM Eligibility, DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Lower 50% - Meals	-.487	-.291	-.425	-.256	-.625***	-.376**
	(.339)	(.301)	(.308)	(.280)	(.225)	(.173)
Adjusted R2	.332	.391	.333	.384	.206	.292
N	8841	8841	8841	8841	8841	8841
Mean of dependent	94.935	94.935	95.264	95.264	97.249	97.249
Upper 50% - Meals	-.262	-.045	-.188	-.027	-.436**	-.275
	(.315)	(.324)	(.292)	(.304)	(.180)	(.172)
Adjusted R2	.467	.460	.486	.481	.350	.342
N	8353	8353	8353	8353	8353	8353
Mean of dependent	95.084	95.084	95.749	95.749	98.448	98.448
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table 2 for notes on the specification.

Table 18: Impact of Competitive BB by FRM Eligibility, Hepatitis B and Varicella

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Lower 50% - Meals	-.492*	-.269	-.565*	-.173
	(.251)	(.226)	(.291)	(.261)
Adjusted R2	.214	.288	.35	.76
N	8841	8841	7620	7620
Mean of dependent	96.392	96.392	96.525	96.525
Upper 50% - Meals	-.360	-.144	-.486	.041
	(.219)	(.235)	(.321)	(.324)
Adjusted R2	.383	.379	.383	.379
N	8353	8353	7167	7167
Mean of dependent	97.068	97.068	97.603	97.603
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table 2 for notes on the specification.

Table 19: Impact of Competitive BB by School API, UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Lower 50% - API	-1.194***	-.716*	.132	.134
	(.374)	(.430)	(.086)	(.109)
Adjusted R2	.524	.526	.407	.422
N	6621	6621	6621	6621
Mean of dependent	92.166	92.166	.502	.502
Upper 50% - API	-.498	.113	.299***	.408***
	(.466)	(.482)	(.104)	(.124)
Adjusted R2	.406	.445	.411	.438
N	8563	8563	8563	8563
Mean of dependent	92.917	92.917	1.862	1.862
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table 2 for notes on the specification.

Table 20: Impact of Competitive BB by School API, PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Lower 50% - API	-.013	-.073	1.075***	.655
	(.040)	(.050)	(.376)	(.433)
Adjusted R2	.055	.085	.524	.525
N	6621	6621	6621	6621
Mean of dependent	.08	.08	7.253	7.253
Upper 50% - API	.050	.022	.149	-.543
	(.041)	(.054)	(.459)	(.465)
Adjusted R2	.111	.126	.408	.445
N	8563	8563	8563	8563
Mean of dependent	.223	.223	4.998	4.998
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table 2 for notes on the specification.

Table 21: Impact of Competitive BB by School API, DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Lower 50% - API	-.511**	-.128	-.342	-.048	-.516**	-.344
	(.234)	(.266)	(.213)	(.244)	(.217)	(.247)
Adjusted R2	.474	.470	.483	.482	.304	.298
N	6621	6621	6621	6621	6621	6621
Mean of dependent	95.172	95.712	95.840	95.840	98.489	98.489
Upper 50% - API	-.328	.087	-.294	.059	-.610***	-.412**
	(.366)	(.342)	(.334)	(.319)	(.234)	(.183)
Adjusted R2	.319	.374	.325	.371	.211	.300
N	8563	8563	8563	8563	8563	8563
Mean of dependent	94.994	94.994	95.331	95.331	97.366	97.366
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table 2 for notes on the specification.

Table 22: Impact of Competitive BB by School API, Hepatitis B and Varicella

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Lower 50% - API	-.487***	-.083	-.496	-.190
	(.178)	(.199)	(.567)	(.645)
Adjusted R2	.351	.347	.332	.327
N	6621	6621	5682	5682
Mean of dependent	97.096	97.096	97.670	97.670
Upper 50% - API	-.424	-.192	-.582*	-.119
	(.261)	(.241)	(.313)	(.275)
Adjusted R2	.218	.293	.372	.390
N	8563	8563	7361	7361
Mean of dependent	96.526	96.526	96.630	96.630
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table 2 for notes on the specification.

Table 23: Google Trends Analysis: Vaccine-Related Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional
"Vaccine"	-.008	-.017*	-.002	.027
	(.027)	(.009)	(.003)	(.028)
Average disease rate	-.809***	.176***	.057**	.578***
	(.127)	(.058)	(.022)	(.117)
Interaction	.010***	.000	-.001	-.009***
	(.003)	(.001)	(.000)	(.002)
Adjusted R2	.906	.872	.488	.870
N	48	48	48	48
Mean of dependent	91.008	2.691	.198	6.102

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The observed sample is California counties from 2007-2008. The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions and % Conditional Exemptions. All regressions include year and county fixed effects, and the following county-level controls: natural log of kindergarten enrollment, population density, employment rate, family poverty rate, share population with college degree of higher, share hispanic, share black, share foreign-born, share population aged 0-4 and share aged 5-9. Standard errors are clustered at the county level.

Table 24: Google Trends Analysis: DTP Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) DTP
"Diphtheria"	.876***	-.432***	.077	-.512**	.967***
	(.221)	(.061)	(.071)	(.206)	(.161)
Tetanus rate	-63.948***	-18.563***	-7.671*	92.981***	-35.841**
	(17.352)	(5.261)	(3.952)	(13.398)	(13.828)
Interaction	-.598	.919***	-.123	-.207	-.924**
	(.570)	(.143)	(.117)	(.601)	(.406)
Adjusted R2	.949	.972	.641	.910	.958
N	38	38	38	38	38
Mean of dependent	91.008	2.691	.198	6.102	92.808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Diphtheria-Tetanus-Pertussis. See Table 1 for notes on the specification.

Table 25: Google Trends Analysis: DTP Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) DTP
"Whooping cough"	.146**	-.123***	-.018	.061	.086
	(.059)	(.042)	(.013)	(.096)	(.109)
Tetanus rate	-2.644	.108	-.961	-3.359	7.793
	(2.731)	(3.726)	(.970)	(12.094)	(8.319)
Interaction	-.569**	-.320*	.048	1.024**	-.783*
	(.209)	(.184)	(.063)	(.486)	(.428)
Adjusted R2	.644	.856	.645	.718	.784
N	44	44	44	44	44
Mean of dependent	91.008	2.691	.198	6.102	92.808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Diphtheria-Tetanus-Pertussis. See Table 1 for notes on the specification.

Table 26: Google Trends Analysis: DTP Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) DTP
"Tetanus"	.001	-.004	-.001	.002	.015
	(.033)	(.017)	(.004)	(.031)	(.033)
Tetanus rate	.710	-11.001	-1.414	6.326	23.317
	(32.097)	(12.129)	(3.000)	(29.745)	(27.502)
Interaction	-.917	.224	.083	.652	-1.592
	(1.042)	(.528)	(.112)	(.867)	(.991)
Adjusted R2	.781	.758	.555	.711	.761
N	46	46	46	46	46
Mean of dependent	91.008	2.691	.198	6.102	92.808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Diphtheria-Tetanus-Pertussis. See Table 1 for notes on the specification.

Table 27: Google Trends Analysis: DTP Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) DTP
"DTP vaccine"	-.058***	.038	-.005*	.066***	-.169***
	(.020)	(.025)	(.003)	(.013)	(.050)
Tetanus rate	-11.592***	6.349*	1.033*	5.197**	-21.294***
	(2.858)	(3.425)	(.510)	(2.404)	(6.457)
Interaction	-.574	-.992***	-.195**	.539	.651
	(.648)	(.312)	(.100)	(.397)	(1.145)
Adjusted R2	.726	.863	.275	.591	.942
N	42	42	42	42	42
Mean of dependent	91.008	2.691	.198	6.102	92.808

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Diphtheria-Tetanus-Pertussis. See Table 1 for notes on the specification.

Table 28: Google Trends Analysis: Polio Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Polio
"Poliomyelitis"	.025	.013	.006	-.056**	.050
	(.026)	(.016)	(.004)	(.024)	(.034)
Average disease rate	-.426***	.143*	-.038*	.212	-.195
	(.100)	(.083)	(.021)	(.154)	(.143)
Interaction	.013***	.004*	-.000	-.011***	-.007***
	(.004)	(.002)	(.001)	(.004)	(.002)
Adjusted R2	.864	.801	.671	.719	.712
N	46	46	46	46	46
Mean of dependent	91.008	2.691	.198	6.102	93.176

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against polio. See Table 1 for notes on the specification.

Table 29: Google Trends Analysis: Polio Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Polio
"Polio vaccine"	-.054	-.058	-.015	.113*	-.007
	(.059)	(.035)	(.009)	(.065)	(.047)
Average disease rate	-.374***	.049	-.057**	.351*	-.187
	(.128)	(.104)	(.026)	(.195)	(.187)
Interaction	.013**	.007*	.001	-.023***	.005
	(.006)	(.004)	(.001)	(.007)	(.004)
Adjusted R2	.751	.848	.613	.662	.722
N	44	44	44	44	44
Mean of dependent	91.008	2.691	.198	6.102	93.176

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against polio. See Table 1 for notes on the specification.

Table 30: Google Trends Analysis: MMR Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) MMR
"Measles"	.062***	-.018**	-.007**	-.047***	.028***
	(.019)	(.009)	(.003)	(.017)	(.008)
Measles rate	3.635	3.011	-.929	-7.187	-13.144**
	(11.775)	(5.954)	(1.514)	(9.433)	(6.233)
Interaction	-.723	-.081	.080	.556	.774
	(.859)	(.454)	(.082)	(.719)	(.487)
Adjusted R2	.873	.828	.597	.718	.880
N	42	42	42	42	42
Mean of dependent	91.008	2.691	.198	6.102	96.125

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Measles-Mumps-Rubella. See Table 1 for notes on the specification.

Table 31: Google Trends Analysis: MMR Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) MMR
"Mumps"	-.007	-.004	-.003	.007	.014
	(.014)	(.007)	(.003)	(.013)	(.009)
Measles rate	-26.401	9.223	4.054**	3.695	-6.980
	(18.900)	(6.711)	(1.590)	(16.742)	(10.124)
Interaction	.641	-.317	-.277**	-.102	.410
	(.123)	(.334)	(.112)	(1.050)	(.392)
Adjusted R2	.846	.825	.690	.679	.808
N	43	43	43	43	43
Mean of dependent	91.008	2.691	.198	6.102	96.125

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Measles-Mumps-Rubella. See Table 1 for notes on the specification.

Table 32: Google Trends Analysis: MMR Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) MMR
"Rubella"	-.015	-.008	-.010***	.031	.028*
	(.020)	(.012)	(.003)	(.023)	(.015)
Measles rate	2.628**	.500	-.446***	-2.586***	.538
	(.984)	(.699)	(.143)	(.807)	(.569)
Interaction	-.088	-.001	.018	.022	.074
	(.264)	(.090)	(.023)	(.257)	(.104)
Adjusted R2	.813	.810	.756	.702	.814
N	40	40	40	40	40
Mean of dependent	91.008	2.691	.198	6.102	96.125

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Measles-Mumps-Rubella. See Table 1 for notes on the specification.

Table 33: Google Trends Analysis: MMR Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) MMR
"MMR vaccine"	-.094*	-.018	.006	.023	-.134**
	(.054)	(.045)	(.019)	(.090)	(.055)
Measles rate	4.643	-.618	1.018	-10.103	6.030
	(8.293)	(4.006)	(1.662)	(6.666)	(8.896)
Interaction	-.035***	-.001	.002	.026**	.001
	(.011)	(.006)	(.002)	(.011)	(.013)
Adjusted R2	.658	.808	.506	.702	.536
N	44	44	44	44	44
Mean of dependent	91.008	2.691	.198	6.102	96.125

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Measles-Mumps-Rubella. See Table 1 for notes on the specification.

Table 34: Google Trends Analysis: Hepatitis B Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Hepatitis B
"Hepatitis B"	-.034	-.015	-.021	.021	.050
	(.085)	(.029)	(.014)	(.079)	(.033)
Hepatitis B rate	-.210**	.007	-.010	.120	-.068*
	(.101)	(.031)	(.016)	(.089)	(.036)
Interaction	.009	.003	.001	-.006	.001
	(.007)	(.002)	(.001)	(.005)	(.002)
Adjusted R2	.907	.829	.548	.855	.875
N	48	48	48	48	48
Mean of dependent	91.008	2.691	.198	6.102	95.518

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Hepatitis B. See Table 1 for notes on the specification.

Table 35: Google Trends Analysis: Hepatitis B Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Hepatitis B
"Hepatitis B vaccine"	.073*	.071***	-.007**	-.103**	-.076***
	(.039)	(.015)	(.003)	(.047)	(.016)
Hepatitis B rate	.006	.080***	-.021***	-.048	-.127***
	(.017)	(.014)	(.003)	(.037)	(.017)
Interaction	-.004***	-.000	.000***	.003***	-.000
	(.001)	(.001)	(.000)	(.001)	(.001)
Adjusted R2	.922	.965	.944	.755	.952
N	40	40	40	40	40
Mean of dependent	91.008	2.691	.198	6.102	95.518

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against Hepatitis B. See Table 1 for notes on the specification.

Table 36: Google Trends Analysis: Varicella Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Varicella
"Chickenpox"	.020	.003	.002	-.040	-.005
	(.056)	(.016)	(.007)	(.042)	(.031)
Average disease rate	-.575***	.133*	.021	.376*	-.363***
	(.198)	(.068)	(.035)	(.202)	(.088)
Interaction	.009	.002	.004	-.008	.001
	(.011)	(.003)	(.001)	(.009)	(.003)
Adjusted R2	.833	.809	.394	.774	.855
N	48	48	48	48	48
Mean of dependent	91.008	2.691	.198	6.102	95.798

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against varicella. See Table 1 for notes on the specification.

Table 37: Google Trends Analysis: Varicella Search Terms

	(1) UTD	(2) PBE	(3) PME	(4) Conditional	(5) Varicella
"Varicella vaccine"	-.134	-.099***	-.021	.256***	.027
	(.110)	(.033)	(.017)	(.082)	(.047)
Average disease rate	-.622**	-.256	-.022	1.153***	.025
	(.251)	(.151)	(.039)	(.341)	(.197)
Interaction	.028 *	.023***	-.001	-.061***	-.014
	(.015)	(.007)	(.002)	(.016)	(.009)
Adjusted R2	.784	.893	.720	.813	.882
N	41	41	41	41	41
Mean of dependent	91.008	2.691	.198	6.102	95.798

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are % kindergarten students who are UTD on all vaccinations, % Personal Belief Exemptions, % Personal Medical Exemptions, % Conditional Exemptions and % vaccinated against varicella. See Table 1 for notes on the specification.

Figure A1: Diffusion of BB across California ZCTAs, June 2000-June 2008.

Figure A2: Diffusion of Internet Service Providers across California ZCTAs, June 2000-June 2008.

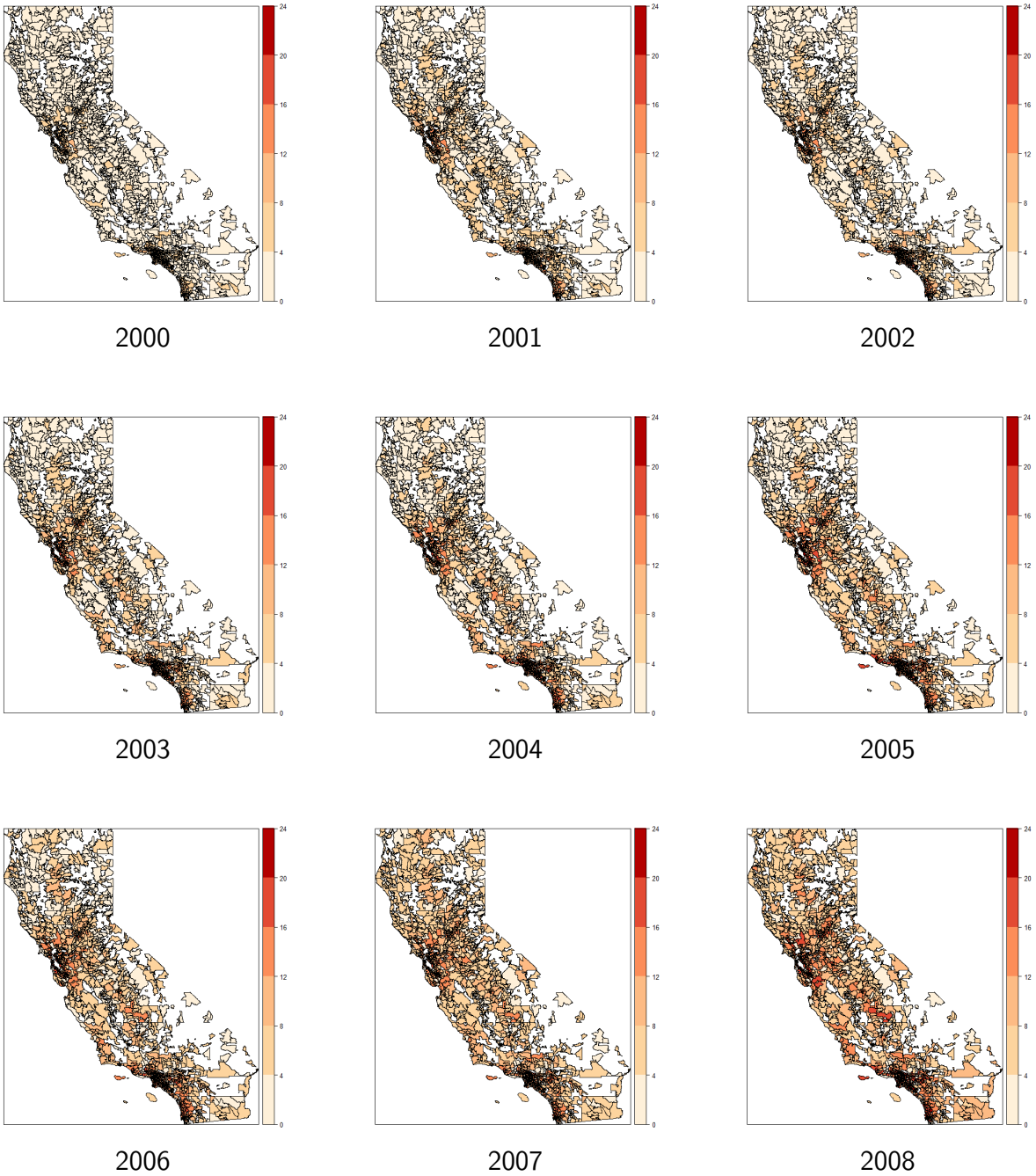
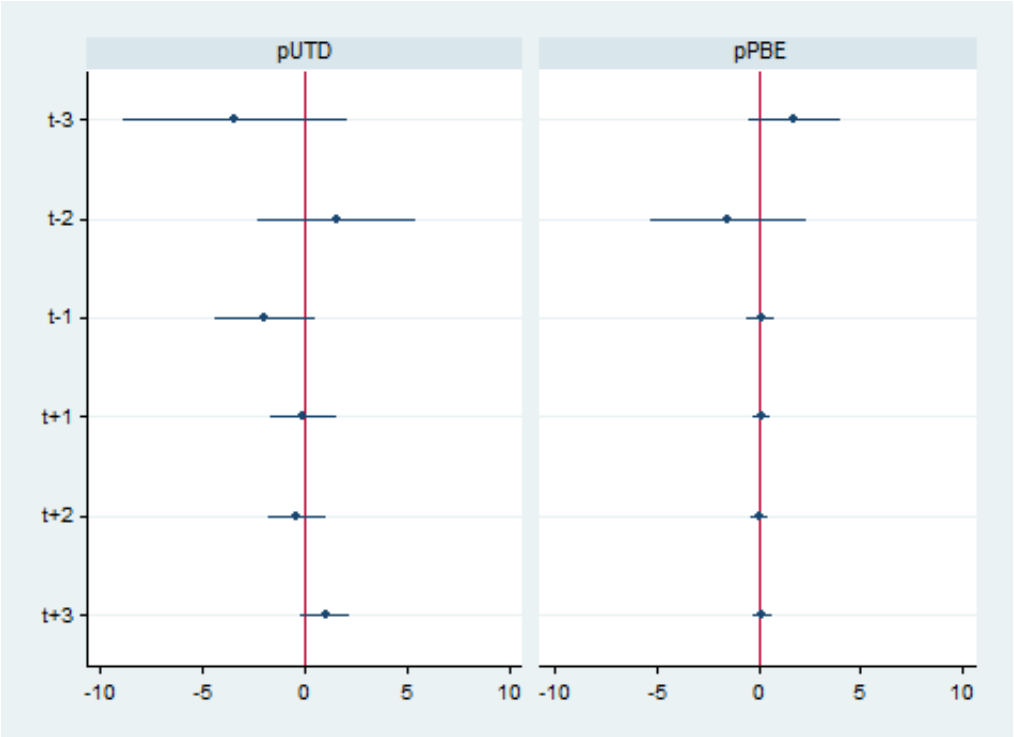
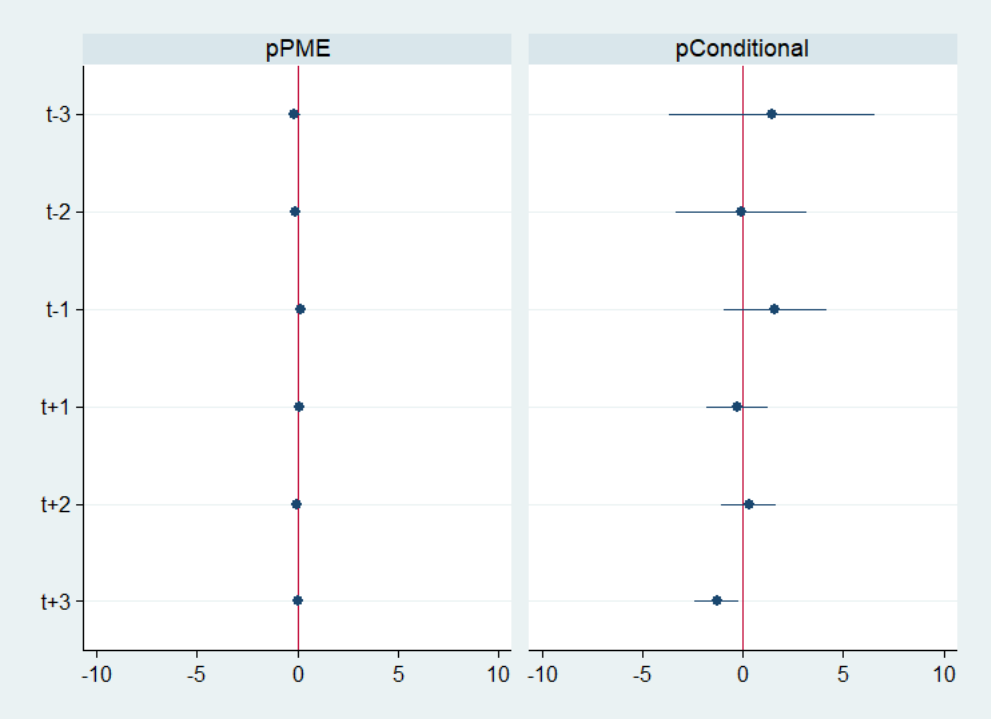


Figure A3: Private schools
Effect of competitive broadband on immunization outcomes over time



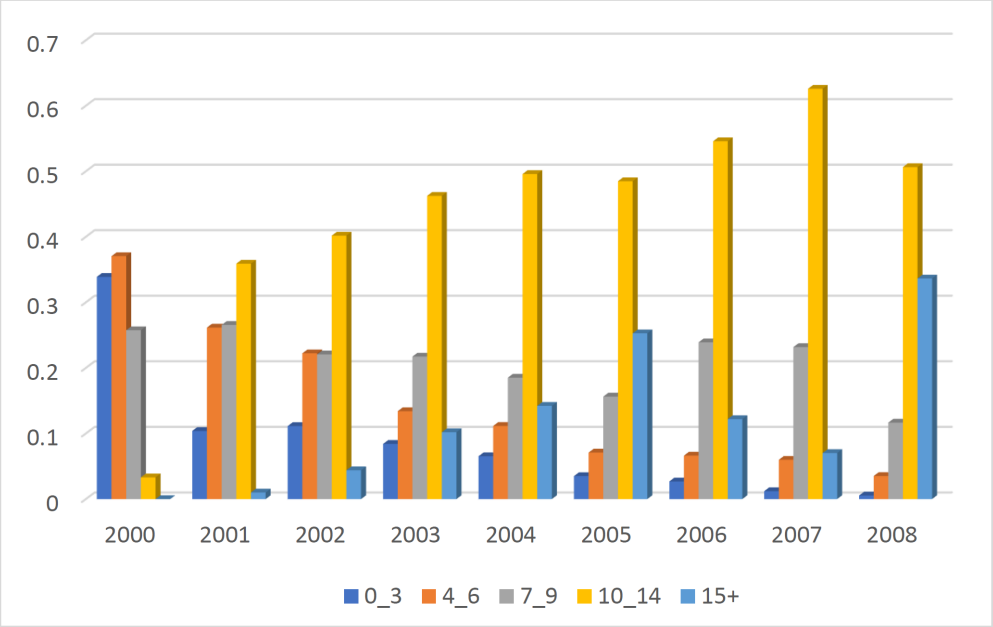
Notes: The left figure plots coefficient estimates for the impact of "competitive broadband" time dummies on percent Up-To-Date. The right figure plots estimates for the impact of the dummies on Personal Belief Exemptions. Time 't' is defined as the first year in which a zipcode obtained access to four or more providers.

Figure A4: Private schools
Effect of competitive broadband on immunization outcomes over time



Notes: The left figure plots coefficient estimates for the impact of "competitive broadband" time dummies on Personal Medical Exemptions. The right figure plots estimates for the impact of the dummies on conditional exemptions. Time 't' is defined as the first year in which a zipcode obtained access to four or more providers.

Figure A5: Distribution of public schools across provider categories, 2000-2008



Notes: Distribution of California public schools across five different categories of broadband provision: 0-3 providers; 4-6; 7-9; 10-14 and 15+.

Table A1: Effect of BB on Public School Outcomes: UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
BB	-.968	-.914	.290*	.364**
	(.740)	(.925)	(.149)	(.180)
Charter=1 × year	-.329	-.708*	.081	.119*
	(.276)	(.393)	(.053)	(.066)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.474	.493	.488	.514
N	16071	16071	16071	16071
Mean of dependent	91.805	91.805	1.699	1.699

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. The school controls are beginning-of-school-year API, growth in API, the School Characteristic Index, pupil turnover, average K-12 class size, proportion of teachers with emergency credentials, proportion of students eligible for free or reduced meals, average educational level of parents, share of English-learning or English-proficient students, and the kindergarten population shares for Native American, Asian, Pacific Islander, Filipino, Hispanic and black. Standard errors are clustered at the zipcode level.

Table A2: Effect of BB on Public School Outcomes: PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
BB	-.014	-.103	.691	.653
	(.085)	(.081)	(.908)	(1.111)
Charter=1 × year	-.013	-.018	.261	.607
	(.012)	(.015)	(.305)	(.445)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.102	.117	.476	.494
N	16071	16071	16071	16071
Mean of dependent	0.172	0.172	6.325	6.325

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table A1 for notes on the specification.

Table A3: Effect of Providers on Public School Outcomes: UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Number of providers	-.071**	-.084**	.009	.010
	(.034)	(.037)	(.008)	(.008)
Charter=1 × year	-.327	-.709*	.080	.119*
	(.275)	(.391)	(.053)	(.066)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.474	.493	.488	.514
N	16071	16071	16071	16071
Mean of dependent	91.805	91.805	1.699	1.699

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table A1 for notes on the specification.

Table A4: Effect of Providers on Public School Outcomes: PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Number of providers	-.000	-.001	.062*	.075**
	(.003)	(.003)	(.033)	(.036)
Charter=1 × year	-.013	-.018	.260	.608
	(.012)	(.015)	(.305)	(.444)
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.102	.117	.476	.494
N	16071	16071	16071	16071
Mean of dependent	0.172	0.172	6.325	6.325

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table A1 for notes on the specification.

Table A5: Effect of Competitive BB on Private School Outcomes: UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Competitive BB	1.375	1.324	.250	.540*
	(.849)	(1.031)	(.276)	(.293)
Religion=1 × year	.048	.090	-.175***	-.163***
	(.184)	(.199)	(.055)	(.055)
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.440	.441	.676	.679
N	11061	11061	11063	11063
Mean of dependent	88.679	88.679	2.574	2.574

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table A1 for notes on the specification.

Table A6: Effect of Competitive BB on Private School Outcomes: PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Competitive BB	-.032	.100	-1.617*	-1.890*
	(.059)	(.071)	(.842)	(.998)
Religion=1 × year	-.012	-.013	.145	.090
	(.013)	(.016)	(.181)	(.197)
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.067	.093	.362	.361
N	11063	11063	11063	11063
Mean of dependent	.188	.188	8.569	8.569

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table A1 for notes on the specification.

Table A7: Effect of Competitive BB on Private School DTP, Polio and MMR Coverage

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Competitive BB	.337	.295	.439	.444	1.383*	1.357
	(.724)	(.956)	(.729)	(.924)	(.839)	(1.013)
Religion=1 × year	.185	.204	.185	.179	.241***	.232***
	(.151)	(.163)	(.151)	(.162)	(.081)	(.088)
District controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X
Adjusted R2	.426	.428	.438	.441	.532	.538
N	11063	11063	11063	11063	11063	11063
Mean of dependent	91.800	91.800	91.853	91.853	95.895	95.895

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table A1 for notes on the specification.

Table A8: Effect of Competitive BB on Private School Hepatitis B and Varicella Coverage

	(1) Hepatitis B	(2) Hepatitis B	(3) Varicella	(4) Varicella
Competitive BB	-.104	-.034	.144	.430
	(.501)	(.572)	(1.334)	(1.521)
Religion=1 × year	.243***	.235**	.125	.177
	(.087)	(.093)	(.130)	(.115)
District controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X
Adjusted R2	.634	.651	.657	.676
N	11063	11063	9134	9134
Mean of dependent	95.361	95.361	94.918	94.918

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table A1 for notes on the specification.

Table A9: Effect of Future Competitive BB on Private School Outcomes

	(1) UTD	(2) PBE	(3) PME	(4) Conditional
Competitive BB	1.137	.202	-.054	-1.288
	(.858)	(.265)	(.061)	(.859)
Future competitive BB	-1.425	.142	.171	.784
	(1.589)	(.581)	(.108)	(1.377)
District controls	X	X	X	X
School, year F.E.	X	X	X	X
Adjusted R2	.439	.676	.068	.361
N	11061	11063	11063	11063
Mean of dependent	88.679	2.574	.188	8.569

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in column (1) is the full coverage rate; in (2), percent Personal Belief Exemptions; in (3), Personal Medical Exemptions; and in (4), Conditional Exemptions. See Table A1 for notes on the specification.

Table A10: Length of Exposure to Competitive BB, Impact on UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Months Competitive BB	.124***	.120	-.005	-.041**
	(.037)	(.086)	(.010)	(.017)
Log Months Squared	-.750***	-.732***	.063	.155**
	(.214)	(.275)	(.058)	(.062)
Adjusted R2	.474	.493	.488	.514
N	16071	16071	16071	16071
Mean of dependent	91.805	91.805	1.699	1.699
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table A1 for notes on the specification.

Table A11: Length of Exposure to Competitive BB, Impact on PME and Conditional

	(1) PME	(2) PME	(3) Conditional	(4) Conditional
Months Competitive BB	-.005	-.003	-.114***	-.076
	(.004)	(.006)	(.035)	(.086)
Log Months Squared	.026	.013	.661***	.564**
	(.020)	(.021)	(.207)	(.266)
Adjusted R2	.102	.117	.476	.494
N	16071	16071	16071	16071
Mean of dependent	0.172	0.172	6.325	6.325
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table A1 for notes on the specification.

Table A12: Length of Exposure to Competitive BB, Impact on DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Months Competitive BB	-.010	-.150**	-.019	-.144**	.016	-.016
	(.026)	(.060)	(.023)	(.056)	(.024)	(.054)
Log Months Squared	.048	.316	.092	.358*	-.087	-.094
	(.153)	(.205)	(.146)	(.195)	(.134)	(.167)
Adjusted R2	.409	.437	.418	.443	.283	.331
N	16071	16071	16071	16071	16071	16071
Mean of dependent	94.287	94.287	94.681	94.681	97.289	97.289
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table A1 for notes on the specification.

Table A13: Length of Exposure to Competitive BB, Impact on Hepatitis B and Varicella

	(1)	(2)	(3)	(4)
	Hepatitis B	Hepatitis B	Varicella	Varicella
Months Competitive BB	.027	.166***	.129***	.086**
	(.018)	(.040)	(.025)	(.034)
Log Months Squared	-.001	.272	-.242*	.081
	(.125)	(.173)	(.139)	(.206)
Adjusted R2	.297	.336	.373	.432
N	16071	16071	13931	13931
Mean of dependent	96.286	96.286	96.581	96.581
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table A1 for notes on the specification.

Table A14: Impact of Competitive BB by Number of Providers, UTD and PBE

	(1) UTD	(2) UTD	(3) PBE	(4) PBE
Competitive BB, 3 cat. =1	-.804***	-.379	.219***	.309***
	(.304)	(.312)	(.070)	(.082)
Competitive BB, 3 cat. =2	-.902***	-.525	.225***	.336***
	(.338)	(.392)	(.076)	(.090)
Adjusted R2	.475	.493	.489	.515
Competitive BB, 4 cat. =1	-.848***	-.452	.226***	.318***
	(.307)	(.317)	(.072)	(.083)
Competitive BB, 4 cat. =2	-.925***	-.573	.229***	.341***
	(.342)	(.397)	(.076)	(.091)
Competitive BB, 4 cat. =3	-1.074**	-.826*	.254***	.370***
	(.422)	(.484)	(.090)	(.106)
Adjusted R2	.475	.493	.489	.515
Competitive BB, 5 cat. =1	-.908***	-.486	.227***	.316***
	(.312)	(.317)	(.071)	(.083)
Competitive BB, 5 cat. =2	-1.045***	-.651	.230***	.339***
	(.347)	(.399)	(.077)	(.091)
Competitive BB, 5 cat. =3	-1.201***	-.916*	.255***	.367***
	(.427)	(.487)	(.091)	(.106)
Competitive BB, 5 cat. =4	-1.691***	-1.468***	.259**	.346***
	(.516)	(.563)	(.112)	(.123)
Adjusted R2	.475	.493	.488	.515
N	16071	16071	16071	16071
Mean of dependent	91.805	91.805	1.699	1.699
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students at a school who have received all required immunizations. In columns (3) and (4), it is percent of students with Personal Belief Exemptions. See Table A1 for notes on the specification.

Table A15: Impact of Competitive BB by Number of Providers, PME and Conditional

	(1)	(2)	(3)	(4)
	PME	PME	Conditional	Conditional
Competitive BB, 3 cat. =1	.014	-.022	.572*	.092
	(.029)	(.036)	(.301)	(.307)
Competitive BB, 3 cat. =2	.022	-.035	.655*	.224
	(.032)	(.039)	(.336)	(.389)
Adjusted R2	.102	.117	.476	.494
Competitive BB, 4 cat. =1	.021	-.016	.602**	.150
	(.030)	(.036)	(.303)	(.311)
Competitive BB, 4 cat. =2	.026	-.031	.671**	.262
	(.033)	(.040)	(.340)	(.394)
Competitive BB, 4 cat. =3	.050	-.007	.771*	.463
	(.041)	(.046)	(.416)	(.479)
Adjusted R2	.102	.117	.476	.494
Competitive BB, 5 cat. =1	.016	-.017	.665**	.186
	(.030)	(.036)	(.307)	(.310)
Competitive BB, 5 cat. =2	.016	-.034	.799**	.346
	(.033)	(.040)	(.345)	(.395)
Competitive BB, 5 cat. =3	.040	-.011	.906**	.560
	(.041)	(.046)	(.421)	(.482)
Competitive BB, 5 cat. =4	.002	-.031	1.430***	1.153**
	(.048)	(.053)	(.506)	(.557)
Adjusted R2	.102	.117	.476	.495
N	16071	16071	16071	16071
Mean of dependent	.172	.172	6.325	6.325
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is percent of public kindergarten students with Personal Medical Exemptions. In columns (3) and (4), it is percent of students with Conditional Exemptions. See Table A1 for notes on the specification.

Table A16: Impact of Competitive BB by Number of Providers, DTP, Polio and MMR

	(1) DTP	(2) DTP	(3) Polio	(4) Polio	(5) MMR	(6) MMR
Competitive BB, 3 cat. =1	-.432*	-.097	-.336	-.041	-.470***	-.255**
	(.229)	(.218)	(.209)	(.203)	(.149)	(.120)
Competitive BB, 3 cat. =2	-.278	.009	-.174	.064	-.386**	-.134
	(.243)	(.280)	(.224)	(.266)	(.161)	(.170)
Adjusted R2	.409	.436	.419	.443	.283	.331
Competitive BB, 4 cat. =1	-.406*	-.099	-.316	-.040	-.457***	-.231**
	(.226)	(.218)	(.205)	(.203)	(.140)	(.117)
Competitive BB, 4 cat. =2	-.265	.007	-.164	.065	-.379**	-.118
	(.244)	(.283)	(.224)	(.268)	(.160)	(.174)
Competitive BB, 4 cat. =3	-.180	-.002	-.097	.068	-.333*	-.033
	(.294)	(.349)	(.263)	(.323)	(.191)	(.239)
Adjusted R2	.409	.436	.419	.443	.283	.331
Competitive BB, 5 cat. =1	-.444*	-.120	-.349*	-.060	-.465***	-.236**
	(.229)	(.219)	(.208)	(.204)	(.141)	(.117)
Competitive BB, 5 cat. =2	-.341	-.040	-.232	.019	-.395**	-.129
	(.244)	(.285)	(.225)	(.270)	(.153)	(.174)
Competitive BB, 5 cat. =3	-.260	-.056	-.168	.015	-.351*	-.046
	(.294)	(.352)	(.266)	(.327)	(.186)	(.240)
Competitive BB, 5 cat. =4	-.568	-.392	-.444	-.308	-.416**	-.126
	(.354)	(.411)	(.329)	(.389)	(.203)	(.271)
Adjusted R2	.410	.437	.419	.443	.283	.331
N	16071	16071	16071	16071	16071	16071
Mean of dependent	94.287	94.287	94.681	94.681	97.289	97.289
School controls	X	X	X	X	X	X
School, year F.E.	X	X	X	X	X	X
School district trend		X		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten DTP coverage rate; in (3) and (4), the polio coverage rate; and in (5) and (6), the MMR coverage rate. See Table A1 for notes on the specification.

Table A17: Impact of Competitive BB by Number of Providers, Hepatitis B and Varicella

	(1)	(2)	(3)	(4)
	Hepatitis B	Hepatitis B	Varicella	Varicella
Competitive BB, 3 cat. =1	-.460***	-.220	-.430	-.254
	(.168)	(.150)	(.308)	(.344)
Competitive BB, 3 cat. =2	-.406**	-.104	-.684**	-.352
	(.175)	(.202)	(.319)	(.340)
Adjusted R2	.297	.336	.374	.392
Competitive BB, 4 cat. =1	-.430***	-.188	-.541*	-.307
	(.149)	(.151)	(.304)	(.336)
Competitive BB, 4 cat. =2	-.333*	-.075	-.475	-.048
	(.182)	(.221)	(.347)	(.371)
Competitive BB, 4 cat. =3	-.391*	-.170	-.632*	-.278
	(.236)	(.260)	(.367)	(.384)
Adjusted R2	.297	.336	.374	.392
Competitive BB, 5 cat. =1	-.438***	-.195	-.431	-.268
	(.162)	(.151)	(.309)	(.346)
Competitive BB, 5 cat. =2	-.404**	-.097	-.718**	-.383
	(.171)	(.208)	(.319)	(.341)
Competitive BB, 5 cat. =3	-.311	.009	-.626*	-.123
	(.211)	(.281)	(.353)	(.374)
Competitive BB, 5 cat. =4	-.364	-.077	-.809**	-.356
	(.246)	(.319)	(.375)	(.389)
Adjusted R2	.297	.336	.374	.392
N	16071	16071	13931	13931
Mean of dependent	96.286	96.286	96.581	96.581
School controls	X	X	X	X
School, year F.E.	X	X	X	X
School district trend		X		X

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable in columns (1) and (2) is the public kindergarten Hepatitis B coverage rate; in (3) and (4), the varicella coverage rate. See Table A1 for notes on the specification.

Chapter 2:
Broadband Internet and Immunization Coverage in U.S. Counties

Carrie Huffaker
Bocconi University

Advisor: Dr. Diego Ubfal

April 16, 2019

1 Introduction

In many local communities across the U.S., coverage rates for different vaccine-preventable diseases have fallen below the herd immunity threshold, leading to new outbreaks. Although most VPDs are controlled, outbreaks not only continue to occur but are worsening over time (Clemmons et al., 2017). Many new parents lack first-hand experience of the diseases that vaccines protect against. Others are concerned by the increase in the number of shots required to attend school or daycare. Yet the diffusion of internet has also played a role in shaping public perceptions of immunization. Since internet access spread to the American public in the early- to mid-1990s, a profusion of vaccine-critical websites have arisen and acquired hold of a vast audience. These websites challenge medical authority, and encourage parents to decide for themselves whether or not to vaccinate their child. There is substantial evidence that parents are turning to the internet for health information and news. Is it possible that there is a relationship between falling local vaccination rates and exposure to the internet? Or has the internet made parents more informed and encouraged vaccine uptake?

In this paper, I explore the relationship between broadband internet access and vaccination rates in 257 U.S. counties from 2000 to 2008. I employ an IV estimation strategy, instrumenting number of Internet Service Providers present in each county with a composite measure capturing the degree to which state-level policies permit access of ISPs to local communities. I find evidence for modest declines in the percentage of toddlers immunized against Haemophilus influenzae (Hib) and in the full coverage rate. The expansion of broadband access led to a decline in Hib coverage of 0.67 points, and a decline in the percentage of fully-immunized toddlers of 0.07 points. In the analysis of heterogeneous effects, I also find evidence of a decline in varicella uptake among those counties with higher family poverty and a greater share of foreign-born evidence.

My paper relates to three strands of literature. The first is literature on impact of new media upon women's health, fertility and other social outcomes. Television has been shown to influence attitudes towards gender relations and childbearing across the developing world, by presenting alternative and idealized lifestyles for women that strongly contrast with those permissible under traditional social norms (La Ferrara et al., 2008; Jensen and Oster, 2009;

La Ferrara, 2016). In the United States, the MTV program *16 and Pregnant* has reduced pregnancy among teenage viewers (Kearney and Levine, 2014).

The second strand of literature addresses how the diffusion of new technologies influences society, and political outcomes in particular. Stromberg (2004) found that during the Great Depression, localities with a higher proportion of radio listeners were allocated more relief funds by politicians. With the introduction of television, the opposite has occurred: Viewers have become less informed, and vote less (Gentzkow, 2006). When news reports provide voters with less information about local politics, they elect local politics elect politicians who are less responsive to constituent needs (Snyder and Stromberg, 2010). Finally, it has been shown that the introduction of Fox News shifted large numbers of voters to the Republican party (DellaVigna and Kaplan, 2007).

The third strand focuses upon the specific technology of broadband internet and how it influences political outcomes. Lelkes (2015) show that the spread of broadband access has increased partisan hostilities in the U.S.. Research focusing on voter turnout in Germany and the U.K. has found a negative impact of broadband (Falck et al., 2014; Gavazza et al., 2016). Larcinese and Miner (2017) show the opposite to hold true in the U.S, while Heblich (2016) also finds a positive effect upon German turnout in more recent years.

My paper represents one of the first attempts to expand the broadband literature and focus upon a different sort of outcome. While the political repercussions of expanded broadband access have been much discussed, this technological change has also revolutionized the way in which consumers research science and health topics and update their beliefs about specific interventions such as immunization. My results suggest that access to the wealth of online health information can result in information overload and lead families to engage in behaviors that put their children's health at risk.

The paper is structured in the following way. In section 2, I discuss the factors that shape demand for immunizations, and how the new information economy- and in particular, the online antivaccine movement- has transformed this demand. In section 3 I introduce the data and empirical strategy. Estimation results are presented in the following section. I demonstrate the exogeneity of the instrument through a number of tests for selection in section 5. The paper concludes with a brief discussion.

2 Immunization Uptake and Broadband Internet Access

Why should we expect a relationship to exist between the diffusion of broadband internet and vaccine uptake? To answer this question, I begin by discussing supply and demand for immunizations in the United States, and how the interaction of these factors shape herd immunity of local populations. Next, I address how the diffusion of internet access, and of broadband in particular, has radically transformed the way in which Americans seek out medical information and make decisions about their own health. The downside of this is that the enormous quantity of online content can generate information overload, a situation in which consumers are unable to distinguish between true and false information. I conclude with a brief review of the wide-ranging literature on how the online antivaccine movement, in particular, has negatively influenced beliefs about and preferences for immunization.

2.1 The Market for Immunizations

Until the passing of the Vaccination Assistance Act by Congress in 1962, many parents were unable to vaccinate their children. The private practices or local health departments responsible for administering the vaccines often lacked the necessary infrastructure to do so, and in some cases families were forced to pay for them out of pocket. Following the passage of the Act and of later legislation, a nationwide immunization program was created to fund state and local health departments and, under the guidance of Public Health Advisers from the Centers for Disease Control (CDC), work towards the goal of universal immunization (Hinman et al., 2011). This public management of local immunization programs means that unlike in developing countries, vaccine uptake in the U.S. is not shaped by supply-side limitations such as shortages of crucial vaccines; variation is an entirely demand-side phenomenon, driven by the preferences of parents.

Underlying mass vaccination campaigns is the concept of *herd immunity*, defined by the CDC as "a situation in which a sufficient proportion of a population is immune to an infectious disease (through vaccination and/or prior illness) to make its spread from person to person

unlikely" (CDC, 2015). Herd immunity depends upon a number of factors, including the effectiveness of the vaccine against disease transmission; the *reproduction number* R_0 (number of secondary cases generated by a typical infectious individual); and the degree to which vaccinated individuals are distributed randomly throughout the population (Fine et al., 2011). Assuming a completely effective vaccine and a randomly mixing, homogeneous population, the critical vaccination level q_c necessary to achieve herd immunity can be calculated as:

$q_c = 1 - \frac{1}{R_0}$

The herd immunity thresholds for six different vaccine-preventable diseases are presented in Table 1. More infectious diseases, likes measles and pertussis, have much higher thresholds.

If a vaccine is shown not to be effective; if unvaccinated individuals are not randomly distributed, but concentrated in specific communities, then herd immunity will fail and populations will become susceptible to disease. While advances in medical knowledge have resulted in the creation of safer, more effective vaccines over time, the concentration of vaccine refusals in specific communities is more difficult to identify and address.

Local-level disparities in immunization rates across U.S. counties reflect two important sets of factors. First are the constraints placed upon families by state law. Vaccination requirements are not established at the national level. The CDC-recommended schedule is a guideline for parents and health care professionals, not a legal requirement. The responsibility for establishing the number of vaccines required of children in order to be admitted to childcare, kindergarten and 7th grade lies with state legislatures. For this reason, requirements vary widely nationwide. Moreover, states may permit parents to opt of requirements by allowing for a variety of exemptions. Medical exemptions are legal in all 50 states; survivors of childhood cancer and other individuals with compromised immune systems cannot be safely vaccinated, for example. Several religious groups, including Christian Scientists and the Amish, forbid vaccination. In 2016, every state save West Virginia and California allowed for this type of objection to be protected by religious exemption. Personal belief, or philosophical, exemptions were allowed in 20 states (NCSL, 2016). Different rules apply during outbreaks of vaccine-preventable diseases; in general, students enjoying such exemptions are excluded from schools for the duration of the outbreak.

A second set of factors shaping idiosyncrasies in local uptake regards the information available to families in making the vaccination decision, and how easily this information can be accessed. As discussed in the next section, with the arrival of broadband internet a new wealth of information in new forms has become available. This information varies widely in quality, however, and the inability of the average consumer to distinguish between what is true and false online can create information overload and result in the formation of beliefs about topics like immunization that have no foundation in science.

2.2 The New Online Information Economy

By the mid-1990s, the advent of the WorldWideWeb, Web browsers and efficient search engines had enabled the American public to access the internet at far higher rates than every before. Privatization of the internet backbone led to the growth of many new ISPs to satisfy the soaring demand. This expansion in usership was eventually followed by an expansion in who was able to create and publish new content on the internet. Blog-publishing tools such as Blogger had appeared by the late 1990s, enabling the non-technical public to build their own websites for the first time. New forms of social media followed, including Facebook (2004), Youtube (2005) and Twitter (2006). The explosion of user-generated content, together with the near-universal online presence of "real-world" organizations and institutions, has produced a democratization of information without precedent. The clear downside to this democratization is the enormous amount of false and misleading information published online. Professional medical advice has been present on the internet since the very beginning- two key examples are WebMD, online since 1996, and Mayo Clinic, online since 1998-but these trustworthy sources can be drowned out by other types of websites with a more pleasing presentation.

Studies carried out by the Pew Research Center reveal the increasing dependence of Americans upon the internet as a source of information. By 2000, 55% of Americans with internet access had looked up health information online; within two years, this figure had risen to 80% (Fox and Rainie, 2000; Fox, 2005). By 2006, a large proportion of the population had also come to rely upon the internet as their primary source of science news and information (Horrigan, 2006). The science study is interesting for two reasons. First, it reveals that individuals with home broadband connections use the internet differently. They are far more likely than

dial-up users to obtain most of their science information from the internet (34% vs. 22%). This result also holds in the context of health: 88% of broadband users researched health information online in 2009, versus 72% of dial-up users (Fox and Jones, 2009). Broadband access is associated with a substitution away from other sources of information like television, magazines and newspapers.

Second, young adults age 18-29 (followed closely by adults age 30-49), are more likely than any other demographic to look up health and science topics on the internet, and to use it as their primary source of information. This reliance has increased over time. The same result holds true for parents of children under age 18. The implications are clear: many young parents learn about health and science online. In many cases- about four in ten- web resources are their principal learning reference. It is possible that these habits may extend to exploiting online resources when deciding whether or not to vaccinate children.

The main explanation for this reliance on the internet, given by 7 of 10 online seekers, is convenience (Horrigan). The internet has drastically lowered search costs, creating a world in which information is cheap and plentiful. Yet online research is characterized by one major drawback: a wealth of unsubstantiated and false "facts" compete with good sources in the online marketplace. Three-quarters of online seekers do not consistently check the date and source of online health information (Fox and Jones). One in four feel overwhelmed by the sheer amount of available health information, while one in five feel confused by what they read online. Similarly, a Pew report on internet use during the 2008 electoral campaign found that it is "usually difficult" for 56% of online adults to distinguish between what is true and what is false on the internet (Smith, 2011).

The online "anti-vaxx" movement has exploited this ambiguity between true and false information in creating an audience. Content analyses of prominent anti-vaccine websites- most of which present themselves as objective explorations of truth- reveal a common narrative placing doctors and other authorities in opposition to concerned parents. Equating the decision to vaccinate with passive, ignorant parenting, and a vaccine-skeptical stance as personal autonomy and empowerment for parents, rejection of the scientific evidence supporting vaccines is seen as "informed choice." The movement encourages parents to assume the role of self-taught experts on vaccinations and their child's health (Kata, 2010). It also accuses the medical

establishment of hiding information about the negative consequences of vaccines in order to maximize profits for themselves and for pharmaceutical companies (ECDC, 2012; Betsch et al., 2012). Anti-vaccine websites frequently reject statistics and scientific evidence- indeed, the scientific process as a whole- on the grounds that the true risks of vaccination are subjective, unknown, and ultimately *unknowable* (Hobson-West, 2007; Brownlie and Howson, 2005).

Experimental evidence suggests that anti-vaccine websites can influence real-world vaccine decisions through changes in the preference for vaccination. Viewing vaccine-critical content increases the perceived risk of vaccines, while decreasing the perceived risk of not vaccinating. The perceived benefit of vaccines also declines. The impact of the personal narrative evidence frequently characterizing such websites is especially strong. Overall, there is a significant connection between exposure to anti-vaccine websites and reduced vaccine intentions (Betsch et al., 2010; Betsch et al., 2011; Kata).

Indeed, beliefs are fundamental to vaccination behavior. A significant share of parents who choose to delay or decline one or more required vaccines believe that vaccines cause both short-term side effects and potentially longer-term, more severe side effects. For these parents, following the CDC-recommended schedule is *more dangerous* than the alternative of no immunizations at all, or delayed ones. Such beliefs are widespread even among parents who adhere to the schedule (Dempsey et al., 2015). Vaccines are perceived, in many cases, as riskier than the diseases they prevent (Saada et al., 2015).

3 Data and Methodology

3.1 Immunization Data

Immunization data is drawn from National Immunization Survey (NIS) data for the years 1995-2008. The NIS is conducted annually, and forms the basis for state- and local-level estimates of immunization rates in the U.S. Toddler-aged children (19-35 months) constitute the population of interest. The survey had two phases during this period. In phase one, a random-digit-dialed telephone survey targeted a list of households with landline telephones and toddler children. In phase two, the child's immunization provider was contacted by mail survey.

The telephone and immunization provider surveys yielded sufficient information to draw up a detailed vaccination history for the child in 62 to 73% of all cases.

The county sample exploited in this paper is composed of the 257 counties for which the NIS sample size exceeded 35 observations at least once in seven biennial periods. The estimated immunization rates thus move every two years (1999-2000, 2001-02, ...). As American territory is divided into 3,143 counties, only a small proportion of these- the relatively populous ones- are included in the sample.

In this paper, the outcomes of interest are eight immunization series, measured from 2000 to 2008 for the population of children aged 19-35 months. The seven individual series analyzed are Diptheria-Tetanus-Pertussis (DTP; at least four shots); polio (three shots); MMR (one shot); Haemophilus influenzae (Hib; three shots); Hepatitis B (three shots); varicella (one shot); and the pneumococcal conjugate vaccine (PCV7; four shots). The *full coverage* rate, or 4:3:1:3:3:1:4 series, is defined as the proportion of children who have received all these shots. Table 2 summaries the county-level dynamics of each series from 2000 to 2008. The two vaccines most recently added to the official CDC schedule, varicella and PCV7, have initially low levels of coverage and experience explosive growth over time. The full coverage rate reflects their dynamics. Older vaccines experience modest growth in coverage rates, with one exception: the Hib vaccine. On average, this vaccine has the highest coverage rate of any series. Yet during this period, the proportion of toddlers immunized against the disease fell by 2.38%, from about 93% to 91%.

The average increase in vaccine coverage nationwide masks considerable heterogeneity at the local level. Returning to the herd immunity thresholds presented in Table 1, we find that the average county does not enjoy herd immunity against pertussis during this period. Indeed, the CDC reports that the number of nationwide pertussis cases rose steadily each year, from 7,867 in 2000 to 13,278 in 2008. Diphtheria also poses a potential threat, although only four cases were reported over this time period.

Measles, mumps and rubella coverage exceed the thresholds for those diseases, on average. Nonetheless, local-level outbreaks of each of these diseases have occurred. Two separate outbreaks in the first half of 2008 resulted in 131 cases of measles; of these, 91% of patients were either unvaccinated or of unknown vaccination status. The majority of these were children.

Throughout the entire U.S., measles cases rose from 86 in 2000 to 140 in 2008 (CDC). The CDC has concluded that the crucial factor in avoiding the trasmission of VPDs is maintaining high coverage rates at the local level: "Measles outbreaks can occur in communities with a high number of unvaccinated persons" (CDC, 2008).

3.2 Broadband Provider Data

In 1996, Congress enacted the first major change to telecommunications law since 1934. The aim of the Telecommunications Act was to remove regulatory barriers to entry and promote local competitiveness of the growing internet services industry. To this end, in March 2000 the FCC began to collect data from providers of *high-speed internet* (also termed "broadband"), defined as over 200 Kpbs in at least one direction.

The semi-annual FCC Broadband Progress Reports detailing the U.S. market for internet services between December 1999 and June 2008 provide the basis for my data analysis. In December 1999 ISPs supplied 2.8 million high-speed lines in the United States, including 1.8 million lines to residential and small business subscribers. By June 2000, this number had increased to 4.3 million HS lines, including 3.1 million to residential and small business subscriptions. One year later, this had grown to 9.6 million HS lines and 7.8 million residential and small business subscriptions (FCC, 2000a, 200b, 2002). By June 2008 there were 132.8 million lines in operation to homes and businesses (FCC, 2009).

Initially, however, the majority of internet subscriptions were not high-speed at all, but dial-up services that utilized the existing infrastructure of telephone lines. Interexchange Carriers, also known as long-distance phone companies, controlled the basic transmission mechanism for internet backbone traffic. Even more important, however, were the Local Exchange Carriers (LECs)- local telephone companies- which possessed the *last mile* of copper wire supplying telephony and internet services to nearly every home and business in the U.S. This monopoly granted the incumbent LECs a considerable amount of market power in the burgeoning market.

Newer technologies enabled data transmission at much higher bandwidths. Digital subscriber lines (DSL) provided increased bandwidth from the existing copper loops of local telephone companies, and unlike dial-up service did not interfere with the carriage of voice service. By

1998, many incumbent LECs had begun to offer DSL service. At the same time, the innovation of broadband access via coaxial cable enabled a new participant- local cable companies- to enter the market. This entrance was eased by the fact that 2/3 of American households already subscribed to cable television. As with LECs, local cable companies enjoyed market power through the ownership of the last mile of coaxial cable. This presented a challenge to outside ISPs. While many LECs and cable companies themselves entered the market as internet providers, outside ISPs depended on the last mile facilities of others for access to retail customers (FCC, 1998). Nonetheless the number of ISPs operating in the U.S. grew quickly over time.

Other sources of broadband service included satellite, wireline, and fixed wireless technologies (FCC, 2000a), although these represent a minority of total supply from 1999 to 2008. Most households accessed the internet through coaxial cable (93% of high-speed lines in 2000; 34.1% in 2008), even as DSL service grew much faster, from only 7% of high-speed lines in 2000 to 27.3% in 2008.

3.2.1 Limitations of the FCC Data

The FCC only collects data from high-speed internet providers, not the universe of providers as a whole. We thus lack information about the dial-up market. However, there is reason to focus on broadband rather than internet access in general. As discussed above in the context of health and science news, broadband access is associated with both a substitution away from other forms of media, and a change in how individuals use the internet. Not only do they spend more time online, but there is extensive evidence of behavioral differences in home broadband users. In addition to looking up information, they are more likely to share files and download media of all types (Horrigan and Rainie, 2002). Even more importantly, broadband users are an important source of user-generated content. To a far greater extent than dial-up users, they post personal content like artwork, photos and stories on the internet; they are also more prone to create their own websites (Horrigan, 2007). It is reasonable to predict that high-speed access will also have a differential impact upon off-line behavior like vaccinations.

The most important limitation is the level at which the FCC collects and reports data. Until June of 2008, the FCC only provided data at the state and zip-code level. Because the CDC

immunization data is reported by U.S. county, it is necessary to aggregate the zip-code data up to the county level. I perform this operation using the Missouri Census Data Center's MABLE/Geocorr web application. Each zip-code observation was matched to one or more counties and then weighted by the proportion of county population that it contributes. County-level figures for number of high-speed service providers therefore represent an extrapolation from the relative data of all constituent zip codes.

A second limitation is that the FCC only reports number of high-speed service providers at the zip-code level, ignoring the demand side of the market. Number of providers as a common proxy for degree of internet penetration in the literature. According to the provider metric, every county in my dataset had internet access by 1999; yet the degree of penetration varies widely, from a minimum of 0.01 to a maximum of 20.38 providers.

Finally, zip codes with one, two or three providers are grouped together into a single class. A Freedom of Information Act request for the actual number of providers in these zip codes was refused on the grounds that the information represented a trade secret, and release of it could cause substantial competitive harm (Exemption 4 to the Act). I assign a value of "2" to these zip codes.

Despite these limitations, there is a strong correspondence between the constructed county measures for the December 1999-June 2008 period and the county-level data reported by the FCC for subsequent periods. In December 2008, the FCC began to report data for three new variables at the county level: Total number of residential high-speed lines, per 1000 households; number of ISPs providing internet to small businesses and households; number of ISPs providing mobile service; and total number of ISPs present in the county.

Tables 3 and 4 demonstrate the interrelationship between the supply- and demand-side variables variables. Reports in Table 3 are the correlations between my constructed provider measure for June 2008 and the four FCC variables in December 2008 and from 2008-2011 (an average of six observations). The correlation between my measure for June 2008 and the December 2008 provider data is quite high, equal to 0.80 (column (1)). As column (2) shows, this strong correlation persists across longer time periods. My measure is also correlated with number of residential providers (ρ= 0.62) and mobile providers (ρ= 0.50).

Residential high speed lines also exhibit a strong, positive correlation with the constructed provider measure (ρ= 0.46). The provider measure therefore represents a good proxy for the unobserved demand side of the market. As more providers enter the market, more households purchase home broadband subscriptions. An increase in the number of ISPs present in a county creates a more competitive market for broadband provision, which translates into expanded access for the local population.

Table 4 reports the same set of correlations, substituting the December 2008 FCC data for my constructed June measure. The relationships are stronger, yet of roughly the same magnitude and display a similar degree of persistence over time. The tables present compelling evidence of two facts. My constructed measure provides a good representation of the true number of providers per county for the 1999-2008 period; and in all periods there is a positive relationship between the supply and demand sides of the market.

The county-year averages plotted in Figures 1 and 2 provide a graphical representation of the relationship between my constructed provider and the eight immunization series from 2000 to 2008. The average number of providers per county increased by 300%, from 2.88 in 2000 to 11.54 in 2008. The two newest immunization series, varicella and PCV7, appear to track growth in number of providers particularly closely; the full coverage rate follows the same overall trend. The DTP, polio, MMR and Hepatitis B series also increase slowly over the course of the decade (as shown in Table 4), though the relationship with number of providers seems less clear. The Hib series follows the opposite trend- falling as broadband provision rises- a result confirmed by the estimation results presented in section 4.

3.3 Instrumenting Broadband Provision: The Technet Index

Although I demonstrated the pattern of internet provision to be exogenous at the extremely disaggregated zipcode level in chapter 1, this does not hold at the level of counties, which are far larger. Thus the effect of access to broadband internet upon vaccine uptake cannot be identified with OLS. Among the earliest actors to adopt high-speed internet were educated individuals with a high degree of technological knowledge, who tended to be clustered in those regions of the country connected to the NSFNET backbone (precursor to the modern

internet). If early adopters have a different understanding and opinion of the benefits and risk of immunization, this would generate a spurious correlation between internet supply and immunization rates. Similarly, reverse causality could arise if families subscribed to broadband internet in order to be able to perform online research about health and science. These families might choose to connect to the internet because of a particular interest in vaccination and other health topics, not vice-versa.

Identification of the treatment effect depends on the ability to isolate that part of variation in internet supply that is exogenously determined. In this paper, I employ Technet's 2002 State Broadband Index. Two previous projects analyzing the impact of broadband internet in the U.S. in this period have employed the same instrument, Lelkes et al. (2015) and Larcinese and Miner (2017). The Technet Network is a national association of executives from companies focusing on fields such as information technology, biotechnology and venture capital. The Technet Index is an attempt to quantify the impact of state policies upon the supply and demand sides of the broadband market. In order to build and maintain the infrastructure necessary to supply end users with broadband, ISPs must be able to access public rights-of-way. Although individual municipalities determine the policies governing ISP access, states can limit their ability to impose excessive regulations, delays and costs. Whether such limits are imposed is fundamental to determining whether or not ISPs are permitted to enter into local markets, and hence the speed with which broadband spreads. The Index also accounts for whether state policies encourage private sector investment and incentivize entrance into underserved areas. The Index is a composite measure of all of these factors (with the greatest weight given to rights-of-way policies). In 2002, the state of Michigan had the highest index score, equal to 144.4, while Wyoming was last with a score of 11.25.

As the first-stage results of the regression analysis presented in section 4 will show, state-level policies do influence local-level provision decisions. Higher Technet scores are associated with a higher number of providers per county, all else equal. However, identification requires that the index itself be exogenous to immunization behavior. I will address the exogeneity assumption in further detail in future sections, and seek to demonstrate that it holds through a number of tests for selection presention in section 5.

3.4 County Control Variables

Many of the same factors have been shown to be associated with both demand for internet and for immunizations. Demographic factors that shape individual demand for internet include age, ethnicity and education. Macroeconomic factors like income and market size are crucial in determining whether ISPs enter into a given market (FCC, 2000).

Attitudes towards vaccines and vaccination behavior are strongly correlated with characteristics like income, ethnicity, age and level of education (Funk et al., 2017). The foreign-born population also matters for vaccination rates. Immigrant populations tend to exhibit systematically different rates of vaccination from native Americans, and to vaccinate their children at different rates.[1] Furthermore, diseases that are no longer endemic in the United States- like measles- continue to be imported from endemic countries by international travelers (Clemmons et al., 2017). Such localized outbreaks have been shown to influence immunization rates in turn, generating a feedback effect (Oster, 2017).

In this paper, I include controls for economic status (median household income, the poverty rate of families with dependent children under age 17, and the employment rate), education (percent of population with a college degree or higher) and demographic make-up (percent foreign-born, black and Hispanic). All controls are taken from the 1990 and 2000 Censuses, and the American Community Surveys of 2005-2008.

Table 5 presents estimation results for a Poisson regression of number of providers on the set of control variables. Larger populations attract more providers. Communities with more foreign-born and Hispanic residents attract fewer. Interestingly, there appears to be a negative association between proportion of college graduates and provision.

Later in the paper, I divide the sample of counties in correspondence to a number of different explanatory variables in order to analyze the heterogeneous effects of broadband access upon immunization. The results are modest. However, they confirm that features such as education, family poverty and population demographics play a role in mediating the effect of broadband.

[1] A Somali-American community in Minnesota had among the highest vaccine coverage rates in the state until 2008, when a growing number of autism diagnoses caused concern about the MMR vaccine. The community entered into communication with the anti-vaccine movement, and coverage rates plummeted. As a consequence, the community experienced a serious outbreak of measles in May of 2017 (Howard, 2017).

3.5 Empirical Model and Estimation

I test the hypothesis that broadband access influences vaccination rates by pooling together three difference models, each referring to a different three-year period: 2003-2006, 2004-2007 and 2005-2008. The model estimated for each period is given by:

$$\Delta v_{is} = \psi \Delta providers_{is} + \beta \Delta X_{is} + \varepsilon_{is}$$

where Δv_{is} is the three-year difference in the immunization series of interest, measured in county *i* in state *s*. The main explanatory variable is change in number of providers in county *i*, $\Delta providers_{is}$. This is instrumented by the Technet score of state *s* in the instrumental variables model. The vector of time-differenced county control variables is given by ΔX_{is}. Standard errors are clustered at the state level.

The pooled model is given by:

$$\Delta v_{ist} = \psi \Delta providers_{ist} + \beta \Delta X_{ist} + \tau_t + \varepsilon_{ist}$$

with t = 2006, 2007 and 2008. The pooled model includes year fixed effects τ_t. Once more, standard errors are clustered at the state level.

The observed difference in outcomes sweeps away time-invariant differences between counties and is thus equivalent to a fixed-effects specification. The key identifying assumptions of the IV model are relevance (demonstrated by the first stage), exogeneity and exclusion. The value of the state-level Technet Index in 2002 must be exogenous to all time-varying unobservable factors correlated with county-level variation in immunization rates. Moreover, the Technet Index should influence immunization rates only through number of providers; any direct influence of the index on immunization behavior would violate the exclusion restriction.

In section 5, I perform a number of tests to evaluate whether the exogeneity restriction holds. First, I analyze whether the state-level immunization and exemption laws in force in 2001 are correlated with the Technet scores. Next, I test for selection on trends in the immunization series from 1995 to 2001 (available for five of eight series). Finally, I test for selection on the 2001 levels of the immunization series. Each of these tests provides evidence that no unobservable variables influenced both immunization behavior and the state broadband regulations summarized by the index. I will also address the exclusion restriction in greater detail in that

section.

4 Broadband Access and County Immunization Coverage

I open this section with a presentation of the OLS estimation results for the eight immunization series. Next, I review the estimates of the instrumental variable model and discuss their relationship to initial predictions about the impact of broadband. Finally, I disaggregate the dataset based upon four different control variables, in order to analyze whether the provider variable has a heterogeneous effect upon different groups. I compare the behavior of counties with "high" and "low" values of educational attainment, family poverty, share of foreign-born residents, and share of black or hispanic residents.

The first column of Table 6 reports the OLS estimate for the 4:3:1:3:3:1:4 full coverage series. The point estimate is positive but not statistically significant. Three of the seven individual series estimates are statistically different from zero: Broadband provision is associated with a drop in coverage for DTP, polio and Hib (Table 7). This does not hold true for the MMR, Hepatitis B, varicella and PCV7 series, however.

I find a negative impact of internet use upon the full coverage rate. An increase in number of providers of one standard deviation causes a very modest drop in the full coverage rate of 0.07 points- amounting to about 0.11% of the mean of the dependent variable. The coefficient estimate is significant at the 5% level.

The Technet Index impacts the number of providers positively. A one standard deviation increase in a state's score causes mean number of providers per county to rise by 0.58, equal to 5% of the dependent variable mean. Moreover, the relevance assumption easily holds for the instrument: The first-stage F-statistic is equal to 26.21.

Turning to the IV results for the seven individual series (Tables 8-10), I find that the drops in DTP and polio coverage predicted by the OLS results are not reflected in the IV estimates. At the county level, broadband access influences only a single immunization rate, that of Haemophilus influenzae. The magnitude of the effect is very small. A one standard deviation

increase in providers causes the Hib coverage rate to fall by 0.67 points, or 0.7% of the sample mean. The estimate is significant at the 10% level. The result suggests that the drop in the full coverage rate for all immunizations may reflect decreased uptake of the Hib vaccine, specifically.

4.1 Heterogeneous Effects of Broadband Provision

For two observable variables- share foreign and share black or Hispanic- I contrast counties in the top tercile with those in the bottom tercile. This was not possible for educational attainment and family poverty, as the Technet instrument was too weak in the case of the bottom terciles to be able to draw any inference. I therefore contrast counties above and below the median in these two cases. The pattern established in the main IV analysis of few significant results persists when one disaggregates the dataset to evaluate heterogeneous effects. However, the estimated effect is somewhat larger in magnitude in the smaller populations.

Broadband access does not drive any behavioral changes for the set of counties with above-median educational attainment (Table 11). Below the median, however, we see drops in both the full coverage and Hib rates; each of these estimates is significant at the 5% (Table 12). Moreover, the full coverage coefficient is much larger that that estimated for the main sample. In this case, vaccine coverage falls by four points, equal to 6.5% of the sample mean. The estimated impact on Hib uptake is equal to 2.1% of that mean.

Counties with above-median family poverty rates experienced decreases in full coverage and varicella uptake, on average (estimates significant at the 10% level; Table 13). This result is quite interesting, given the dramatic increase in both series throughout the period. While average households increasingly chose to invest in immunization, broadband access instead discouraged it in those counties with higher concentrations of poverty. Relatively wealthy counties saw no change in overall coverage, but did experience a decline in Hib uptake that is signficant at the 5% level (Table 14).

Broadband exercised a negative impact on varicella uptake for counties in the top tercile of foreign-born residents (Table 15). For counties with the greatest share of native-born residents, Hib rates declined due to expanded broadband access. Both estimates are significant at the

10% level (Table 16).

Counties with the lowest proportion of black and Hispanic residents experienced no change in immunization rates as a result of broadband (Table 17). Hib uptake fell for those counties with the highest share of these two ethnic groups (significant at the 10% level; Table 18).

Putting these results together, we obtain a clearer picture of how broadband provision influenced immunization coverage during this period, and which population groups drove the change. The negative treatment effect upon the 4:3:1:3:3:1:4 coverage rate was concentrated in those counties with relatively fewer college graduates and higher family poverty- in other words, counties undergoing economic decline. Although these types of counties have adopted broadband more slowly (largely due to the reluctance of ISPs to enter such unlucrative markets), once they do, broadband use has a dramatic impact upon immunization behavior. More highly educated and richer counties have not seen the same effect, with one important exception.

A wide cross-section of county types have experienced declines in Hib coverage due to broadband access. Counties that are less educated, wealthier, with more native-born residents, and with more black and Hispanic residents have all experienced this effect. That all of these diverse, and not necessarily overlapping, population groups were significantly impacted has contributed to the result we saw in the main analysis, a fall in Hib coverage in the general population. Indeed, the decline in uptake of this vaccine, which was initially higher than that of any other series (Figure 2), appears to be the main driving force behind the decline in the full coverage rate resulting from broadband.

Poorer counties, as well as those with a greater foreign-born share, both experienced declines in the varicella coverage rate due to broadband. These declines were not experienced by other types of counties, however, and were therefore insufficient to drive a change in the general population.

5 Exogeneity of the Technet Index

I now review the results of a number of tests evaluating whether the Technet Index is exogenous from time-varying unobservable factors correlated with immunization trends, the crucial assumption allowing us to interpret the coefficient estimates as a true treatment effect. First, I examine the state-level immunization laws that were in force for admission to childcare centers in the year 2001. Data on these requirements is drawn from the Immunization Action Coalition webpage detailing state-by-state requirements.[2] For each immunization series, I construct a dummy variable for whether that vaccine was required in the year 2000. I sum the dummy across all series to obtain a count variable of the number of immunizations required by each state. I also consider the number of doses each state regards as sufficient for immunization against a given vaccine-preventable disease; again summing across all series, I obtain a count variable of minimum number of doses each state requires in order for a child to be considered fully immunized. Finally, I consider whether a state allows temporary medical, permanent medical, religious or philosophical exemptions.

Coefficient estimates from the regression of the Technet Index on these variables are reported in Table19. As the results make clear, there is no correlation between state right-of-way laws and those laws governing number of vaccines, number of doses, and which exemptions are permissible. Although state legislatures are responsible for drawing up both types of laws, policies concerning public health and those dealing with information technology are determined in an independent way.

In the next two tests, I look at past trends and levels in immunization behavior. If these are found to be correlated with the Technet Index, then it is probable that there is some underlying force that drives both. In Table 20, the Technet Index is regressed on the 1995 to 2001 trends in DTP, polio, MMR, Hib and Hepatitis B (data on varicella, PCV7 and the full coverage rate are not available for this period). There is no relationship between prior immunization trends and Technet scores. In Table 21, the Technet Index is regressed on 2001 uptake for the four vaccines included in Table 20, in addition to varicella. Once more, state policies regarding broadband access are found to be completely exogenous from county-level

[2] Available at http://www.immunize.org/laws/.

immunization behavior.

Taken together, the three tests provide strong evidence for the exogeneity of the Technet Index. Relevance was demonstrated by the first-stage F-statistics of the IV regression results. One final concern remains regarding the validity of the instrument: The exclusion restriction. State laws promoting the expansion of broadband must not directly influence vaccine uptake. Right-of-way and supply-side initiatives target the behavior of Internet Service Providers, not of individual consumers, nor do they impact the provision of vaccinations by local health clinics. Because these regulations are independent of the market for preventative healthcare, driving immunization behavior only through the incentives to adopt broadband, exclusion- while impossible to verify empirically- is likely to be satisfied.

6 Discussion

In this paper, I have employed an instrumental variables estimation strategy to investigate the effect of broadband internet upon county immunization rates. The 2002 Technet Index, a composite measure of state legislation governing the ability of ISPs to access public rights-of-way and enter local markets, was employed to instrument the number of broadband providers per county. Several clear findings emerge. First, expanded access to broadband internet does not have a strong impact on immunization rates at the county level. Coefficient estimates are small in magnitude, and frequently significant at only the 10% level. Second, the most important impact of broadband is upon Hib uptake. In the main analysis, I find that a one standard deviation increase in number of providers providers causes the Hib coverage rate to fall by 0.67 points. This decline in Hib coverage is the driving force behind the negative effect observed for the 4:3:1:3:3:1:4 full coverage series; an increase in number of providers of one standard deviation causes an extremely small drop of 0.07 points in the full coverage rate.

The situation changes for the analysis of heterogeneous effects; these estimates tend to be larger in magnitude, and more significant. I find that those counties with lower educational attainment, more native-born residents, and more black and Hispanic residents all experienced a drop in the Hib coverage rate as a result of broadband access. Surprisingly, the result also holds for counties with lower family poverty rates. The decline in the full coverage rate is

instead due to counties with lower educational attainment and higher poverty. Finally, counties that are poorer or have more foreigners tended to see a drop in varicella coverage, although this treatment effect is not significant in the general population.

Haemophilus influenzae was the only immunization series to undergo a decline from 2000 to 2008. Broadband access appears to have played a very modest role in this decline. At the same time, the full coverage rate rose dramatically- in contrast to the negative and significant effect that I find for broadband internet. This result suggests that although internet use caused a small decline in vaccine uptake, larger and more important forces drove a change in the opposite direction. In fact, every treatment effect estimate that resulted significant was negative. Immunization coverage rose over time for the set of U.S. counties, but this increase was not due to the increase in good health information available on the internet; rather, the "information overload" effect appears to have prevailed, causing certain types of families to discount the expert opinion of doctors and government health departments and instead trust alternative sources of information about their childrens' health.

Adopting the same Technet Index as an instrument, together with a matching estimation strategy, Larcinese and Miner find that broadband internet increased voter turnout in the 2008 presidential election. A within-pair standard deviation increase in number of providers caused turnout to increase by about 2%. Lelkes et al., focusing upon the outcome of partisan hostility, find that broadband provision increased partisan hostility by 0.01 points from 2000 to 2004 and by 0.02 points from 2004 to 2008, relative to a baseline value of about 0.65. Both of these estimates are larger than what I observe for immunization rates in the of U.S. counties, but consistent in magnitude with the heterogeneous analysis results. At the county level, broadband internet appears to have a more dramatic impact on political outcomes than upon immunization rates, on average. As the estimation results in chapter 1 of this thesis suggest, broadband internet is more important at the very local level, for specific subsets of the larger population, and for other types of behaviors such as vaccine exemption rates.

References

[1] Abbate, J. 1999. *Inventing the Internet*. Cambridge: MIT Press.

[2] Andersen, T.B., J. Bentzen, C. Dalgaard and P. Selaya. 2011. "Does the Internet Reduce Corruption? Evidence From U.S. States and Across Countries." The World Bank Economic Review, 25(3): 387-417.

[3] Andersen, T.B., J. Bentzen, C. Dalgaard and P. Selaya. 2012. "Lightning, IT Diffusion, and Economic Growth Across U.S. States." *The Review of Economics and Statistics*, 94(4): 903-924.

[4] Betsch, C., N.T. Brewer, P. Brocard, P. Davies, W. Gaissmaier, N. Haase, J. Leask, F. Renkewitz, B. Renner, V.F. Reyna, C. Rossmann, K. Sachse, A. Schachinger, M. Siegrist and M. Stryk. 2012. "Opportunities and Challenges of Web 2.0 for Vaccination Decisions." *Vaccine*, 30(25): 3727-3733.

[5] Betsch, C., F. Renkewitz, T. Betsch and C. UlshÃ¶fer. 2010. "The Influence of Vaccine-Critical Websites on Perceiving Vaccination Risks." *Journal of Health Psychology*, 15(3): 446-55.

[6] Betsch C., C. UlshÃ¶fer, F. Renkewitz and T. Betsch. 2011. "The Influence of Narrative v. Statistical Information on Perceiving Vaccination Risks." *Medical Decision Making*, 31(5): 742-53

[7] Brownlie, J. and A. Howson. 2005. "Leaps of Faith and MMR: An Empirical Study." *Sociology*, 39(2): 221-239.

[8] Campante, F., R. Durante and F. Sobbrio. 2013. "Politics 2.0: The Multifaceted Effect of Broadband Internet and Political Participation." NBER Working Paper 19029.

[9] Centers for Disease Control and Prevention. 2008. "Update: Measles– United States, January-July 2008." Morbidity and Mortality Weekly Report.

[10] Centers for Disease Control and Prevention- National Center for Immunization and Respiratory Diseases. 2015. "Vaccines and Immunizations: Glossary." https://www.cdc.gov/vaccines/terms/glossary.html

[11] Clemmons, N.S., G.S. Wallace, M. Patel and P.A. Gastanaduy. 2017. "Incidence of Measles in the United States, 2001-2015." *Journal of the American Medical Association*, 318(3): 1279-1281.

[12] Delamater, P.D., T.F. Leslie and Y.T. Yang. 2017. "Change in Medical Exemptions for Immunization in California After Elimination of Personal Belief Exemptions." *Journal of the American Medical Association*, 318(9): 863-864.

[13] DellaVigna, S. and E. Kaplan. 2007. "The Fox News Effect: Media Bias and Voting." *The Quarterly Journal of Economics*, 122(3): 879–924.

[14] Dempsey, A.F., S. Schaffer, D. Singer, A. Butchart, M. Davis and G.L. Freed. 2011. "Alternative Vaccination Schedule Preferences Among Parents of Young Children." *Pediatrics*, 128(5): 848-856.

[15] European Center for Disease Prevention and Control. 2012. "Communication on Immunisation – Building Trust." Stockholm: ECDC.

[16] Fangjun, Z., A. Shefer, J. Wenger, M. Messonnier, L.Y. Wang, A. Lopez, M. Moore, T.V. Murphy, M. Cortese and L. Rodewald. 2014. "Economic Evaluation of the Routine Childhood Immunization Program in the United States, 2009." *Pediatrics*, 133(4): 577-585.

[17] Falck, O., R. Gold and S. Heblich. 2014. "E-lections: Voting Behavior and the Internet." *American Economic Review*, 104(7): 2238-2265.

[18] Federal Communications Commission. 1998. "First Broadband Progress Report."

[19] Federal Communications Commission. 2000a. "Federal Communications Commission Releases Data on High-Speed Services for Internet Access."

[20] Federal Communications Commission. 2000b. "Deployment of Advanced Telecommunications Capability: Second Report."

[21] Federal Communications Commission. 2002. "Third Broadband Progress Report."

[22] Federal Communications Commission. 2008. "Fifth Broadband Progress Report."

[23] Fine, P., K. Eames and D.L. Heymann. 2011. "Immunizations: A Rough Guide." *Clinical Infectious Diseases*, 52(7): 911-916.

[24] Fox, S. 2005. "Health Information Online." Pew Internet & American Life Project.

[25] Fox, S. 2006. "Online Health Search 2006." Pew Internet & American Life Project.

[26] Fox, S. and S. Jones. 2009. "The Social Life of Health Information." Pew Internet & American Life Project.

[27] Fox, S. and L. Rainie. 2000. "The Online Health Care Revolution." Pew Internet & American Life Project.

[28] Funk, C. 2015. "Public and Scientists' Views on Science and Society." Pew Research Center.

[29] Funk, C., B. Kennedy and M. Hefferon. 2017. "Vast Majority of Americans Say Benefits of Childhood Vaccines Outweigh Risks." Pew Research Center.

[30] Gavazza, A., M. Nardotto and T. Valletti. 2016. "Internet and Politics: Evidence from U.K. Local Elections and Local Government Policies."

[31] Gentzkow, M. 2006. "Television and Voter Turnout." *The Quarterly Journal of Economics*, 121(3): 931-972.

[32] Greenstein, S. 2015. *How the Internet Became Commercial*. Princeton: Princeton University Press.

[33] Hinman, A.R., W. A. Orenstein and A. Schuchat. 2011. "Vaccine-Preventable Diseases, Immunizations, and MMWR — 1961–2011." *Morbidity and Mortality Weekly Report Supplements*, 60(04): 49-57.

[34] Hobson-West, P. 2007. "'Trusting Blindly Can Be the Biggest Risk of all': Organised Resistance to Childhood Vaccination in the UK." *Sociology of Health and Illness*, 29(2):198-215.

[35] Horrigan, J.B. 2006. "Home Broadband Adoption 2006." Pew Internet & American Life Project.

[36] Horrigan, J.B. 2006. "The Internet as a Resource for News and Information about Science." Pew Internet & American Life Project.

[37] Horrigan, J.B. 2007. "Broadband: What's All the Fuss About?" Pew Internet & American Life Project.

[38] Horrigan, J.B. and L. Rainie. 2002. "The Broadband Difference: How Online Behavior Changes With High-speed Internet Connections." Pew Internet & American Life Project.

[39] Howard, J. 2017. "Anti-vaccine Groups Blamed in Minnesota Measles Outbreak." *CNN*, http://edition.cnn.com/2017/05/08/health/measles-minnesota-somali-anti-vaccine-bn/index.html.

[40] Jensen, R. and E. Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *The Quarterly Journal of Economics*, 124(3): 1057–1094.

[41] Jolley, D. and K.M. Douglas. 2014. "The Effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions." *Plos One*, 9(2).

[42] Kata, A. 2010. "A Postmodern Pandora's Box: Anti-Vaccination Misinformation on the Internet." *Vaccine*, 28: 1709-1716.

[43] Kearney, M.S. and P.B. Levine. 2014. "Media Influences on Social Outcomes: The Impact of MTV's 16 and Pregnant on Teen Childbearing." NBER Working Paper 19795.

[44] La Ferrara, E. 2016. "Mass Media and Social Change: Can We Use Television to Fight Poverty?" *Journal of the European Economic Association*, 14(4): 791-827.

[45] La Ferrara, E., A. Chong and S. Duryea. 2008. "Soap Operas and Fertility: Evidence from Brazil." Inter-American Development Bank Working Paper 633.

[46] Lahart, J. 2009. "There Go the Servers: Lightning's New Perils." *The Wall Street Journal*, March 22.

[47] Larcinese, V. and L. Miner. 2017. "The Political Impact of the Internet on U.S. Presidential Elections."

[48] Lelkes, Y., G. Sood and S. Iyenger. 2015. "The Hostile Audience: The Effect of Access to Broadband Internet on Partisan Effect." *American Journal of Political Science*, 61(1): 5-20.

[49] Lewandowsky, S., G.E. Gignac and K. Oberauer. 2013. "The Role of Conspiracist Ideation and Worldviews in Predicting Rejection of Science." *Plos One*, 8(10).

[50] Meckel, R.A. 2004. "Levels and Trends of Death and Disease in Childhood, 1620 to the Present." *Children and Youth in Sickness and Health: A Handbook and Guide*, ed. J. Golden, R.A. Meckel and H.M. Prescott. Westport, CT: Greenwood Press.

[51] Mesch, G.S. and K.P. Schwirian. 2014. "Confidence in Government and Vaccination Willingness in the U.S." *Health Promotion Internation*, 30(2): 213-221.

[52] National Conference of State Legislatures. 2016. "States with Religious and Philosophical Exemptions from State Immunization Requirements." http://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx

[53] National Telecommunications and Information Administration. 1998. "Falling Through the Net II: New Data on the Digital Divide."

[54] Nichols, T. 2017. *The Death of Expertise*. Oxford: Oxford University Press.

[55] Oster, E. 2017. "Does Disease Cause Vaccination? Disease Outbreaks and Vaccination Response."

[56] Otto, S. 2016. *The War on Science.* Minneapolis: Milkweed Editions.

[57] Poland, G.A. and R.M. Jacobson. 2011. "The Age-Old Struggle Against the Antivaccinationists." *The New England Journal of Medicine*, 364(2): 97-99.

[58] Saada, A., T.A. Lieu, S.R. Morain, B.J. Zikmund-Fisher and E. Wittenberg. 2015. "Parents' Choices and Rationales for Alternative Vaccination Schedules: A Qualitative Study." *Clinical Pediatrics*, 54(3): 236-243.

[59] Smith, A. 2011. "The Internet and Campaign 2010." Pew Internet & American Life Project.

[60] Snyder, J.M. and D. Stromberg. 2010. "Press Coverage and Political Accountability." *Journal of Political Economy*, 118(2): 355-408.

[61] Stromberg, D. "Radio's Impact on Public Spending." 2004. *The Quarterly Journal of Economics*, 119(1): 189–221.

[62] Technet. 2002. "The State Broadband Index: An Assessment of State Policies Impacting Broadband Deployment and Demand." https://illinoisbroadbanddeployment.pbworks.com/f/State_Broadband_Index.pdf

Table 1: Herd Immunity Threshold, by Vaccine-Preventable Disease

	R_0[a]	q_c[b] (%)
Diphtheria	6-7	83-86
Pertussis	12-17	92-94
Polio	5-7	80-86
Measles	12-18	92-94
Mumps	4-7	75-86
Rubella	6-7	83-86

[a] Reproduction number
[b] Herd immunity threshold, calculated as $1 - \frac{1}{R_0}$
Data: Fine, 1993.

Table 2: Coverage Rate by Vaccine, 2000-2008

	Mean.	Min.	Max	% Change
4:3:1:3:3:1:4 series	52.9	11.6	80.1	+83.8%
DTaP/DTP *4 doses Diphtheria-Tetanus-Pertussis*	84.7	66.2	97.0	+2.1%
Polio *3 doses Polio*	91.5	78.1	96.8	+4.56%
MMR *1 dose Measles-Mumps-Rubella*	92.1	84.9	96.6	+1.18%
Hib *3 doses Haemophilus influenzae*	92.9	79.2	97.3	-2.38%
Hepatitus B *3 doses Hepatitis B*	91.3	76.7	96.6	+5.34%
Varicella *1 dose Varicella*	81.5	29.2	96.5	+49.55%
PCV7 *4 Pneumococcal*	61.2	12.6	91.0	+85.1%

Table 3: Comparison of Constructed Provider Data With FCC Measures

	2008	2008-11
Residential HS lines	0.460	0.434
Total providers	0.800	0.791
Residential providers	0.616	0.604
Mobile providers	0.496	0.662

Notes: Values are correlation of the constructed measure of total number of internet providers per county in June 2008 with FCC variables in December 2008 and 2008-2011 (average of six observations).

Table 4: Comparison of FCC Provider Data With Other FCC Measures

	2008	2008-11
Res. HS lines	0.534	0.512
Total providers	1	0.983
Residential providers	0.804	0.800
Mobile providers	0.491	0.798

Notes: Values are correlation of the FCC-reported total number of internet providers per county in December 2008 with the listed variables in December 2008 and 2008-2011 (average of six observations).

Table 5: Poisson Model of Number of Providers

	(1)
	Number of providers
Log population	.799*
	(.435)
Log income	.101
	(.289)
Family poverty rate	-.002
	(.002)
Employment rate	.007
	(.006)
College degree or higher	-.033***
	(.009)
Share foreign	-.034**
	(.013)
Share black	.002
	(.012)
Share Hispanic	-.051***
	(.014)
N	1923
Mean	7.572

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The regression includes year and county fixed effects. Standard errors are clustered at the county level.

Table 6: Full Immunization Results

	(1) OLS	(2) IV 1st	(3) IV 2nd
Technet		.021**	
		(.003)	
Number of providers	.086		-2.068**
	(.243)		(.933)
F-statistic		26.21	
N	362	362	362
Mean of dependent	63.922	11.133	63.922

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variable is the 4:3:1:3:3:1:4 immunization series. All specifications include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 7: Individual Series- OLS Results

	(1) DTP	(2) Polio	(3) MMR	(4) Hib	(5) Hepatitis B	(6) Varicella	(7) PCV7
Number of providers	-.237***	-.101*	-.001	-.194***	-.066	-.066	.066
	(.086)	(.056)	(.050)	(.067)	(.055)	(.127)	(.260)
Adjusted R2	.231	.169	.092	.312	.280	.380	.658
N	494	494	494	494	494	494	360
Mean of dependent	85.170	93.084	92.007	91.813	93.139	88.903	74.218

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the individual immunization series. All specifications include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 8: DTP, Polio and MMR- IV Results

	(1) DTP: IV 1st	(2) DTP: IV 2nd	(3) Polio: IV 1st	(4) Polio: IV 2nd	(5) MMR: IV 1st	(6) MMR: IV 2nd
Technet	.016**		.016**		.016**	
	(.003)		(.003)		(.003)	
Number of providers		-.304		.175		.032
		(.296)		(.218)		(.225)
F-statistic	17.79		17.79		17.79	
N	362	362	494	494	494	494
Mean of dependent	11.133	85.170	11.133	93.084	11.133	92.007

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are coverage rates for DTP (columns 1 and 2), polio (columns 3 and 4), and MMR (columns 5 and 6). All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 9: Hib and Hepatitis B- IV Results

	(1)	(2)	(3)	(4)
	Hib: IV 1st	Hib: IV 2nd	Hepatitis B: IV 1st	Hepatitis B: IV2
Technet	.016**		.016**	
	(.003)		(.003)	
Number of providers		-.412*		.024
		(.214)		(.237)
F-statistic	17.79		17.79	
N	494	494	494	494
Mean of dependent	11.133	91.813	11.133	93.139

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are coverage rates for Hib (columns 1 and 2) and Hepatitis B (columns 3 and 4). All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 10: Varicella and PCV7- IV Results

	(1) Varicella: IV 1st	(2) Varicella: IV 2nd	(3) PCV7: IV 1st	(4) PCV7: IV 2nd
Technet	.016**		.021**	
	(.003)		(.003)	
Number of providers		-.324		-.377
		(.393)		(.606)
F-statistic	17.79		26.14	
N	494	494	360	360
Mean of dependent	11.133	88.903	11.133	74.218

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are coverage rates for varicella (columns 1 and 2) and pneumococcal (columns 3 and 4). All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 11: Heterogeneous Effects by Education- Above the Median

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-1.424	.214	-.037	.372	-.272	-.086	-1.014	-.584
	(1.195)	(.556)	(.336)	(.485)	(.358)	(.368)	(.672)	(1.140)
F-statistic	13.87	14.29	14.29	14.29	14.29	14.29	14.29	14.29
N	236	237	237	237	237	237	237	237
Mean of dependent	65.617	86.516	93.291	92.495	92.315	93.336	86.609	76.112

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 12: Heterogeneous Effects by Education- Below the Median

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-1.163**	-.478	.088	.093	-.546**	.070	.059	-.325
	(.562)	(.292)	(.164)	(.186)	(.220)	(.226)	(.389)	(.679)
F-statistic	11.91	11.72	11.72	11.72	11.72	11.72	11.72	11.46
N	153	152	152	152	152	152	152	152
Mean of dependent	61.100	82.900	92.720	91.182	90.937	92.793	87.730	71.012

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 13: Heterogeneous Effects by Family Poverty- Above the Median

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-1.221*	-.321	-.054	-.091	-.339	-.028	-.864*	.251
	(.717)	(.364)	(.209)	(.200)	(.272)	(.292)	(.520)	(.804)
F-statistic	13.03	12.86	12.86	12.86	12.86	12.86	12.86	12.84
N	179	178	178	178	178	178	178	177
Mean of dependent	60.853	84.283	92.680	91.870	91.888	92.822	88.962	70.479

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 14: Heterogeneous Effects by Family Poverty- Below the Median

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-1.366	-.207	-.196	.267	-.659**	-.106	-.113	-1.173
	(.927)	(.415)	(.243)	(.395)	(.301)	(.267)	(.502)	(.931)
F-statistic	10.68	11.08	11.08	11.08	11.08	11.08	11.08	11.06
N	185	186	186	186	186	186	186	185
Mean of dependent	66.86	85.941	93.365	92.210	91.738	93.398	89.409	77.677

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 15: Heterogeneous Effects by Foreign Share- Top Tercile

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-1.675	.018	.182	-.109	-.156	-.578	-1.668*	-1.263
	(1.139)	(.502)	(.347)	(.446)	(.297)	(.411)	(.894)	(1.103)
F-statistic	22.05	17.80	17.80	17.80	17.80	17.80	17.80	22.54
N	167	167	167	167	167	167	167	167
Mean of dependent	62.467	85.468	92.527	92.169	92.619	92.713	89.983	71.948

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 16: Heterogeneous Effects by Foreign Share- Bottom Tercile

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-.277	-1.159	1.076	-.119	-.913*	.703	-.090	1.622
	(1.401)	(.847)	(.706)	(.490)	(.533)	(.531)	(.967)	(1.546)
F-statistic	2.16	11.11	11.11	11.11	11.11	11.11	11.11	2.28
N	78	78	78	78	78	78	78	78
Mean of dependent	60.135	84.638	93.418	91.813	91.922	93.348	86.819	70.262

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 17: Heterogeneous Effects by Share Black or Hispanic- Top Tercile

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-.658	.137	-.052	.282	-.336*	.189	-.210	-.427
	(.562)	(.318)	(.186)	(.186)	(.200)	(.228)	(.321)	(.630)
F-statistic	20.36	20.70	20.70	20.70	20.70	20.70	20.70	20.70
N	186	185	185	185	185	185	185	188
Mean of dependent	63.071	84.454	92.700	92.032	91.895	92.959	89.803	72.630

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 18: Heterogeneous Effects by Share Black or Hispanic- Bottom Tercile

	(1) Full Coverage	(2) DTP	(3) Polio	(4) MMR	(5) Hib	(6) Hepatitis B	(7) Varicella	(8) PCV7
Number of providers	-2.100	-.840	1.031	-1.317	.240	.053	1.112	.309
	(2.381)	(1.115)	(.900)	(1.060)	(.901)	(.703)	(2.097)	(2.037)
F-statistic	1.33	12.01	12.01	12.01	12.01	12.01	12.01	1.40
N	79	79	79	79	79	79	79	76
Mean of dependent	64.335	85.584	93.451	91.946	91.674	93.241	87.149	75.577

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables are the 4:3:1:3:3:1:4 series and the individual immunization series. All regressions include year fixed effects. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Table 19: Check for Selection on Immunization Laws

	(1) Technet	(2) Technet	(3) Technet
Required doses	-.825		
	(2.212)		
Required vaccines		-1.496	
		(4.017)	
Temporary medical			3.416
			(6.932)
Permanent medical			1.225
			(13.499)
Religious			1.588
			(16.093)
Philosophical			15.157
			(9.262)
R-squared	.003	.003	.073
N	51	51	51
Mean of dependent	31.569	31.569	31.569

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables is the 2002 Technet index. Immunization regulations refer to the year 2001. Standard errors are clustered at the county level.

Table 20: Check for Selection on 1995-2001 Trends in Immunization

	(1) Technet	(2) Technet	(3) Technet	(4) Technet	(5) Technet
DTP	.174				
	(.575)				
Polio		-.618			
		(.822)			
MMR			-.353		
			(.724)		
Hib				-.644	
				(.956)	
Hepatitis B					.304
					(.343)
R-squared	.001	.003	.001	.002	.004
N	162	162	162	162	162
Mean of dependent	31.569	31.569	31.569	31.569	31.569

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: The dependent variables is the 2002 Technet index. The immunization trends refer to the years 1995-2001. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the county level.

Table 21: Check for Selection on 2001 Immunization Levels

	(1) Technet	(2) Technet	(3) Technet	(4) Technet	(5) Technet	(6) Technet
DTP	-.473					
	(.654)					
Polio		-.569				
		(.800)				
MMR			-2.419			
			(1.530)			
Hib				-.845		
				(1.414)		
Hepatitis B					-.391	
					(.918)	
Varicella						-.323
						(.314)
Adjusted R2	.143	.142	.158	.141	.140	.144
N	210	210	210	210	210	210
Mean of dependent	31.569	31.569	31.569	31.569	31.569	31.569

$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Notes: The dependent variables is the 2002 Technet index. The immunization series are measured in the year 2001. The county-level controls are log population, log median household income, the family poverty rate, the employment rate, percent of population with bachelors degree or higher, share foreign, share black and share Hispanic. Standard errors are clustered at the state level.

Figure 1: Number of Providers vs. Immunization Series Trends

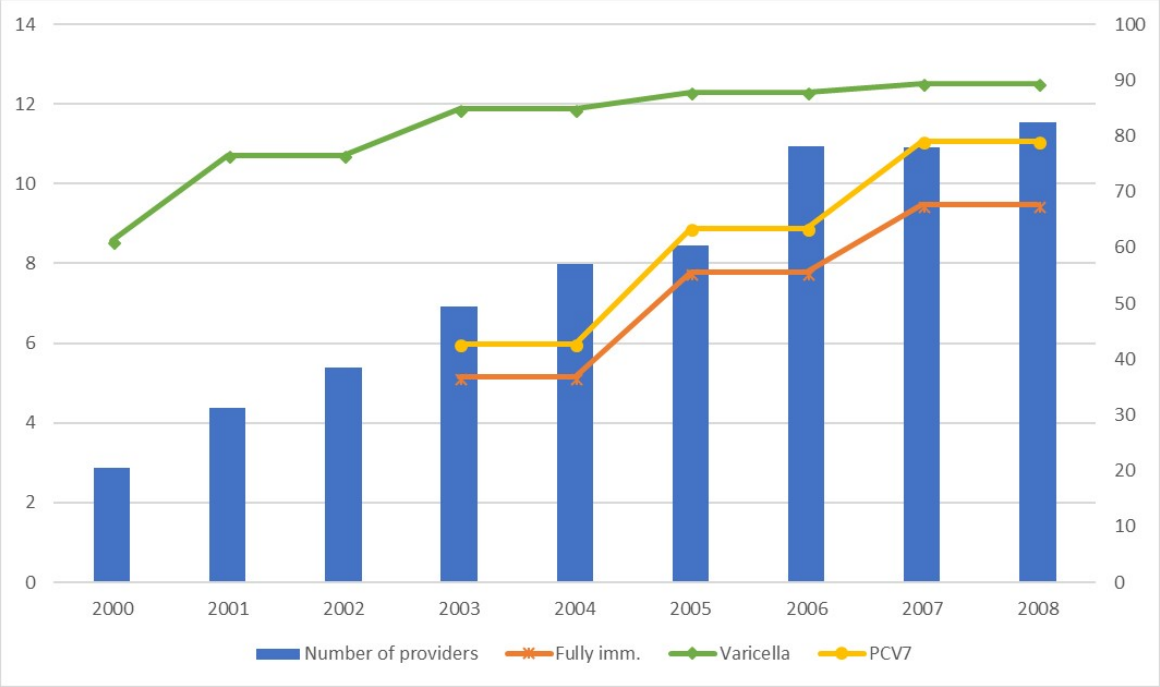
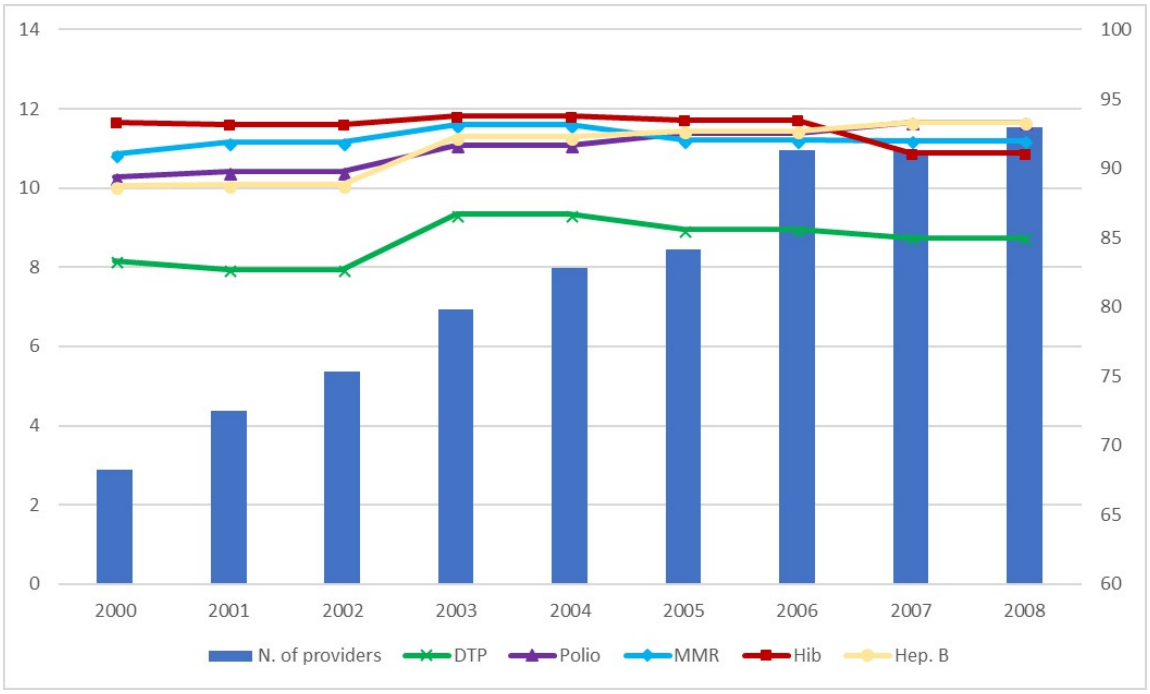


Figure 2: Number of Providers vs. Immunization Series Trends



Chapter 3:
Spillovers in Local-Level Conflict

Carrie Huffaker
Bocconi University

Advisor: Dr. Diego Ubfal

April 16, 2019

1 Introduction

Over the last three decades, war has claimed two millions lives in battle, and tens of millions more lives among civilians. Nearly 1000 unique conflicts have been fought since 1989. The vast majority of these conflicts are intrastate, that is civil wars, with frequent intervention by international actors. The Syrian and Iraqi civil wars and rise of Islamic State have caused hundreds of thousands of deaths and displaced millions more. The worst refugee crisis since World War II is still unfolding.[1]

Since the pioneering work of Saleyhan and Gleditsch (2006), which examined the role of refugees in the spread of civil war, researchers have focused new attention on the role of transnational forces in the outbreak of new conflict and intensification of old ones. An important step forward in this field was the innovation of spatial autocorrelation models, originally developed in the fields of epidemiology, biology and geography. These models are used to investigate interdependence in the outcomes of connected actors.

The fundamental component of spatial econometric models is the spatial weight matrix, *W*. By defining the weight matrix in different ways we can create new networks of spatial dependence. The unique power of the spatial weight matrix lies in the ability to explicitly evaluate distinct hypotheses about how conflict spreads between different localities. The two main forms of proximity I focus upon in this paper are geography and ethnicity.

In this paper, I focus on spillovers of local-level conflict across the African continent from 1997 to 2015. Employing data from the Armed Conflict Location and Event Data (ACLED) project, I analyze three different outcomes: Low-intensity *political conflict* such as rioting and protests, *violent conflict* and finally, *total number of conflict events*. I show that during this period, there was an important spatial autocorrelation component to the spread of conflict. When a given region experiences an intensification of conflict, so do its geographic neighbors. Spillover effects are even more powerful when the two regions are linked not by geography but by shared ethnic identities. Ethnicity is a powerful tool for the spillover of conflict even once geographic proximity is accounted for. All types of conflict demonstrate positive, significant spatial spillovers.

[1] These figures are estimated on the basis of data from the Uppsala Conflict Data Program (UCDP), 2016.

I show that political forces play a role in the diffusion of conflict. In particular, the more distant a region is from the national capital, the more insulated it is from the spread of conflict. Moreover, when a district gains greater access to central power, the entire neighborhood experiences greater political conflict in the form of protests and rioting.

Past studies of local spillovers include Harari and La Ferrara (2014), who focus upon how rainfall shocks generate waves of conflict at an extremely disaggregated (cell) level of analysis; and Konig *et al.* (2015), who analyze strategic dependencies in fighting networks during the Second Congo War. My unique contribution to the literature is to investigate the role of ethnicity and political power in the process of diffusion.

The organization of the paper is as follows. In the next section, I address spatial dependence in outcomes and describe different mechanisms through which it might occur. I also address what sort of districts might be particularly vulnerable to conflict spillovers. In section 3, I introduce the three main spatial models to be estimated. These are the Spatial Autoregressive Model, the Spatial Durbin Model and the Spatial Error Model. In section 4, I describe the data and methodology used in the paper. The presentation of estimation results from all of the different models follows in the next section. Robustness checks are performed in section 6, and conclusions follow in section 7.

2 Theoretical framework

2.1 Spatial dependence in outcomes

The first challenge in analyzing a complex phenomenon like diffusion is to clearly define and identify it. As Gleditsch (2007) writes, "We would like to know not only that conflict in Rwanda increases the risk of conflict in Zaire, but *what it is* about conflict in neighboring states that increases the risk of conflict in another state" (emphasis added). Are the outcomes of Rwanda and the Congo linked because agents travel and communicate across borders? Are actors in one state emulating strategies that they've seen successfully enacted in another? Or are agents in each country behaving in a completely autonomous way, as many past studies of conflict implicitly assume? A clear dichotomy arises between interaction and interdependence,

in which case it is appropriate to speak of diffusion; and independent outcomes, in which case it is not.

The clustered adoption of specific practices can occur under three distinct circumstances (Elkins and Simmons, 2005). First, clusters can reflect *coordinated effort* on the part of actors. When different actors engage in strategic interaction, their outcomes are then jointly determined. Many suicide bombings across the Middle East and elsewhere have been carried out by militants loyal to the Islamic State, a clear example of international cooperation towards a common end. The adoption of the euro is a more benign example. Coordinated effort may also arise from coercion. The universal adoption of burqas by women in the lands controlled by Islamic State, and more generally the mass religious conversions that have followed many territorial conquests, are both examples of clustering-by-coercion. In all cases, both actors are actively involved in transmitting contagion.

Second, clustered activity may arise due to *uncoordinated interdependence*: One transition- for example, the outbreak of war in one region- increases the probability that war will start in a second region, through *no conscious intention* on the part of the initiators. The two principal channels of this uncoordinated interdependence are *adaptation* and *learning* (Ibid). Groups adapt to changed circumstances when the initial adoption of a practice increases their own payoff from adopting it; they learn when the initial adoption imparts new information about the cost-benefit trade-off of such an action. When adaptation and learning occur, outcomes are correlated even when actors do not directly interact with each other. The diffusion process tends to be one-sided: the destination actor responds to a new situation which the actor at the origin has created.

Coordinated effort and uncoordinated interdependence both represent theoretically interesting situations of diffusion. In the first situation, the origin and destination actively collude; both actors are directly involved. In the uncoordinated case, the origin may not even be aware that contagion has occurred. The key feature that separates *absence of diffusion* from either scenario is that of *independence*. Groups with similar economic, political and cultural attributes may respond to *like shocks* in a *similar way*. Many of the risk factors associated with conflict, such as poverty and ethnic fractionalization, are clustered in certain geographic regions; moreover, such regions are vulnerable to the same sorts of shocks (Fowler, 2011). In such cases,

behavior appears to demonstrate dependence, yet there is no causal link in the spread of a behavior from one observation to another. Each region operates autonomously.

Evidence of diffusion may arise for atheoretical reasons. Spatial processes may operate on a larger scale than the single unit of observation. Administrative boundaries such as states and provinces often fail to correspond to the neighborhoods that give rise to the variables that we can observe. Moreover, latent, unobservable or unmeasurable features that explain variation in the outcome might be spatially autocorrelated. In both cases, observed relationships may be statistically significant, even in the absence of an underlying theoretical explanation (Buhaug and Gleditsch, 2008; LeSage, 2004).

The identification of diffusion requires us to distinguish between like regions responding to like circumstances, statistical anomalies, and true relationships of spatial dependence, which are characterized by interaction and interdependence. It can be empirically difficult to distinguish between the cases of coordinated and uncoordinated interdependence. We usually lack the information needed to establish clear and definitive ties between actors, or to impute intentions to their behavior. Few states would admit to purposely destabilizing inimical neighboring regimes by aiding rebel movements, for example. In the coordinated case, one agent [2]

In this paper I define diffusion as *spatial dependence in the outcomes of two connected districts.* It is a process linking together an initial stimulus in one region with behavior in another, through some sort of medium or conduit and through the choices of specific social agents (Solingen, 2012). Which mechanisms favor transmission, what characteristics render regions vulnerable, and how regions are linked together therefore jointly determine when spillovers will occur, and when they will not.

2.2 Spillover mechanisms

Conflict can spread through two types of channels, soft or hard. *Soft channels* produce changes in information and incentives, which in turn spark adaptation and learning. When

[2]Yet evidence of such aggressive strategies is widespread, especially in Africa. In 2012, the Hague's Special Court for Sierra Leone convicted former president Charles Taylor of Liberia of supplying and encouraging rebel atrocities in Sierra Leone, the most egregious example of this strategy. Throughout the 1980's, South Africa's apartheid regime worked to overthrow newly elected Robert Mugabe in Zimbabwe, contributing to his paranoid mistrust of Europeans and whites (Meredith, 2007).

actors observe a successful (failed) revolt abroad, they use this new information to update their priors about the probability of success and thus are more (less) prone to start a revolt in their own country. This is known as the demonstration effect.

Even as opposition movements observe events abroad, a second diffusion process may occur simultaneously: Authoritarian learning on the part of rulers (Lynch, 2014; Brownlee et al., 2015; Danneman and Ritter, 2014). Anticipating the shifting information available to their own opposition movements, regimes may engage in preemptive repression in order to avoid the same undesirable outcome of coup d'etat. The Arab spring uprisings and ensuing civil wars in Yemen, Syria and Libya are examples of simultaneous learning on the part of opposition movements and authoritarian regimes. Conflict can therefore spread to new destinations through the behavior of either side, the rulers and their allies or opposing forces.

The demonstration effect and authoritarian learning require no interaction between origin and destination agents: The destination must merely be aware of and able to observe the origin (the rationale behind severe restrictions on internet access in North Korea, see Zeller (2006). Yet a second soft channel- that of communication- requires the active participation of both origin and destination. Effective communication is essential to overcoming one central strategic challenge to collective action, that of coordinating potential participants. The expansion of social networks and mobile coverage has enabled protest movements to grow in Russia and Africa (Enikolopov *et al.*, 2016; Manacorda and Tesei, 2016); cellular communication has also aided counterinsurgency efforts in Iraq, by providing a safe and secure means for civilians to provide anonymous tips about insurgent violence (Shapiro and Weidmann, 2015). Communication between actors in different locations allows them to share information and coordinate their efforts on a larger scale, leading to interdependence between outcomes in the two locations. Such links therefore represent a second indirect channel of diffusion. Weidmann (2015) finds that communication networks aid the international spread of ethnic violence.

While soft channels are often unobservable, *hard channels* of diffusion are the highly visible fruits of warfare: The steady flow of fighters, arms and other materiel across administrative borders; the flood of humanity escaping out of conflict zones and into refugee camps. While the first type of flow is difficult to measure, a number of studies have found a significant link between refugees and diffusion (Salehyan and Gleditsch, 2006; Salehyan, 2007; Buhaug and

Gleditsch, 2008; Bohnet, 2012; Ruegger, 2013). Studies of migration due to environmental stress identify several important mechanisms: Migrants upset the existing balance of ethnic groups in the house country, heightening intergroup tensions; they also exacerbate competition for resources such as jobs and housing (Reuveny, 2007; Warnecke et al., 2010). Conflict refugees present additional challenges: They abound in grievances and have a low opportunity cost of fighting, and refugee camps often provide sanctuary to foreign rebels (increasing risk in country of origin as well). [3]

Hard channels are particularly important for the diffusion of violent conflict. Because transportation costs are increasing in distance, hard channels are more likely to operate at close geographic proximity, and are less important for diffusion across longer distances.

Soft channels are less constrained by distance. Social, political and cultural allies are likely to be especially susceptible to the soft channels. As we will see in the next section, collective identities provide a potential basis for group membership. There is likely to be greater communication between groups that share similar languages and history; on the basis of their shared identity, they may observe and emulate each other in a special way. Channels such as communication have been shown to be crucial for the transnational diffusion of ethnic conflict (Weidmann, 2015).

2.3 Who is at risk, and why?

Social agents are responsible for propagating diffusion. Yet not all types of agent are likely to do so. The characteristics of origin localities and of potential destinations, and the links that bind them together, are crucial in determining which districts are vulnerable and where conflict is likely to spring up next. Much of the foundational work in conflict spillovers modelled probability of outbreak in a target country as a function of own characteristics and those of neighboring countries (potential originators of conflict). The distribution of ethnic groups making up the population, whether these groups had kin abroad, political institutions, and type of conflict have all been shown to be of first-order importance.

[3] Cross-border rebel sanctuaries have played a role on both sides of the Rwandan civil war. The Rwandan Patriotic Front, composed of Tutsi refugees, invaded Rwanda from Uganda in 1990 and 1994; refugee camps in Kivu (eastern Congo) hosted many of the Hutu paramilitaries and government officials who carried out the 1994 genocide, and their presence contributed to the outbreak of war in Congo in 1998. See Gourevitch, 1998.

States exhibiting polarization- when a large ethnic minority confronts an ethnic majority- are at greater risk of contagion (Forsberg, 2008), although there is also support for fractionalization- a large number of small groups- as a risk factor (Gleditsch, 2007). Similarly, states are vulnerable when their domestic ethnic groups transcend national borders. The dangerous influence of transnational ethnic kin (TEK) groups is a strong result found across the literature. Such heightened risk of conflict exists even in the absence of contagion; the very existence of TEK threatens domestic stability, whether or not these TEK are engaged in conflict in their own states. The threat in increasing in the size TEK groups, but only up to a certain point; very large groups (such as the Russians or Chinese) are able to suppress rebellion among their kin in neighboring countries (Cederman *et al.*, 2009; Cederman *et al.*, 2013).

Access to power also matters. Dyadic and triadic analyses show that when a domestic ethnic minority gains access to central power, the probability of conflict decreases; downgrading or complete alienation power instead increases risk. Discriminated and separatist minorities pose the greatest threat of all. Moreover, TEK excluded from power in their own country are more dangerous for the domestic arena, perhaps because they have less of an incentive to maintain regional stability (*Ibid*). Similarly, separatist conflicts abroad- especially those involving TEK- are more destabilizing than are conflicts over central power (Buhaug and Gleditsch, 2008; Brathwaite, 2014). Analyzing Cold War era international disputes, Woodwell (2004) finds that shared ethnic majorities create conflict over group leadership and ideology, while ethnic groups that are a minority in one state and a majority are seen as a threat to state sovereignty and territorial integrity.

Many of the same risk factors for independent outbreaks of conflict- including poverty, ethnic divisiveness and instability- also place a country at risk of contagion. Yet controlling for domestic factors alone cannot fully capture this risk. As Gleditsch (2007) shows, "the likelihood of civil war in an extremely unfavorable region (is) several hundred percent higher than the risk of conflict in very favorable neighborhoods." Such assertions are clearly sensitive to one's definition of neighborhood.

Past studies have adopted the criteria of geographic contiguity, distance, and finally relationships based upon other characteristics. Danneman and Ritter (2013) analyze contagion at the national level, using spatial weight matrices based upon distance, religion and culture, and

political similarities.

While political institutions and policies are endogenous to conflict processes, ethnicity, language and religion are slow-moving variables that provide a strong contrast to contiguity-based measures of similarity. Moreover, such variables have economic and theoretical meaning. Spolaore and Wacziarg (2015) show that more closely related populations tend to go to war more frequently, due to similar preferences over rival goods and the lower cost of ruling over similar groups (creating an "economies-of-scale" incentive for territorial conquest that Alesina and Spolaore (1997) also address). These results are robust to redefining *relatedness* in religious, linguistic and cultural terms, which Spolaore and Wacziarg demonstrate in a 2016 paper to be both closely related and easily summarized by the genetic distance between populations. Ethnicity is therefore an objective characteristic with clear biological and historical foundations.[4]

This attribute shapes behavior because individuals derive utility not only from their own actions, but those of other group members. They enjoy benefits and bear costs that accrue to them personally, yet are at the same time sensitive to the power, status and reputation of the group as a whole (Akerlof, 2000). When group membership becomes part of the self, "events that harm or favor an in-group by definition harm or favor the self, and the self might thus experience emotions on behalf of the in-group" (van Stekelenburg and Klandermans, 2013). Not all group memberships have salience for the individual all the time. A distinction can thus be established between collective social identities that are still latent, and therefore separate from the self, and those that shape individual behavior. Sambanis and Shayo (2013) present a model of endogenous social identification in which individuals choose whether to align themselves with a national or subnational identity, as a function of group status and the permeability of group boundaries. This process of individual and social alignment- the *politicization* of identity- is a crucial step in transforming group membership into collective action: The riots, protests and violence under analysis here (Simon and Klandermans, 2001). Yet most models of diffusion have failed to allow for the degree to which self-identity and collective social identities coincide or contrast.

Politicized identities are a source of conflict when power structures are unclear or unstable

[4] Hirshleifer (1998) argues that kinship and group membership arose through evolution as a "returns to scale" in contending for ultimate biological goods such as access to food and reproduction.

(*Ibid*). Rationalist explanations for war posit that peace will break down only in situations of incomplete or asymmetric information. States uncertain about their opponent's ability or willingness to use force are compelled to use force themselves, as the bargaining that could fend off war becomes difficult if not impossible (Gartzke, 1999). As discussed above, refugee flows create uncertainty by upsetting a region's existing demographic balance. Shocks concentrated abroad, but involving ethnic kin- either their arrival to power in a neighboring state, or their engagement in a war- can be similarly disruptive. The politicization of identity means that such shocks can alter information and incentives, activating the soft channels of diffusion. This can occur even when the destination is not directly impacted by the original conflict; even when two states are geographically distant.

When transnational kin become embroiled in conflict, domestic ethnic groups may perceive a change to their status quo that encourages them to act. Rival groups may exercise preemptive force to stave off the future strengthening of their opponent (Danneman and Ritter, 2013). Each sort of shock creates uncertainty about the true strength of a group, a problem of asymmetric information that can lead to bargaining failure and the outbreak of conflict (Jackson and Morelli, 2009).

The importance of identity politics can hold true even when local-level conflicts, such as those analyzed here, have very little to do with macro-level cleavages. As Kalyvas (2003) argues, civil wars are the joint production of the strategic actions of central actors and the opportunistic actions of local ones. While the latter recruit and mobilize supporters, central actors inject extra muscle into local conflicts, creating a "concatenation of multiple, often disparate local cleavages, more or less loosely arranged around the master cleavage." Nor must local identities fully align with collective ones in order for these actors to benefit from group success (Hardin, 1995). Finally, targeted violence against civilians can render nonparticipation in a conflict so costly that civilians have little choice but to participate: An alternative solution to the collective action problem (Kalyvas and Kocher, 2007). In this way, violence contributes to the hardening of collective identities and master cleavages, as civilians seek refuge in those groups less likely to do them harm.

2.4 Hypotheses

As the existing body of literature demonstrates, conflict of all kinds spills across space. Such diffusion effects remains significant even when the spatial clustering of risk factors is accounted for. Mechanisms may be soft or hard, and different types of mechanisms are likely to favor different types of conflict. Originators of conflict, as well as targets, may be linked by geography or by shared ethnic communities. Whether these communities have access to power in either state is an important consideration. Finally, ethnicity represents an important motor of collective action, even when local actors are pursuing their own self-interest.

I focus upon six basic hypotheses that arise out of transnational literature, yet have never been tested on regional-level data. The first two contrast geographic and ethnic ties in their ability to serve as mediums for diffusion. The third addresses different types of conflict. The final two hypotheses address the role of political power and peripherality.

(H1) Local-level conflict processes display geographical contagion, even once the spatial clustering of risk factors is accounted for.

(H2) Ethnicity is an important channel for contagion, beyond the effect of geographic contiguity.

(H3) Diffusion of political conflict, like protests and demonstrations, is likely to be driven by different determinants than diffusion of violent conflict. In particular, geographic contiguity- which favors hard channels of diffusion- is likely to be more important for violent conflict.

(H4a) Districts that are isolated and peripheral will be more vulnerable to spillovers than will districts with access to political power.

(H4b) Conflict shocks originating from powerful districts, versus marginalized ones, will provoke a more powerful response.

3 Modelling contagion

When spatial dependence is present, OLS estimates are biased and inconsistent, as well as inefficient.[5]

As in time series analysis, the first step in spatial modelling is to test whether such an autoregressive approach is even appropriate, by analyzing residuals from an ordinary regression for residual spatial correlation. Moran's I and Geary's C are the two primary tests used to evaluate global and local spatial autocorrelation in data (Moran, 1950; Geary, 1954). When a data generating process exhibits significant autocorrelation, the fundamental tool of spatial econometrics to model this dependence is the spatial weight matrix, W. Like the adjacency matrix in graph theory, W represents links between the elements of a given system or network.

The key feature of the W matrix is its flexibility. By defining and redefining the matrix according to different criteria, it is possible to compare and contrast competing spatial hypotheses in an analytically rigorous way. As Corrado and Fingleton (2011) argue, "where spatial econometrics takes a lead is in its ability to identify and test theory relating to explicit spatial dependence mechanisms, as embodied in the parameterization of a W matrix." Neighborhoods defined by contiguity or by some limiting distance from the target (1 km, 750 km, etc.) enable us to test *direct* mechanisms. These are likely to coincide with the hard channels of diffusion discussed above. Refugees flee to to the nearest international border, in general. When sea, air and road networks are relatively underdeveloped- as in Africa- the transport of men and arms becomes more costly. For such channels, distance and nearness matter.

We can test indirect mechanisms by defining W in more imaginative ways. Religious neighbors share a common faith; ethnic neighbors, common culture and practices; political neighbors, common policy preferences. If these relationships remain significant even when contiguity and spatial distance been accounted for, then we have direct evidence of their validity in determining joint outcomes across space.

Once the W matrix has been defined, there are three main ways of modelling spatial autocorrelation. Each one serves a distinct purpose. By introducing lags of the dependent into the set

[5]When spatial lags in the dependent or in the regressors are significant (the SAR and Spatial Durbin models, discussed below), OLS estimates are biased; while in the SEM case of spatially correlated errors, the OLS estimator is merely inefficient (Elhorst, 2003).

of regressors, we can establish whether spatial diffusion of the outcome is actually occurring. Spatially lagged disturbances control for latent variables that are correlated across space and influence the dependent. Finally, lagged covariates allow for dependence in the outcome of any single region upon not only its own characteristics, but also those of its neighbors. Together, the complete framework of the three models provides three different spatial parameters (λ, γ and ρ), each governing the effect of a distinct type of spillover.

The most basic version of the spatial model posits first-degree autocorrelation in the dependent variable, in correspondence to the temporal AR(1) process. This is the Spatial Autoregressive Model, or SAR, sometimes referred to as the Spatial Lag Model. Given a dataset of N observations and an NxN spatial matrix, W, the SAR model is given by:

$$(I_N - \lambda W)\, y = X\beta + \varepsilon$$

$$y = \lambda W y + X\beta + \varepsilon$$

In reduced form:

$$y{=}(I_N - \lambda W)^{-1} X\beta + (I_N - \lambda W)^{-1}\varepsilon$$

where I_N is the NxN identity matrix, λ is the autoregressive parameter, and y is the vector of dependent variables. The dependent variable is a function of both observable characteristics $(X\beta)$ and unobservables (ε).

By lagging the dependent variable, the SAR model specifies spatial effects as a diffusion process. The model captures both strategic interaction and uncoordinated interdependence. A significant λparameter provides basic evidence that outcomes of neighboring units are jointly determined, and contagion is actually occurring (Sparks, 2013a; Elhorst, 2013). Rebellion in one region might cause neighboring regions to update their probabilities of success, create a new environment that neighbors must adapt to, or spark destabilizing flows of refugees or arms. All of these impacts would be captured by the SAR parameter. As λ rises, spatial effects grow more persistent, approaching a limiting "spatial random walk" scenario in which a shock to the outcome of a single district propagates through the entire system.

The Spatial Durbin Model (SDM) is an extension of the SAR. In this case, the observable characteristics of neighboring districts into the information set. Estimates for the SDM model

are presented in the Appendix.

As a robustness check I estimate the SAR in combination with a third model of spatial dependence: The Spatial Error Model (SEM). In the SEM, the error term is assumed to be spatially correlated:

$$y = X\beta + u$$

$$u = \rho W u + \varepsilon$$

$$y = X\beta + (I_N - \rho W)^{-1}\varepsilon$$

The error term has two components, the spatially autocorrelated u and the random, homoskedastic ε. The spatial error parameterρ measures the strength of spillovers among the model residuals of neighboring observations (Sparks, 2013b).

The SEM framework allows us to control for latent variables that are unobservable or difficult to quantify, and that might be correlated with conflict. The basic units of analysis- the administration 1 districts- are of different shapes and sizes, and their borders are possibly endogenous. The same holds true for the homelands of different ethnic groups. Moreover, the spatial process under examination may not coincide with any of these borders. Finally, unobservable idiosyncratic shocks might follow a spatial diffusion process (Elhorst, 2013; Millo, 2014). All of these possibilities give rise to spatial correlation that is distinct from the dependence in outcomes described by a SAR process. Inclusion of the SEM component allows us to distinguish between the effects of th two types of spatial dependence.

By combining the SAR and SEM models, we can analyze whether spatial diffusion in outcomes (as represented by the SAR parameter λ) persists, even after controlling for the influence of spatially autocorrelated unobservables (represented by the SEM parameter ρ). If violence breaks out first in region A and then in neighboring region B, it may appear that B is emulating A, a SAR effect. However, a shock such as plant disease, extreme weather, or a rise in brigandry- all of which would be unobservable to the outside observer- might have first impacted A and then B, threatening livelihoods and lowering the opportunity cost of violence. In introducing the spatially correlated error term, we discern whether true spatial diffusion is at work: Thus avoiding the trap of similar localities responding to similar circumstances.

This type of combined analysis is commonly known as the SAC model. Given a panel dataset of N observations and T time periods, the SAC model with k covariates is given by:

$$y = \lambda (I_T \otimes W) y + X\beta + u$$

where y is an NTx1 matrix of observed dependents, I_T is the TxT identity matrix, X is the NTxk matrix of covariates and u is the NTx1 vector of spatially autocorrelated residuals. As before, $\lambda$is the SAR parameter and β is the kx1 coefficient vector.

The model with individual effects has a two-part disturbance vector:

$$u = (\iota_T \otimes \mu) + \varepsilon$$

$$\varepsilon = \rho (I_T \otimes W) \varepsilon + \nu$$

where ι_T is a Tx1 vector of ones and μ is the Nx1 vector of time-invariant individual effects. As before, ρ is the SEM parameter. The vector of spatially autocorrelated idiosyncratic errors is given by ε, while $\nu_{it} \sim IID(0, \sigma_\nu^2)$ by hypothesis.

4 Data and methodology

Table 1 presents district-year descriptive statistics for the main dependent and independent variables, described below. Over 18 years, 764 districts were observed, for a total of 13,752 district-year observations.

4.1 Dependent variables

The outcome of interest is conflict event occurrences in Africa from 1997-2015, taken from the Armed Conflict Location and Event Data Project (ACLED). ACLED dataset observations are drawn from news, international organization, NGO and security reports. Each observation includes type of event, actors involved, number of fatalities and geographic coordinates of the location of the event.

The ACLED represents the most complete set of conflict data available worldwide. However, it has several limitations. One arises from the nature of the sources. Events that are not

reported upon are excluded from the dataset. Second, if many actors are involved in an event, some may be inadvertently excluded from the dataset. Third, the classification of an event is necessarily subjective. Some events may fall into multiple categories.

The dataset creators address these weaknesses by requiring three reviewers to analyze each conflict event. Moreover, data is published immediately on processing, allowing the public to report possible errors. By including year and country fixed effects, I can correct for unobservable variation in reporting across time and across countries. However, I cannot correct for unobservable factors that vary by country and year.

I create three conflict variables based upon the ACLED data. The first dependent variable, n_events_{it}, is total number of conflict events of any type that occurred in district i in year t. *Political conflict* is defined as the forms of conflict characteristic of most parliamentary democracies. The two types of event included in political conflict are protests and rioting. These events are lower in intensity and non-violent in nature, in general.[6] *Violent conflict* includes all of the remaining categories: "Violence against civilians," "Battle-No change of territory," "Battle- government regains territory," "Battle- Non-state actor overtakes territory" and "Remote violence."

The third hypothesis is that political and violent conflict events are driven by different types of shocks. By disaggregating conflict occurrences in this way, I test whether different mechanisms are indeed at play, and whether one type of conflict is more easily transmitted.

By utilizing counts as the outcome, rather than a simple dummy for conflict onset- as most past studies have done- I focus on the *prevalence* of conflict in the population of African districts, rather than *incidence* or onset. Such an approach allows us to exploit variation on the intensive margin. A count variable allows us to distinguish, for example, between a few contained instances of unrest, and a full-scale civil war that generates numerous casualties and massive refugee flows. Moreover, districts that have experienced conflict in recent years- and are thus already counted as "infected" in the incidence sense- may undergo drastic changes in terms of level and type of conflict, due to spillovers. A prevalence approach allows us to

[6] Riots and protests resulted in 0.32 fatalities, on average, with a maximum value of 631 (rioting between Christians and Muslims in Nigeria in 2008). Eight fatalities occurred per violent conflict event, on average, with a maximum value of 25,000: in 1997, the Alliance of Democratic Forces for the Liberation of Congo-Zaire massacred Hutu refugees from Rwanda.

exploit to the fullest extent all of the information available in our dataset.

4.2 Weight matrices

The first hypothesis is that geographically contiguous regions are subject to the diffusion of conflict events; the second, that regions related by ethnicity experience diffusion. I propose to test hypotheses (1) and (2) through the construction of for different spatial weight matrices, G , E, G_R and E_R. The first matrix contains information on *geographic contiguity*: $w_{i,j} = 1$ if districts i and j are neighbors, and 0 if they are not. The second matrix defines neighborhoods on the basis of *ethnic contiguity*: $w_{i,j} = 1$ if the same ethnic group lives in both districts, and 0 if there are no shared ethnic groups.

The final two matrices isolate the effects of geography and ethnicity by purging relationships based upon the other variable. The *residual geographic contiguity* matrix defines a neighborhood of regions sharing geographic but not ethnic ties. This matrix is derived as $G_R = G - E$. The *residual ethnicity matrix* defines the neighborhood of non-contiguous regions sharing ethnic ties, and is derived as $E_R = E - G$.

These four weight matrices can also be used to test hypothesis (3), that due to the higher cost of transporting fighters and resources across long distances, geographical contiguity should be more important to the diffusion of violent conflict. If this hypothesis holds true, than the estimated SAR parameters for the violence model should be larger and more significant when neighborhoods are defined by geography rather than by the less clearly-defined force of shared ethnicity.

Figure 1 maps out the neighborhoods defined by each of the weight matrices. The ethnic contiguity matrix (in grey) is a slightly sparser version of the geographic contiguity matrix (in green), while the residual contiguity matrix (in purple) is sparser still. The residual ethnicity matrix (not pictured) is quite similar to the ethnicity one.

The weight matrices are defined based upon the 768 administrative areas of Africa contained in the Global Administrative Areas (GADM) database. These are the basic units of observation. They are the largest sub-national districts in each country, corresponding to states, provinces and governorates. Both their size and shape vary widely from country to country. Burundi,

one of the smallest countries in Africa, has 17 such regions; while the DRC, one of the largest, has only nine. In order to account for this heterogeneity, I employ the Spatial Error Model with district-level fixed effects.

Four of the 768 regions have no direct neighbors, and are excluded from the dataset. The geographic contiguity matrix G therefore has dimensions of 764x764. However, spillovers may occur between these regions because they tend to share a large number of ethnic groups- not because they are near in the physical sense. To isolate the role of geography, the residual contiguity matrix G_R excludes 172 more regions to attain dimensions of 592x592. The SAR, Durbin models and employing the G and G_R matrices provide the evidence we need to test the first hypothesis.

The ethnic contiguity matrix, E, is based on the Geo-referencing Ethnic Power Relations (GeoEPR) dataset. The GeoEPR is a spatial polygon dataframe the delineates the settlement areas of ethnic groups across the world from 1946 to 2013. International borders play an important role in the composition of the GeoEPR dataset. For example, Hutus in Rwanda, Burundi and the DRC are counted as three distinct ethnic groups. There are two defining features of the GeoEPR dataset. First, it maps *politically relevant* ethnicities: Groups which have sought political representation at the national level, or been discriminated against by the state. Second, the GeoEPR defines ethnicity broadly, in terms of language, religion, ancestry, culture and phenotype (Vogt et al., 2015). In order to identify common ethnic identities that cross borders, I merge the GeoEPR with the Transnational Ethnic Kin (TEK) dataset, which is constructed from the same source data as the GeoEPR, using the same methodology. This dataset matches each national sub-grouping with a larger international group, should one exist.

The ethnic contiguity matrix built from these two sources is 699x699. The E weight matrix provides a basis for testing hypothesis (2), but is not sufficient for identification of the true effect of ethnicity. Because most ethnic groups are scattered across a close cluster of districts, parameter estimates might be capturing relationships of direct contiguity, rather than shared preferences and behaviors based upon a shared identity. The residual ethnicity matrix excludes four regions and has the dimensions 694x694.

All four types of neighborhoods provide the basis for evaluating hypothesis (3), that the mechanisms driving the spread of riots and protests and those causing violent conflict to

spread are different. In particular, G and G_R should be more fundamental to violent conflict; E and E_R, to riots and protests.

Table 3 summarizes the four weight matrices. As the weight matrix becomes progressively smaller, the proportion of nonzero links rises, while both the average and maximum number of links per district fall. Under all four regimes, the most connected region is Kivu in eastern Congo: one of the most conflict-prone places in the entire world.

In the analysis, the row-stochastic form (each row normalized to sum to one) of each weight matrix is used. Normalization helps to ensure invertibility of the term $(I_N - \rho W)$- where ρ may be replaced by any of the three spatial weight parameters), and positive-definiteness of the variance-covariance matrix (LeSage, 2008).

4.3 Independent variables

The final two hypotheses address the impact of political factors upon the diffusion process. Hypothesis (4a) is that conflict shocks originating from districts with political power will be less vulnerable to conflict than will isolated or ostracized districts. Hypothesis (4b) posits that conflict shocks originating from powerful districts will provoke a more powerful response in terms of conflict spillovers. The spatial matrix is not a suitable approach in this case. Ideological affiliations and types of political systems tend to be fluid and fast-changing. The spatial weight matrix is instead assumed to be exogenous and constant over the entire period of study. In order to evaluate these hypotheses, I introduce explanatory variables and interaction terms into regression equation that identify powerful and peripheral districts in a precise way.

The first two indicators measure a district's *access to central power*. The $capital_{it}$ variable equals 1 if the district hosts the national capital, and 0 otherwise. The $access_{it}$ variable equals 1 if either the current head of government or head of state was born in the district. Place of birth of the two most powerful state officials has economic and political significance because patronage networks are often targeted to benefit the specific ethnic group and region of origin of leaders (De Juan and Bank, 2015).[7] In Zimbabwe, for example, Robert Mugabe's

[7] Francois et al. (2015) finds that the extent of patronage is limited by the risk of outside rebellion, resulting in a fairly equal distribution of cabinet shares across different ethnic groups;. However, the country leader's ethnic group always receives a power premium.

own ethnic group- the Shona- enjoy special redistributive privileges, while the homeland of Zimbabwe's second largest ethnic group, Matabeleland, has been systematically ignored and neglected (Meredith, 2007).

The next indicators instead address *challenges to central power*. The $outside_{it}$ variable indicates whether the current head of state or government of a foreign nation was born in district i. This is a common occurrence due to the shift in international borders that followed decolonization in Africa, and is potentially indicative of a district with outside loyalties (though not necessarily one isolated from the government of its own state). For example, both presidents of Djibouti have been foreign-born, in Somalia and Ethiopia.

The $distance_{it}$ variable instead measures the distance (in km) of each district centroid from the national capital. The state's ability to exercise power is often deteriorating in this distance. As this measure grows, districts become more and more isolated from the apparatus of power, and less integrated into the nation as a whole. It therefore indicates peripheral and marginalized districts, districts with a greater innate potential to challenge state power or even attempt to break away.

One drawback to the $distance_{it}$ and $capital_{it}$ indicators is that they present little variation over time. Over the entire period studied, two countries- Nigeria and Tanzania- moved their national capitals (in 1991 and 1996). Two more- Eritrea and South Sudan- became independent (in 1993 and 2011), which also registers as a transfer of the capital. [8]

By interacting the conflict variables with $capital_{it}$ (or $access_{it}$) on the one hand, and $outside_{it}$ and $distance_{it}$ on the other, we can precisely distinguish between those shocks originating from regions with access to, or peripheral from, central power. This distinction is important for addressing hypothesis (4b), that is that shocks coming from powerful districts and from peripheral ones will provoke a different response in terms of conflict behavior.

The set of capitals is taken from the *CIA World Factbook* (2016); the Latitude.to website, which calculates the GPS coordinates of any address in the world, provided their precise location. The two top politicians of each state are taken from the *World Factbook* and from

[8] Sovereignty over the capital of Western Sahara, Laayoune, is claimed by both Morocco and the Polisario Front, and the day-to-day activities of the independence movement are generally directed from refugee camps located elsewhere. Because of its lack of sovereignty, Western Sahara is counted as having no capital at all in the dataset. See the *CIA World Factbook* (2016).

the *World Statesmen* website, an online encyclopedia that catalogs the leaders of states and territories. Data on the districts of origin of politicians was obtained from various sources, including the websites of national parliaments, which often contain detailed biographies of MPs; international press agencies; and online newspapers.

4.4 Control variables

The control variables are nightlight data, population, population density, rainfall and deviations from average rainfall. These account for variation in how large and sophisticated locals economies are, and the opportunity cost of fighting.

Nightlight satellite data has been used as a proxy for local levels of development in numerous studies (Henderson et al., 2012; Bickenbach et al., 2013; Mellander, et al., 2013; Nordhaus and Chen, 2014). Shortland (2013) uses variation in nighlight output to gauge the impact of Somalia's civil war upon local economic development; De Juan and Bank (2015), as a proxy for selective redistribution. For this study, nightlight data is taken from NOAA's Version 4 DMSP-OLS Nighttime Lights Time Series. The Average Visible, Stable Lights, & Cloud Free Coverages dataset was used. The yearly average of nightlight data was calculated for each district.

Because nightlight satellite data may simply be proxying urbanization, I also include controls for district population and population density. These data are taken from NASA's Socioeconomic Data and Applications Center (SEDAC). This data is only available at five-year intervals. Data for 2005 and 2010 are taken from the Gridded Population of the World, v4 dataset, while data for 1995 and 2000 are taken from the Global Rural-Urban Mapping Project (GRUMP), v1. Observations from 1997 to 2000 are assigned the 1995 values; from 2001 to 2004, the 2000 values, and so forth. In all cases, the UN-adjusted population counts are used. The original population density variable is persons/km; due to large variability in the size of districts, I rescale by dividing by 1,000. I rescale the population variable by dividing by 100,000.

Many recent papers employ rainfall data from satellites and gauges to demonstrate how climatic volatility influences local livelihoods and hence, the risk that regions will be struck by violence. Rainfall measures are particularly suited to explaining conflict incidence in sub-

Saharan Africa: The one region of the world where a majority of livelihoods still depend upon rain-fed agriculture. Environmental instability plays a much weaker role elsewhere (Miguel; Gleditsch, 2012).

Because extreme deviations- positive and negative- from average monthly rainfall levels are more important to explaining conflict than are rainfall levels, I employ the former as a control in all specifications. Both measures are calculated on the basis of data from the CPC Merged Analysis of Precipitation (CMAP), available from the National Oceanic and Atmospheric Administration (NOAA) of the U.S. Department of Commerce.

4.5 Estimation and identification challenges

Individual fixed effects, μ_i, are introduced into all three models in order to control for time-invariant heterogeneity among the observed regions. When units of observation are irregular in shape and size, as in our case, the random effects assumption is inappropriate (Elhorst, 2003). Hausman tests to compare the two models confirm that the random effects assumption is inappropriate for this dataset; therefore, all models are estimated with fixed effects.

The F.E. model is unbiased if there are no unobservables that vary across both time and space. I introduce year dummies to control for time-varying influences that are constant across individuals. A number of specification tests- including Wooldridge's test for unobserved effects in panel models, the Lagrange multiplier tests, and the F test- confirm that the correctly specified model requires both individual and time fixed effects.

The estimation process of the fixed effects spatial panel model is a two-step iterative procedure, alternating between GLS for the nuisance parameters (β and σ_ν^2) and maximization of the concentrated likelihood for the parameters of interest (the three spatial parameters, as well as the variance ratio $\phi = \sigma_\mu^2/\sigma_\varepsilon^2$). These two steps- GLS and ML- are repeated until a convergence criteria for parameter estimates is satisfied. See Millo and Piras (2012) and Croissant and Millo (2008) for further details on the estimation procedure used by plm and splm, the two packages in R dedicated to panel data and spatial panel data analysis.

Maximum likelihood estimation of the F.E. spatial panel model imposes a number of restrictions on the stochastic term, including normality, homoskedasticity, and the absence of serial corre-

lation. Several of the assumptions prove to fail in this dataset. Applying the Breusch-Pagan test, the null hypothesis of homoskedasticity is rejected for several of the 18 cross-sections. Moreover, the Breusch-Godfrey test and the Wooldridge test for AR(1) errors, as well as the the Baltagi, Song, Jung and Koh LM test (specific to spatial panels) reveal the error terms to be serially correlated. The within transformation required by fixed effects leads to residuals with negative serial autocorrelation, but in this case the dependence applies to the original data as well.

One solution to serial correlation is to introduce time-lags of the dependent. However, when the fixed effects transformation is applied this leads to correlation between the regressors and the error term, leading to a bias (the "Nickell Bias") in the estimate of the SAR parameter of order 1/T (Nickell, 1981). For my panel dataset (18 years long) the Nickell Bias is of the order 0.06. Both the time and spatial lag coefficients are impacted. However, coefficient estimates are biased downwards. Therefore, specifications that include time-lagged dependents can be interpreted as a useful lower bound on the true effect.

When serial correlation and heteroskedasticity is present, standard errors tend to underestimate the true standard deviation, and therefore cannot be used a basis for statistical inference. Serial correlation also leads to inconsistent estimates in the SAR model. I allow for both serial and spatial autocorrelation in the disturbances of all spatial specifications. The simplified framework discussed above assumed that the non-spatially autocorrelated part of the error followed the distribution $\nu_{it} \sim IID\left(0, \sigma_{\nu}^{2}\right)$. In this paper, the ν are instead taken be serially autocorrelated. The complete model for the disturbance, presented in Millo (2013), has an individual, time-invariant component μ (as before) and an idiosyncratic component that is both spatially and serially autocorrelated. The disturbance vector has three parts:

$$u = (\iota_T \otimes \mu) + \varepsilon$$

$$\varepsilon = \rho\left(I_T \otimes W\right)\varepsilon + \nu$$

$$\nu_t = \psi\nu_{t-1} + e_t$$

The new parameter ψ captures serial correlation.

All models presented in the paper include year and country fixed effects. Standard errors are clustered at the district level.

4.5.1 The reflection problem

Employing the spatial lag of conflict as a dependent variables, as all of the models presented here do, presents a special challenge to identification. The reflection problem, introduced by Manski (1993), closely relates to the discussion of clustered risk factors versus true interdependence. When districts in a neighborhood engage in similar behavior, does this reflect the dependence of individual behavior upon the propensity of neighbors to adopt the same action (endogenous effects)? Or are neighboring districts acting independently, yet responding to similar circumstances in a similar way (correlated effects)? To paraphrase Manksi, when we observe both a district and its neighbor to engage in conflict simultaneously, is the influence of the neighbor responsible for the district's behavior, is the neighbor reflecting the district, or are both district and neighbor responding to a common shock?[9]

The ability to distinguish between correlated yet independent behavior of neighbors, and true spillover effects, depends upon two key assumptions. First, the functional form must be correctly specified. Second, the spatial weight matrix must reflect the true pattern of dependence in the network. When these assumptions are satisfied, the spillover effect of conflict in one region upon conflict in another is identified and maximum likelihood provides a consistent estimate.

5 Results

In this section, I present estimation results. OLS estimates of the panel data model are discussed in section 5.1, while the SAR model, estimated with ML, is discussed in 5.2. In the following section, I introduce political interactions into the SAR model in order to better evaluate the impact of these variables.

[9] "Does the mirror cause the person's movements, does the image reflect the person's movements, or do the person and the image move together in response to a common external stimulus?' (Manski, 1996).

5.1 OLS estimates of the non-spatial model

The OLS estimates provide a baseline for comparison with the later spatial models. Table 3 presents coefficient estimates for the number of events, political conflict and violence models. In column (1), I regress events on four control variables: the nightlight variable, population, population density and deviations from average monthly rainfall. The four political indicators- *capital*, *access*, *outside*, and *distance*- are also included. In column(2), time-lagged events are introduced. As discussed above, this coefficient will be smaller in magnitude than the true value. In column (3), I regress riots on the set of basic controls, introducing time-lagged riots in column (4). The two violence regressions are presented in columns (4) and (5). All models are estimated using the 764x764 geographic contiguity dataset.

The population and the nightlights indicators are positively associated with all types of conflict: Larger, more economically developed regions experience more riots, more violence, and greater conflict overall. Excessive rainfall and flooding also leads to more conflict, but loses its significance when the lagged dependent is introduced. Finally, the coefficient on the time-lagged dependent is large and highly significant in all cases.

Distance is the most important of all the political variables. As the distance between the district and the national capital grows, the level of violence and overall conflict it experiences goes down; distance however has no impact on political conflict. Access to outside power leads to less violence (significant at the 10% level). Finally, access to central power leads to more conflict events of any type (significant at the 10% level). The capital dummy is not significant in any regression.

Much as traditional specification tests evaluate whether the Gauss-Markov assumptions are met, OLS residuals can be also analyzed for spatial autocorrelation. Moran's I varies between -1 (negative autocorrelation, so that like observations of the dependent repel each other) and +1 (positive autocorrelation, so that like observations cluster together). For each of the 18 cross-sections that make up the panal dataset, the residuals from specification (1) tested positive for spatial correlation, with an average Moran's I of 0.228 and an average p-value of 0.001.

Figure 2 maps out the residuals from the estimation of (2) on the entire panel. This map

provides clear additional evidence of spatial clustering. Broad swaths of the continent display negative residuals, in blue. In green are regions for which the model underestimates violence, resulting in positive residuals: these include much of the Sahara, coastal Egypt and Sudan, the Horn of Africa and Madagascar.

5.2 SAR model estimates

Regression results for the SAR model are presented in Table 4. The first coefficient average impact of spatial lags in the dependent upon a district's own outcome. It is positive and significant at the 1% level across all specifications.

Panel A presents estimates based upon the contiguity weights matrix, G. Rioting and protests display the greatest spatial autocorrelation of any conflict event, although the magnitude of the coefficient falls when the time-lag is introduced (a conservative underestimate, given the Nickell Bias). This holds true for all types of conflict. Spatial autocorrelation is much lower in the residual contiguity model (panel B), suggesting that the estimated geographic network effect was confounded by shared ethnic ties.

Ethnicity is an equally important source of conflict spillovers, as shown in panel C. Regions that share ethnic groups are particularly vulnerable to the spillover of rioting and protests. Even when ethnicity is purged of the effect of direct geographical ties (panel D), riots display a high degree of spatial autocorrelation.

The estimation results of Table 4 support the first two hypotheses. Both geography and ethnicity are important channels for the diffusion of conflict. Observed patterns of violence across districts are not independent, but the outbreak in one district actually discourages its spread elsewhere. When conflict intensifies in one district, there is a positive neighborhood effect.

Hypothesis (3) posited that spillovers of violent conflict would be more significant for those neighborhoods defined by geographical contiguity. This is not what we find. We find, in fact, that violence spreads more easily through ethnic networks. Ethnicity is more important than geography for the diffusion of all types of conflict. Yet its impact is particularly strong when we consider the sort of conflict prone to spread through soft, indirect channels: Riots and

protests.

The political violence SAR coefficient for the neighborhood of districts defined by ethnicity, equal to 0.62, has two possible interpretations. First, when the average number of riots/ protests in an ethnically-linked districts increases by one, the impact on the single district will be an increase of 0.62. Second, when the number of riots/ protests in a district *i* increases by one, the average district *j* will undergo a shock of 0.62/N, where N is the number of *j*'s ethnic neighbors. In the same panel, we find that when the weighted average of violent events increases marginally in an ethnic neighborhood, the average district undergoes a violence shock equal to 0.41.

Estimation results relative to the contiguity and residual contiguity SAR models with political indicators are presented in Table 5a; relative to the ethnicity and residual ethnicity models, in Table 5b. The addition of political variables does not modify the magnitude of the SAR coefficients; nor does their statistic significance fall. In the contiguity models, distance from the capital is associated with a slight reduction in violent events and in overall conflicts, much as we found in the OLS model with political variables. This effect is significant only in the regressions with time-lagged dependents, however.

Distance exercises a negative impact upon conflict intensity, regardless of how the network is defined. The result holds for both violent conflict and overall number of conflict events, but not for political conflict; distance from the capital is orthogonal to the outbreak of riots and protesting.

There is weak evidence that access to outside power increases political violence (in the ethnicity model), while decreasing violence (in the residual ethnicity model). But this result is not robust.

5.3 At the intersection of conflict and political power

The SAR model provides mixed evidence for hypotheses (4a) and (4b). There is little evidence that more peripheral regions are more vulnerable to conflict; in fact, the opposite appears to hold true, as regions more distant from the capital experience lower levels of conflict. Conflict in capital regions does not appear to generate strong spillovers; nor is access to power in an outside district of particular importance. However, The Durbin model results suggest

that spillovers from districts with access to central power may be of some importance to the frequency of political violence (Appendix).

In this section, I introduce interactions between the political variables and time-lagged conflict outcomes. In Tables 6a and 6b, each outcome is interacted with the capital dummy, access to central power, access to outside power, and distance from the capital. The dependent variable in each case is number of conflict events. The capital coefficient is positive and highly significant across all three specifications in which it is included. The interaction with rioting is negative and significant at the 10% level. Access to central power and its interactions are not significant.

Although the outside dummy is not significant, each of its interactions is negative and significant: Obtaining access to outside power- in terms of a native inhabitant of the region acceding to rule in a foreign country- depresses number of conflict events in the following year, when this access to outside power is paired with an increase in conflict in the current year.

Finally, distance from the capital is consistently negative and highly significant, as previous regressions have shown. However, an increase in distance from the capital does not depress future conflict when it is combined with a higher frequency of conflict in the current year.

Political violence is the outcome in Tables 6c and 6d. Only the interaction between political violence and distance from the capital is significant (and negative).

Finally, violence is the outcome in Tables 6e and 6f. There is a strong positive association between the capital dummy and a violence, though none of the interaction terms are significant. All of the outside interactions are negative and significant, and once again distance is negative and significant. Thus the violence estimates mirror those obtained for overall conflict.

The political interactions suggest that the capital district is uniquely vulnerable. While conflict shocks originating from the capital do not seem to significantly influence conflict, it is regions most exposed to central power- not the most isolated and peripheral- that are subject to the greatest amount of conflict. The farther a district is from the central capital, the less vulnerable it is to conflict. Finally, access to power in a foreign country, combined with past experiences of conflict, appear to depress future levels of conflict.

6 Robustness Checks

In this section, I perform several robustness checks. First, I allow for spatial autocorrelation of the error component, estimating a combined SAR-SEM (SAC) model. This model allows us to account for latent variables that follow a spatial diffusion process and are correlated with conflict. Second, I restrict each network to either internal and external neighbors, in order to isolate how conflict spreads within individual countries and across international borders.

The SAC model results are presented for the contiguity and residual contiguity networks in Table 7a, and for the ethnicity and residual ethnicity networks in Table 7b. In each case, the SAR parameters for number of events and for violence grow drastically in magnitude, while the SAR parameter for political conflict grows much smaller. Spillover effects remain strongest for those neighborhoods defined by ethnicity. Finally, the SAR parameter becomes very small and loses its significance in those specifications that include the time-lagged dependent; yet given the downward Nickell Bias, these results should be taken with a grain of salt.

The SAR parameter appears to become smaller in the case of political conflict because a large amount of positive spatial autocorrelation is picked up the SEM parameter. Conversely, the estimated effects for number of events and violence grow stronger because they are no longer offset by the large, negative impact of the latent variables. Overall conflict and violent incidents, on the one hand and political conflict on the other appear to be driven by distinct spatial processes. While overall number of events and violence demonstrate strong diffusion effects, much of the spatial correlation in political conflict may in fact be driven by unobservable spatially correlated factors.

One more result of the combined SAR-SEM analysis is noteworthy. In each previous analysis, parameter estimates grew smaller in magnitude when the residual contiguity and residual ethnicity networks were considered. The result still holds for the contiguity and residual contiguity models. When we compare the ethnicity and residual ethnicity results, however, we find that two estimates grow in magnitude: Those relative to political conflict (without a time lag) and violent conflict (with time lag). When unobservable, spatially correlated factors are taken into consideration, these two types of conflict propagate more readily through the network of districts connected only by ethnicity but not geography.

Tables 8a-8d present the next set of robustness check results. Each network is restricted to the set of internal (within the same state) and external (across different states) neighbors. Internal and external contiguity results are presented in Table 8a, residual contiguity results in 8b, ethnicity results in 8c and residual ethnicity results in 8d.

The contiguity and residual contiguity networks continue to display significant spatial autocorrelation when the dataset is restricted to internal neighbors. The networks of cross-border geographically contiguous neighbors do not, however. It is clear that the pattern of geographical diffusion found up to now was almost entirely driven by internal neighbors, not by conflict spilling over international borders.

However, the external contiguity results provide new information about the role of political variables in determining conflict. Within the network of external neighbors, the capital district and access to outside power are both positively associated with conflict.The farther a district is located from the capital, the less conflict it experiences, as before (a result that also holds for the network of internal neighbors).

Neighborhoods defined by internal ethnic and residual internal ethnic links continue to demonstrate strong spatial autocorrelation. The two external ethnicity SAR parameters are extremely small and significant only for political conflict and violence; in the case of political conflict, I find evidence of negative spillovers. In parallel to the contiguity results, the positive spatial correlation of conflict across ethnic neighbors appears driven entirely by within-country dynamics. Finally, distance from the capital continues to lower the intensity of conflict experienced by a district in terms of violent conflict and overall number of events.

7 Conclusions

The evidence presented in this paper supports two conclusions. First, capital districts are exposed to greater amounts of conflict. The farther a district is from the capital, the less vulnerable it is to spillovers. However, it is not clear that conflict shocks originating from the capital are more dangerous to neighbors.

Second, both geographic and ethnic ties are important conduits for diffusion, but ethnicity is

a more powerful force than geography.

Ethnicity appears to convey not only a shared sense of identity, but moreover new information about the desirability of engaging in conflict, and particularly in political conflict in the form of rioting and protests. A possible realignment of incentives is occurring. While refugee flows are by far the most visible and contentious consequence of modern warfare, these flows are frequently directed at the nearest international border. The degree to which refugees, combatants and arms succeed in finding their way to noncontiguous regions inhabited by ethnic kin is an open question. It is clear that neighborhoods defined by ethnicity, as opposed to geography, are engaging in an altogether different sort of calculus. Conflict spillovers are transforming these neighborhoods in profound and far-reaching ways. Spillovers from ethnic neighbors are- for better or for worse- destabilizing existing political systems,and remaking societies; remaking, even, how societies see themselves.

8 Appendix

The Spatial Durbin Model (SDM) introduces the observable characteristics of neighboring districts into the information set. Actors do not behave in isolation, whether they are individuals, districts or entire nations. In undertaking decisions, actors will always account for the larger strategic setting in which they are embedded. Unlike the SAR model, which implicitly assumes that only shocks to neighboring dependents matter, or traditional models that account only for "own" attributes, the Durbin model allows for the attributes of others to directly enter into an individual region's decision-making process. Spatial Durbin estimates are of particular value because they result unbiased even when the true data generating process is a SAR or SEM.

The spatial Durbin model is given by:

$$y = y + X\beta + WX\gamma + u$$

where WX is the vector of spatially lagged covariates and γ is the Durbin parameter vector. The coefficient γ_r measures the average impact of a change in variable x_{ir} upon the outcome of region j, $i \neq j$. This parameter reflects indirect (neighborhood) effects, while traditional regression coefficients capture direct effects (LeSage, 2008; Glass et al., 2012). However, the situation is complicated by the fact that the indirect effects of any type of shock will eventually lead back to the district of origin itself. In evaluating the total impact (direct + indirect) of a variable, it is necessary to account for the entire cycle of spillovers and feedback effects that an initial change can generate.

Suppose a politician from region A becomes president. Suddenly, region A has much greater power and influence on the national level. Levels of violence may fall in A, as residents reap the benefits of a shift in the geographic distribution of selective redistribution in favor of their own region. As residents travel outside the region to buy and sell goods, or send remittances, this new wealth spreads to other regions, potentially lowering violence there, as well. At the same time, region B has lost access to the privileges of executive power. As B suffers from lower levels of patronage, violence may rise there and in neighboring regions, which are also negatively affected. If A and B are neighbors, the situation becomes yet more complicated. In each case, feedback effects are generated, which sometimes operate in a different direction from the initial, direct impact of the change in x_{ir}. The average total impact measure reflects

the long-run equilibrium effect of such a change upon the entire network. Evaluation of the direct, indirect, and long-run impact of shocks to the dependent is also possible in the SAR model.

8.1 Spatial Durbin model estimates

Further evidence for the impact of the political variables is provided by the spatial Durbin models (Table 10a-10d). Each of the four tables presents Durbin estimates for a different set of neighborhoods: Contiguity and residual contiguity in 10a and 10b, ethnicity and residual ethnicity in 10c and 10d. In each table, the dependents are number of conflict events (column 1), number of riots (column 2), and number of violent events (column 3). The time-lagged dependent is not included in the regressions.

In all regressions, the SAR parameter maintains the same magnitude as in previous regressions, and remains highly significant. However, the combination of spatial autocorrelation in the dependent and independent variables performs quite poorly for one dependent, in particular: violence. With an adjusted R-squared equal to 0 in all of the violence regressions, it is clear that the Spatial Durbin is not at all able to explain the variation across time of this outcome. It performs slightly better when number of events and rioting are the modelled outcomes. Moreover, the lagged independents are never significant for the two contiguity models. This result suggests that neighborhood effects are far more powerful and important for those neighborhoods defined by ethnic ties, rather than geographical nearness.

The Durbin coefficient, given by γ, represents the indirect effect of marginal change in a indicator; that is, the average impact upon a district's neighbors of policies the district chooses to enact, or of shocks it undergoes. In the ethnicity and residual ethnicity models, three of the Durbin coefficients are statistically significant at the 5% level. Of the spatially lagged controls, only population and population density are significant, while rainfall deviations are significant at the 10% level.

When the weighted average of population in an ethnic neighborhood increases by 100,000, the average district experiences about one more incident of violent conflict. The increase is 1.33 for the residual ethnicity neighborhood. Thus there is a positive association between

population and violence. However, population density has the opposite effect. An increase of 1000 residents/km in the average density of an ethnic neighborhood leads to 13.20 fewer conflict events for the typical region. Alternatively, when population density for one district increases by 1000 residents/km, all districts in the neighborhood experience a decrease of 13.20/N conflict events, where N is average number of links. Although these shocks to population and population density may seem unrealistically large, half a million Hutus fled from Rwanda to Eastern Congo in the wake of the 1994 genocide. The UNHCR estimates that 1.1 million Syrians and 450,000 Palestinians were registered as refugees in Lebanon in 2014, a country of only 4.1 million inhabitants.

Of the spatially lagged political indicators, access to central power is positively and significantly related to number of riots. When a native inhabitant of an ethnically neighboring region becomes president or prime minister, the network as a whole experiences 3.43 more riots per year, on average.

References

[1] Acemoglu, D. and J. A. Robinson. 2009. *Economic Origins of Dictatorship and Democracy*. Cambridge: Cambridge University Press.

[2] Aragon, F.M., P. Chuhan-Pole and B.C. Land. 2015. "The Local Economic Impacts of Resource Abundance: What Have We Learned?" World Bank Policy Research Working Paper No. 7263.

[3] Belhaj, M., Y. Bramoullé and F. Deroïan. 2014. "Network Games Under Strategic Complementarities." *Games and Economic Behavior*, 88: 310–319.

[4] Besley, T. and T. Persson. 2011. "The Logic of Political Violence," *Quarterly Journal of Economics*, 126(3): 1411-1445.

[5] *Black's Law Dictionary: Free Online Legal Dictionary*. 1910. 2nd ed, available at: http://thelawdictionary.org/.

[6] Blanford, N. 2015. "For Syria's Druze, survival hinges on choosing the right ally." *Al Jazeera America*, July 2, available at: http://america.aljazeera.com/articles/2015/7/2/druze-struggle-for-survival-in-syria.html.

[7] Braithwaite, A. 2010. "Resisting Infection: How State Capacity Conditions Conflict Contagion." *Journal of Peace Research*, 47(3): 311-319.

[8] Bramoullé, Y., R. Kranton and M. D'Amours. 2014. "Strategic Interaction and Networks." *American Economic Review*, 104(3): 898-930.

[9] Brathwaite, K.J.H. 2014. "Repression and the Spread of Ethnic Conflict in Kurdistan." *Studies in Conflict & Terrorism*, 37: 473-491.

[10] Brownlee, J., T. Masoud and A. Reynolds. 2015. *The Arab Spring: Pathways of Repression and Reform*, 1st ed., Oxford: Oxford University Press.

[11] Buhaug, H. 2010. "Climate Not to Blame for African Civil Wars." Proc. Natl. Acad. Sci. USA, 107(38).

[12] Buhaug, H. and K.S. Gleditsch. 2008. "Contagion or Confusion? Why Conflicts Cluster in Space." *International Studies Quarterly,* 52(2): 215-233.

[13] Buhaug, H., N.P. Gleditsch and O.M. Theisen. 2008. "Implications of Climate Change for Armed Conflict." Paper prepared for the Social Dimensions of Climate Change Program. Washington, DC: World Bank, Social Development Department, available at: http://siteresources. worldbank. org/INTRANETSOCIALDEVELOPMENT/ Resources/SDCCWorkingPaper_Conflict.pdf.

[14] Burke, M.B., E. Miguel, S. Satyanath, J. A. Dykema and D. B. Lobell. 2009. "Warming increases the risk of civil war in Africa." Proc. Natl. Acad. Sci. USA, 106(49).

[15] Caselli, F., M. Morelli and D. Rohner. 2015. "The Geography of Interstate Resource Wars." *The Quarterly Journal of Economics,* 130: 267-315.

[16] Cederman, L.-E., L. Girardin, and K.S. Gleditsch. 2009. "Ethnonationalist Triads: Assessing the Influence of Kin Groups on Civil Wars." *World Politics*, 61: 403-437.

[17] Cederman, L.-E., A. Wimmer and B. Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics*, 62(1): 87-119.

[18] Cederman, L.-E., K.S. Gleditsch, I. Salehyana and J. Wucherpfennig. 2013. "Transborder Ethnic Kin and Civil War." International Organization, 67(2): 389-410.

[19] Collier, Paul and Anke Hoeffler. 2002. "Greed and Grievance in Civil War." World Bank, CSAE Working Paper 2002-01.

[20] Collier, P. and A. Hoeffler. 1998. "On the Economic Causes of War." Oxford Economic Papers, 50: 563-573.

[21] Dal Bó, E. and P. Dal Bó. 2011. "Workers, Warriors, And Criminals: Social Conflict In General Equilibrium." *Journal of the European Economic Association*, 9(4): 646-677.

[22] Danneman, N. and E.H. Ritter. 2014. "Contagious Rebellion and Preemptive Repression." *Journal of Conflict Resolution*, 58(2): 254-279.

[23] Diamond, J.M. 2006. "Collapse: How Societies Choose to Fail or Succeed." London: Penguin Books.

[24] Diamond, J.M. 1997. "Guns, Germs, and Steel: The Fates of Human Societies." New York: W. W. Norton & Company.

[25] Dube, O. and J. Vargas. 2013. "Commodity Price Shocks and Civil Conflict: Evidence From Colombia." *Review of Economic Studies*, 82(3).

[26] Esteban, J., M. Morelli and D. Rohner. 2015. "Strategic Mass Killings," forthcoming in *Journal of Political Economy*.

[27] Fearon, J.D. and D.D. Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review*, 97: 75-90.

[28] Fosberg, E. "Polarizatin and Ethnic Conflict in a Widened Strategic Setting." *Journal of Peace Research*, 45: 283-300.

[29] Franke, J. and T. Ozturk. 2015. "Conflict Networks." *Journal of Public Economics*, 126(C): 104-113.

[30] Gilbert, Martin. 1985. T*he Holocaust: A History of the Jews of Europe During the Second World War*. New York: Henry Holt and Co.

[31] Gleditsch, N.P. 2007. "Transnational Dimensions of Civil War." Journal of Peace Research, 44(3): 293-309.

[32] Gleditsch, N.P. 2012. "Whither the Weather? Climate Change and Conflict." *Journal of Peace Research*, 49(1): 3-9.

[33] Gurr, T.R. 2000. *Peoples Versus States: Minorities at Risk in the New Century*. Washington, D.C.: United States Institute for Peace Press.

[34] Harari, M. and E. La Ferrara. 2013. "Conflict, Climate and Cells: A Disaggregated Analysis." CEPR Discussion Paper No. DP9277.

[35] Hegre, H., T. Ellingsen, S. Gates and N.P. Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change and Civil War, 1816-1992." *American Poltical Science Review*, 95(1):33-48.

[36] Heydemann, S. and R. Leenders. 2014. "Authoritarian Learning and Counterrevolution." In *The Arab Uprisings Explained: New Contentious Politics in the Middle East*, M. Lynch, ed., New York: Columbia University Press.

[37] Hegre, H. and N. Sambanis. "Sensitivity Analysis of Empirical Results on Civil War Onset." *Journal of Conflict Resolution*, 50(4): 508-535.

[38] King, Jeremy. 2002. *Budweisers into Czechs and Germans*. Princeton, NJ: Princeton University Press.

[39] Konig, M.D., D. Rohner, M. Thoenig and F. Zilibotti. 2015. "Networks in Conflict: Theory and Evidence from the Great War of Africa." CEPR Discussion Paper No. DP10348.

[40] Lake, D.A. and D. Rothchild. 1998. *The International Spread of Ethnic Conflict: Fear, Diffusion, and Escalation*. Princeton, NJ: Princeton University Press.

[41] *The Long Way Home.* 1998. Documentary film, M.J. Harris dir.

[42] Lynch, M. 2014. *The Arab Uprisings Explained: New Contentious Politics in the Middle East*, New York: Columbia University Press.

[43] Mann, M. 2005. *The Dark Side of Democracy: Explaining Ethnic Cleansing*. Cambridge: Cambridge University Press.

[44] Maves, J. and A. Braithwaite. "Autocratic Institutions and Civil Conflict Contagion." *The Journal of Politics*, 75(2): 478-490.

[45] Meredith, M. 2007. *Mugabe: Power, Plunder, and the Struggle for Zimbabwe's Future*, New York: Public Affairs.

[46] Miguel, E., S. Satyanath and E. Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy*, 112(4): 725-753.

[47] Morelli, M. and D. Rohner. 2015. "Resource Concentration and Civil Wars." *Journal of Development Economics*, 117: 32-47.

[48] Reuveny, Rafael. 2007. "Climate Change-Induced Migration and Violent Conflict." *Political Geography,* 26: 656-673.

[49] Robinson, J. A., D. Acemoglu, and S. Johnson. 2003. "An African Success Story: Botswana." In *In Search of Prosperity: Analytic Narratives on Economic Growth*, D. Rodrik, ed., 80-119. Princeton: Princeton University Press.

[50] Robinson, J. A, D. Acemoglu, and S. Johnson. 2005. "Institutions as a Fundamental Cause of Long-Run Growth." In *Handbook of Economic Growth* 1A: 386-472.

[51] Ross, M.L. 2013. "The Politics of the Resource Curse: A Review." Forthcoming in the Handbook on the Politics of Development, C. Lancaster and N. van de Walle, ed.

[52] Salehyan, I. and K.S. Gleditsch. 2006. "Refugee Flows and the Spread of Civil War." *International Organization,* 60(2): 335-366.

[53] Sambanis, N. 2002. "A Review of Recent Advances and Future Directions in the Quantitative Literature on Civil War." *Defense Economics*, 13(1): 215-243.

[54] Simons, Marlise. 2010. "International Court Adds Genocide to Charges Against Sudan Leader." *The New York Times*, July 12, available at: http://www.nytimes.com/2010/07/13/world/africa/13hague.html?_r=1.

[55] Stearns, J. 2012. *Dancing in the Glory of Monsters: The Collapse of the Congo and the Great War of Africa,* New York: Public Affairs.

[56] Tainter, J. 1988. *The Collapse of Complex Societies*. Cambridge: Cambridge University Press.

[57] van der Ploeg, F. 2010. "Natural Resources: Curse of Blessing?" CESifo Working Paper Resources and Environment No. 3125.

[58] VICE News. 2014. *The Islamic State*, documentary film, available at: https://news.vice.com/video/the-islamic-state-full-length.

[59] Wood, G. 2015. "What ISIS Really Wants." *The Atlantic*, March, available at: http://www.theatlantic.com/magazine/archive/2015/03/what-isis-really-wants/384980/.

FIGURE 1.

The contiguity (green), ethnicity (grey) and residual contiguity (purple) weight matrices

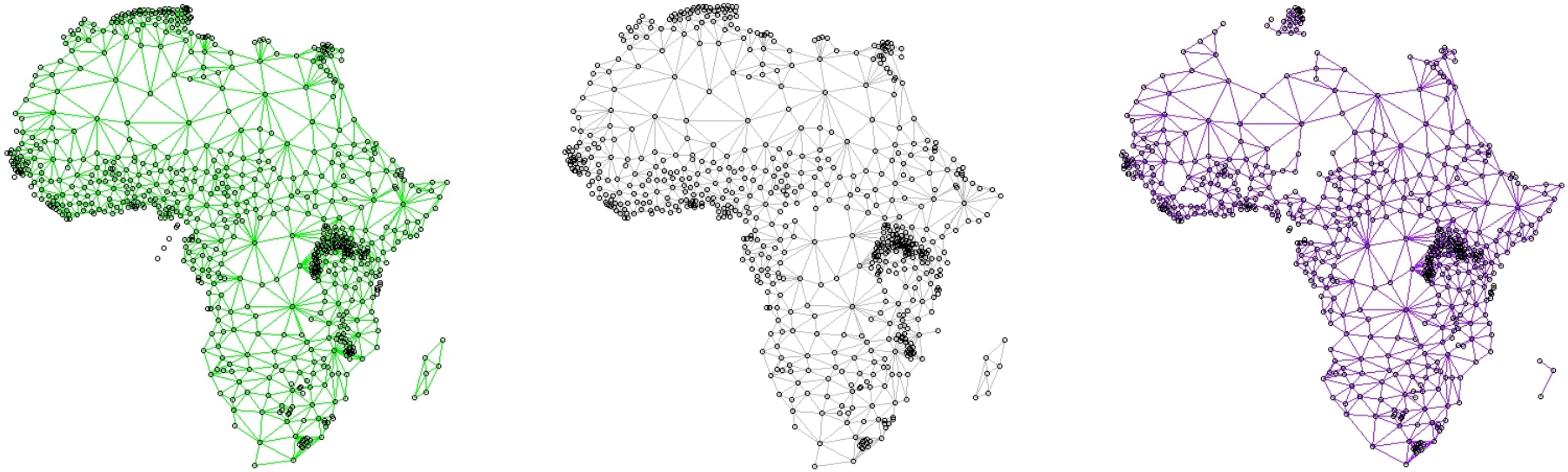


FIGURE 2.

Spatial residuals from the OLS model

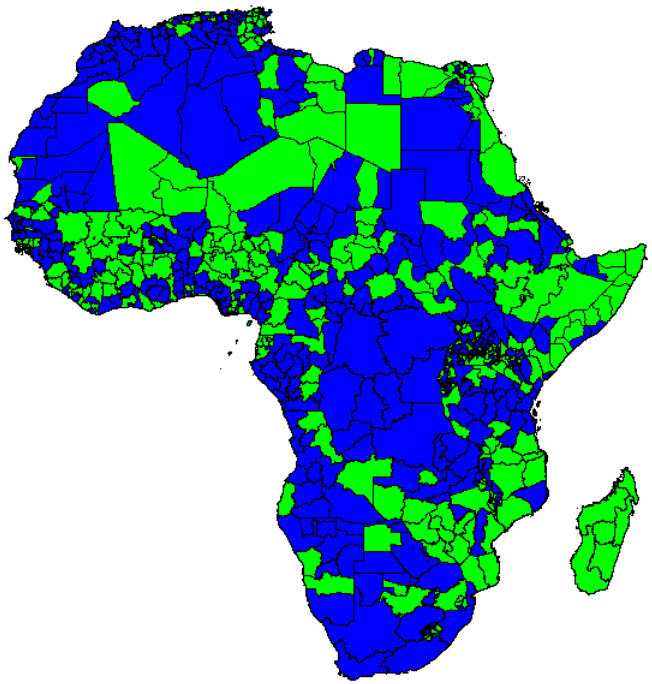


TABLE 1.

Descriptive statistics

Statistic	Mean	St. Dev.	Min	Max
Number of Events	8.27	34.35	0	879
Riots	2.09	12.07	0	652
Violent Events	5.50	26.41	0	785
Fatalities	42.10	900.85	0	66,565
Population	1,083,917	1,679,260	9.12	23,628,270
Population Density	247.59	882.37	0.08	17,268.71
Rainfall	3.57	3.67	0	25.17
Deviations in Rainfall	-1.46	1.97	-6.47	5.22
Nightlights	5.39	7.18	0	62.35
Capital	0.07	0.25	0	1
Access to Central Power	0.13	0.33	0	1
Access to Outside Power	0.01	0.08	0	1
Distance from Capital	321.77	266.25	1.03	1,436.06

TABLE 2.

The four spatial networks

	E	E_R	G	G_R
N	699	694	764	592
% nonzero links	2.49	2.07	0.7	0.53
ave. number of links	17.4	14.34	5.37	3.14
max. number of links	95	64	18	15

TABLE 3.

OLS model: Contiguity

	(1) N. Events	(2) N. Events	(3) Riots	(4) Riots	(5) Violence	(6) Violence
Population	1.61**	0.67**	0.96**	0.41*	0.56+	0.22+
	(0.56)	(0.24)	(0.36)	(0.19)	(0.29)	(0.12)
Pop. density	2.33	-0.61	-0.71	-0.72	2.62	0.07
	(7.39)	(2.44)	(1.68)	(0.67)	(6.27)	(2.09)
Nightlights	1.29***	0.72***	0.74***	0.36***	0.45*	0.34**
	(0.28)	(0.15)	(0.12)	(0.07)	(0.19)	(0.11)
Rainfall deviations	0.68***	0.00	0.21***	0.02	0.54***	0.07
	(0.20)	(0.14)	(0.05)	(0.04)	(0.16)	(0.12)
Capital	-20.03	-6.97	-1.21	-0.89	-22.86	-7.79
	(20.91)	(7.78)	(2.69)	(1.24)	(18.09)	(6.66)
Access	3.20	1.93+	1.10	0.70	1.58	0.91
	(2.24)	(1.16)	(1.30)	(0.63)	(1.30)	(0.75)
Outside	-3.39	-1.59	-1.32	-0.16	-2.00	-1.62+
	(2.83)	(1.04)	(1.30)	(0.68)	(1.63)	(0.92)
Distance	-24.85**	-14.03***	0.46	-0.10	-22.54**	-12.80***
	(11.90)	(4.40)	(1.00)	(0.56)	(10.73)	(3.95)
Dependent t-1		0.71***		0.68***		0.70***
		(0.03)		(0.09)		(0.05)
R-squared	0.03	0.48	0.06	0.43	0.01	0.46
N	764	764	764	764	764	764

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 4.

SAR Models

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violence	(6) Violence
Contiguity						
Spatial autocorr.	0.40***	0.22***	0.48***	0.36***	0.43***	0.25***
	(0.05)	(0.03)	(0.08)	(0.06)	(0.07)	(0.04)
Dependent t-1		0.65***		0.62***		0.64***
		(0.04)		(0.08)		(0.06)
R-squared	0.04	0.57	0.07	0.39	0.02	0.63
N	764	764	764	764	764	764
Resid. Contiguity						
Spatial autocorr.	0.25***	0.13***	0.26***	0.20***	0.25***	0.13***
	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)	(0.03)
Dependent t-1		0.69***		0.80***		0.67***
		(0.04)		(0.08)		(0.06)
R-squared	0.04	0.61	0.08	0.58	0.04	0.64
N	592	592	592	592	592	592
Ethnicity						
Spatial autocorr.	0.41***	0.25***	0.62***	0.46***	0.41***	0.26***
	(0.05)	(0.04)	(0.10)	(0.08)	(0.05)	(0.04)
Dependent t-1		0.64***		0.59***		0.61***
		(0.03)		(0.07)		(0.04)
R-squared	0.06	0.55	0.08	0.41	0.01	0.58
N	699	699	699	699	699	699
Resid. Ethnicity						
Spatial autocorr.	0.31***	0.20***	0.52***	0.38***	0.27***	0.17***
	(0.04)	(0.03)	(0.10)	(0.07)	(0.04)	(0.02)
Dependent t-1		0.64***		0.62***		0.60***
		(0.03)		(0.08)		(0.04)
R-squared	0.06	0.55	0.08	0.42	0.84	0.87
N	694	694	694	694	694	694

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 5a.

SAR Models with political variables: contiguity and residual contiguity

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violence	(6) Violence
Contiguity						
Spatial autocorr.	0.40***	0.22***	0.48***	0.36***	0.43***	0.25***
	(0.05)	(0.03)	(0.08)	(0.06)	(0.07)	(0.04)
Dependent t-1		0.65***		0.62***		0.64***
		(0.04)		(0.08)		(0.06)
Capital	-4.44	0.01	1.50	0.67	-10.88	-2.88
	(16.86)	(6.65)	(3.03)	(1.45)	(14.15)	(5.55)
Access	2.85	1.82	1.10	0.73	1.18	0.74
	(2.01)	(1.13)	(1.19)	(0.60)	(1.11)	(0.71)
Outside	-2.92	-1.41	-0.90	0.03	-2.11	-1.69
	(2.81)	(1.18)	(1.38)	(0.65)	(1.83)	(1.21)
Distance	-17.47*	-10.73***	0.412	-0.09	-15.60*	-9.69***
	(9.79)	(3.71)	(1.02)	(0.63)	(8.55)	(3.27)
R-squared	0.03	0.53	0.07	0.39	0.00	0.55
N	764	764	764	764	764	764
Resid. Contiguity						
Spatial autocorr.	0.25***	0.13***	0.26***	0.20***	0.25***	0.13***
	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)	(0.03)
Dependent t-1		0.68***		0.80***		0.67***
		(0.04)		(0.08)		(0.06)
Capital	-8.82	-2.37	3.17	0.65	-16.68	-5.57
	(18.86)	(7.23)	(1.93)	(0.88)	(16.54)	(6.47)
Access	1.72	1.27	0.60	0.45	0.38	0.27
	(2.34)	(1.24)	(1.51)	(0.57)	(1.12)	(0.68)
Outside	-1.04	-0.42	-1.23	-0.15	-0.51	-0.60
	(2.86)	(1.62)	(1.79)	(0.76)	(1.94)	(1.27)
Distance	-21.01*	-12.18***	0.41	-0.06	-19.15**	-11.38***
	(11.15)	(4.15)	(1.01)	(0.49)	(10.03)	(3.85)
R-squared	0.02	0.57	0.08	0.58	0.00	0.57
N	592	592	592	592	592	592

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 5b.

SAR Models with political variables, ethnicity and residual ethnicity

	(1) N. Events	(2) N. Events	(3) Riots	(4) Riots	(5) Violence	(6) Violence
Ethnicity						
Spatial autocorr.	0.41***	0.25***	0.62***	0.46***	0.40***	0.26***
	(0.05)	(0.04)	(0.10)	(0.08)	(0.05)	(0.04)
Dependent t-1		0.64***		0.59***		0.61***
		(0.03)		(0.07)		(0.04)
Capital	-4.62	0.53	1.66	0.86	-11.24	-2.79
	(18.21)	(7.49)	(3.06)	(1.48)	(15.86)	(6.71)
Access	3.05	1.62	0.59	0.33	1.73	0.86
	(2.12)	(1.24)	(1.17)	(0.58)	(1.17)	(0.81)
Outside	-2.98	-1.46	0.07	0.73+	-2.79	-2.19
	(2.61)	(1.48)	(0.65)	(0.44)	(1.98)	(1.49)
Distance	-17.63	-10.33**	-0.37	-0.60	-15.51	-9.23**
	(10.79)	(4.33)	(0.83)	(0.52)	(9.73)	(4.05)
R-squared	0.04	0.50	0.08	0.41	0.00	0.49
N	699	699	699	699	699	699
Resid. Ethnicity						
Spatial autocorr.	0.31***	0.20***	0.52***	0.38***	0.26***	0.16***
	(0.04)	(0.03)	(0.10)	(0.07)	(0.04)	(0.02)
Dependent t-1		0.64***		0.62***		0.60***
		(0.03)		(0.08)		(0.04)
Capital	-14.18	-5.22	0.89	0.27	-19.48	-8.28
	(19.98)	(8.56)	(3.03)	(1.39)	(17.62)	(8.00)
Access	3.01	1.59	0.61	0.34	1.76	0.89
	(2.22)	(1.27)	(1.25)	(0.61)	(1.23)	(0.83)
Outside	-4.25	-2.12	-0.55	0.52	-3.10+	-2.41
	(2.60)	(1.56)	(0.77)	(0.51)	(1.83)	(1.50)
Distance	-21.29*	0.64***	-0.11	-4.14	-19.68*	-12.07**
	(11.88)	(0.03)	(0.802)	(4.73)	(10.86)	(4.87)
R-squared	0.04	0.49	0.08	0.42	0.02	0.57
N	694	694	694	694	694	694

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6a.

Political interactions: number of events

	(1)	(2)	(3)	(4)	(5)	(6)
			N. Events			
Spatial autocorr.	0.25***	0.35***	0.30***	0.25***	0.36***	0.31***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
N. events t-1	0.65***			0.61***		
	(0.04)			(0.06)		
Political t-1		1.13***			0.87***	
		(0.18)			(0.08)	
Violence t-1			0.78***			0.70***
			(0.08)			(0.08)
Capital	17.54***	23.78***	22.12***			
	(2.46)	(2.30)	(5.42)			
Capital x n. events	-0.04					
	(0.05)					
Capital x political		-0.33+				
		(0.20)				
Capital x violence			-0.11			
			(0.12)			
Access				1.25	2.68	1.21
				(1.05)	(1.64)	(1.61)
Access x n. events				0.04		
				(0.10)		
Access x political					0.01	
					(0.16)	
Access x violence						0.12
						(0.21)
R-squared	0.53	0.15	0.35	0.55	0.17	0.36
N	699	699	699	699	699	699

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6b.

Political interactions: number of events, continued

	(1)	(2)	(3)	(4)	(5)	(6)
			N. Events			
Spatial autocorr.	0.25***	0.36***	0.30***	0.25***	0.34***	0.30***
	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)
N. events t-1	0.64***			0.49***		
	(0.03)			(0.06)		
Political t-1		0.88***			-0.11	
		(0.11)			(0.37)	
Violence t-1			0.76***			0.57***
			(0.07)			(0.13)
Outside	0.94	-0.54	0.10			
	(0.97)	(1.46)	(1.76)			
Outside x n. events	-0.35***					
	(0.04)					
Outside x political		-0.62*				
		(0.28)				
Outside x violence			-0.56***			
			(0.09)			
Distance				-16.09***	-10.05*	-11.07***
				(7.05)	(2.65)	(2.20)
Distance x n. events				0.03*		
				(0.01)		
Distance x political					0.24*	
					(0.10)	
Distance x violence						0.03+
						(0.02)
R-squared	0.55	0.16	0.36	0.44	0.08	0.30
N	699	699	699	699	699	699

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 6c.

Political interactions: riots

	(1)	(2)	(3) Political	(4)	(5)	(6)
Spatial autocorr.	0.58***	0.46***	0.61***	0.59***	0.45***	0.62***
	(0.09)	(0.08)	(0.10)	(0.09)	(0.08)	(0.10)
N. events t-1	0.05**			0.09**		
	(0.02)			(0.03)		
Political t-1		0.58***			0.60***	
		(0.10)			(0.09)	
Violence t-1			0.03**			0.03**
			(0.01)			(0.01)
Capital	-7.33	1.65+	0.18			
	(5.15)	(0.91)	(3.82)			
Capital x events	0.21*					
	(0.11)					
Capital x political		0.01				
		(0.13)				
Capital x violence			0.06			
			(0.08)			
Access				0.05	0.37	0.31
				(0.76)	(0.52)	(1.08)
Access x n. events				0.03		
				(0.06)		
Access x political					-0.01	
					(0.12)	
Access x violence						0.03
						(0.03)
R-squared	0.15	0.42	0.09	0.13	0.41	0.09
N	699	699	699	699	699	699

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6d.

Political interactions: riots, continued

	(1)	(2)	(3) Political	(4)	(5)	(6)
Spatial autocorr.	0.59***	0.46***	0.62***	0.59***	0.43***	0.62***
	(0.10)	(0.08)	(0.10)	(0.10)	(0.08)	(0.10)
N. events t-1	0.11**			0.25		
	(0.04)			(0.16)		
Political t-1		0.59***			0.30+	
		(0.07)			(0.16)	
Violence t-1			0.04**			0.04
			(0.01)			(0.06)
Outside	0.43	0.74	0.33			
	(0.55)	(0.45)	(0.71)			
Outside x n. events	-0.04					
	(0.05)					
Outside x political		-0.00				
		(0.12)				
Outside x violence			-0.05			
			(0.04)			
Distance				.81***	-0.17	0.04
				(.14)	(0.35)	(0.04)
Distance x n. events				-0.00		
				(0.03)		
Distance x political					0.07*	
					(0.03)	
Distance x violence						-0.00
						(0.01)
R-squared	0.13	0.41	0.09	0.13	0.41	0.09
N	699	699	699	699	699	699

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 6e.

Political interactions: violence

	(1)	(2)	(3) Violence	(4)	(5)	(6)
Spatial autocorr.	0.28***	0.40***	0.26***	0.27***	0.41***	0.26***
	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
N. events t-1	0.44***			0.45***		
	(0.05)			(0.07)		
Political t-1		0.45**			0.20***	
		(0.16)			(0.06)	
Violence t-1			0.62***			0.62***
			(0.05)			(0.08)
Capital	13.90**	12.99***	12.51***			
	(4.71)	(1.85)	(1.68)			
Capital x n. events	-0.13					
	(0.12)					
Capital x political		-0.30+				
		(0.16)				
Capital x violence			-0.06			
			(0.05)			
Access				1.78+	1.57	1.10
				(0.92)	(1.11)	(0.76)
Access x n. events				-0.08		
				(0.08)		
Access x political					0.04	
					(0.11)	
Access x violence						-0.03
						(0.13)
R-squared	0.39	0.02	0.54	0.42	0.01	0.58
N	699	699	699	699	699	699

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 6f.

Political interactions: violence, continued

	(1)	(2)	(3) Violence	(4)	(5)	(6)
Spatial autocorr.	0.29***	0.40***	0.27***	0.28***	0.39***	0.26***
	(0.04)	(0.05)	(0.04)	(0.04)	(0.05)	(0.04)
N. events t-1	0.41***			0.15		
	(0.05)			(0.22)		
Political t-1		0.23***			-0.40	
		(0.06)			(0.33)	
Violence t-1			0.61***			0.46***
			(0.04)			(0.06)
Outside	0.16	-1.48	-0.10			
	(1.13)	(1.43)	(1.22)			
Outside x n. events	-0.26***					
	(0.05)					
Outside x n. political		-0.68*				
		(0.28)				
Outside x violence			-0.43***			
			(0.06)			
Distance				-13.15*	-9.01**	-8.91***
				(6.79)	(3.75)	(2.63)
Distance x n. events				0.12		
				(0.09)		
Distance x riots					0.02	
					(0.02)	
Distance x violence						0.03
						(0.02)
R-squared	0.41	0.01	0.58	0.24	0.00	0.39
N	699	699	699	699	699	699

$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 7a.

SAC model: contiguity and residual contiguity

	(1) N. Event	(2) N. Event	(3) Political	(4) Political	(5) Violence	(6) Violence
Internal R. Contig.						
Spatial autocorr.	0.38***	0.23***	0.26***	0.16**	0.34***	0.20***
	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)	(0.03)
Dependent t-1		0.63***		0.81***		0.62***
		(0.05)		(0.08)		(0.07)
Capital	-1.79	1.36	3.39+	0.83	-12.11	-3.71
	(21.69)	(10.07)	(2.00)	(0.89)	(19.08)	(8.76)
Access	2.21	1.61	0.83	0.61	0.70	0.43
	(2.32)	(1.33)	(1.60)	(0.61)	(1.02)	(0.68)
Outside	0.95	1.57	-0.64	0.41	0.80	0.98
	(2.23)	(1.15)	(1.51)	(0.66)	(1.76)	(0.75)
Distance	-16.09**	-9.82***	-0.47	-0.27	-12.84*	-8.42***
	(6.53)	(2.49)	(1.09)	(0.42)	(6.55)	(2.66)
R-squared	0.03	0.60	0.08	0.59	0.00	0.61
N	536	536	536	536	536	536
External R. Contig.						
Spatial autocorr.	0.00	0.00	-0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Dependent t-1		0.68***		0.90***		0.64***
		(0.03)		(0.10)		(0.05)
Capital	-9.68	-3.47	7.50**	2.03*	-20.90	-9.02
	(27.86)	(11.27)	(2.30)	(0.97)	(24.48)	(10.57)
Access	-1.44	0.19	-2.37	-0.84	-0.10	0.63
	(3.46)	(2.00)	(2.00)	(0.62)	(1.76)	(1.05)
Outside	3.46	-1.29	-0.89	-0.81	3.82*	0.02
	(2.33)	(2.76)	(1.42)	(1.21)	(1.49)	(1.43)
Distance	-16.08**	-9.92***	-0.39	-0.23	-12.90**	-8.56***
	(6.61)	(2.53)	(1.08)	(0.41)	(6.61)	(2.68)
R-squared	0.02	0.49	0.06	0.61	0.01	0.42
N	307	307	307	307	307	307

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 7b.

SAC model: ethnicity and residual ethnicity

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violence	(6) Violence
Internal R. Ethnic.						
Spatial autocorr.	0.34***	0.22***	0.42***	0.29***	0.25***	0.16***
	(0.04)	(0.03)	(0.09)	(0.06)	(0.03)	(0.02)
Dependent t-1		0.60***		0.59***		0.59***
		(0.03)		(0.06)		(0.05)
Capital	7.90	8.01	-0.08	-0.10	-2.21	1.98
	(24.76)	(12.01)	(3.81)	(1.94)	(19.89)	(9.43)
Access	1.92	0.59	0.58	0.32	1.35	0.40
	(2.22)	(1.24)	(1.31)	(0.66)	(1.38)	(0.90)
Outside	-2.36	-0.86	-0.20	0.64+	-1.83+	-1.46
	(1.78)	(1.74)	(0.89)	(0.36)	(1.01)	(1.46)
Distance	-15.05**	-9.17***	-0.48	-0.51	-12.73*	-8.54**
	(7.07)	(3.00)	(0.98)	(0.43)	(7.15)	(3.38)
R-squared	0.05	0.46	0.10	0.43	0.00	0.44
N	594	594	594	594	594	594
External R. Ethnic.						
Spatial autocorr.	0.02	0.01	-0.04***	-0.01**	0.04+	0.03*
	(0.02)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)
Dependent t-1		0.67***		0.91***		0.61***
		(0.03)		(0.04)		(0.04)
Capital	-10.38	-3.47	4.61*	0.84	-19.54	-8.39
	(20.09)	(8.08)	(1.96)	(0.80)	(17.61)	(7.85)
Access	2.51	1.26	0.08	-0.03	1.91	1.00
	(2.57)	(1.43)	(1.55)	(0.46)	(1.43)	(0.94)
Outside	-1.29	-0.48	-0.61	0.24	-1.10	-0.94
	(1.87)	(1.38)	(0.75)	(0.35)	(1.45)	(1.34)
Distance	-15.21**	-9.22***	-0.19	-0.27	-13.18*	-8.85***
	(7.12)	(3.06)	(1.05)	(0.52)	(7.15)	(3.37)
R-squared	0.01	0.52	0.10	0.73	0.00	0.42
N	622	622	622	622	622	622

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

TABLE 8a.

SAR model: internal and external contiguity

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violent	(6) Violent
Internal Contiguity						
Spatial autocorr.	0.45***	0.27***	0.44***	0.33***	0.45***	0.27***
	(0.04)	(0.04)	(0.07)	(0.05)	(0.05)	(0.04)
Dependent t-1		0.62***		0.61***		0.61***
		(0.04)		(0.07)		(0.07)
Capital	8.24	6.52	2.50	3.43**	-6.21	-0.73
	(18.82)	(8.51)	(3.10)	(1.05)	(16.19)	(7.25)
Access	3.31+	2.13+	1.16	0.79	1.48	0.93
	(1.92)	(1.14)	(1.15)	(0.59)	(1.02)	(0.70)
Outside	-1.35	-0.39	-0.26	0.27	-0.79	-0.77
	(1.60)	(1.27)	(0.77)	(0.49)	(1.01)	(1.25)
Distance	-13.74**	-8.88***	-0.23	-0.34	-10.77**	-7.41***
	(5.46)	(2.04)	(1.20)	(0.59)	(5.33)	(2.12)
R-squared	0.01	0.59	0.06	0.54	0.00	0.60
N	762	762	762	762	762	762
External Contiguity						
Spatial autocorr.	-0.00	0.00	-0.02	0.00	-0.00	0.01
	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)
Dependent t-1		0.67***		0.91***		0.64***
		(0.03)		(0.09)		(0.05)
Capital	-0.58	0.40	8.21***	1.91*	-13.48	-5.63
	(24.80)	(9.89)	(1.63)	(0.79)	(21.87)	(9.26)
Access	0.06	0.69	-1.29	-0.41	0.56	0.72
	(2.57)	(1.44)	(1.36)	(0.44)	(1.47)	(0.84)
Outside	2.47	-0.67	-1.06	-0.81	3.10**	0.39
	(1.77)	(2.05)	(1.42)	(1.08)	(1.14)	(1.01)
Distance	-13.39**	-8.64***	0.09	0.08	-10.77**	-7.45***
	(5.60)	(2.16)	(1.24)	(0.64)	(5.42)	(2.17)
R-squared	0.02	0.51	0.02	0.61	0.02	0.43
N	447	447	447	447	447	447

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 8b.

SAR model: internal and external residual contiguity

	(1) N. Event	(2) N. Event	(3) Political	(4) Political	(5) Violence	(6) Violence
Internal R. Contig.						
Spatial autocorr.	0.38***	0.23***	0.26***	0.16**	0.34***	0.20***
	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)	(0.03)
Dependent t-1		0.63***		0.81***		0.62***
		(0.05)		(0.08)		(0.07)
Capital	-1.79	1.36	3.39+	0.83	-12.11	-3.71
	(21.69)	(10.07)	(2.00)	(0.89)	(19.08)	(8.76)
Access	2.21	1.61	0.83	0.61	0.70	0.43
	(2.32)	(1.33)	(1.60)	(0.61)	(1.02)	(0.68)
Outside	0.95	1.57	-0.64	0.41	0.80	0.98
	(2.23)	(1.15)	(1.51)	(0.66)	(1.76)	(0.75)
Distance	-16.09**	-9.82***	-0.47	-0.27	-12.84*	-8.42***
	(6.53)	(2.49)	(1.09)	(0.42)	(6.55)	(2.66)
R-squared	0.03	0.60	0.08	0.59	0.00	0.61
N	536	536	536	536	536	536
External R. Contig.						
Spatial autocorr.	0.00	0.00	-0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Dependent t-1		0.68***		0.90***		0.64***
		(0.03)		(0.10)		(0.05)
Capital	-9.68	-3.47	7.50**	2.03*	-20.90	-9.02
	(27.86)	(11.27)	(2.30)	(0.97)	(24.48)	(10.57)
Access	-1.44	0.19	-2.37	-0.84	-0.10	0.63
	(3.46)	(2.00)	(2.00)	(0.62)	(1.76)	(1.05)
Outside	3.46	-1.29	-0.89	-0.81	3.82*	0.02
	(2.33)	(2.76)	(1.42)	(1.21)	(1.49)	(1.43)
Distance	-16.08**	-9.92***	-0.39	-0.23	-12.90**	-8.56***
	(6.61)	(2.53)	(1.08)	(0.41)	(6.61)	(2.68)
R-squared	0.02	0.49	0.06	0.61	0.01	0.42
N	307	307	307	307	307	307

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 8c.

SAR model: internal and external ethnicity

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violent	(6) Violent
Internal Ethnicity						
Spatial autocorr.	0.41***	0.24***	0.53***	0.39***	0.40***	0.25***
	(0.05)	(0.05)	(0.10)	(0.07)	(0.04)	(0.04)
Dependent t-1		0.61***		0.56***		0.60***
		(0.03)		(0.05)		(0.05)
Capital	16.71	12.87	3.11	2.00	6.15	8.16
	(20.18)	(9.69)	(3.66)	(1.99)	(16.59)	(7.66)
Access	2.85	1.57	0.52	0.34	1.60	0.76
	(2.27)	(1.39)	(1.21)	(0.60)	(1.27)	(0.89)
Outside	-1.31	-0.38	0.36	0.86**	-1.70	-1.37
	(1.72)	(1.44)	(0.70)	(0.32)	(1.06)	(1.36)
Distance	-13.95**	-8.75***	-0.83	-0.76*	-10.85*	-7.17***
	(6.26)	(2.50)	(0.98)	(0.44)	(6.11)	(2.56)
R-squared	0.04	0.48	0.07	0.39	0.00	0.47
N	628	628	628	628	628	628
External Ethnicity						
Spatial autocorr.	0.02	0.01	-0.05***	-0.02***	0.05+	0.03*
	(0.02)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)
Dependent t-1		0.67***		0.91***		0.62***
		(0.03)		(0.04)		(0.04)
Capital	-10.23	-3.31	4.64*	0.84	-19.17	-8.02
	(20.04)	(7.97)	(1.97)	(0.81)	(17.45)	(7.64)
Access	2.44	1.30	0.10	-0.02	1.83	1.01
	(2.63)	(1.46)	(1.58)	(0.47)	(1.46)	(0.96)
Outside	-1.32	-0.50	-0.65	0.23	-1.15	-0.98
	(1.88)	(1.38)	(0.75)	(0.35)	(1.48)	(1.35)
Distance	-13.88**	-8.51***	-0.53	-0.52	-10.90*	-7.30***
	(6.25)	(2.49)	(1.05)	(0.53)	(5.60)	(2.59)
R-squared	0.01	0.52	0.10	0.73	0.00	0.43
N	611	611	611	611	611	611

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 8d.

SAR model: internal and external residual ethnicity

	(1) N. Events	(2) N. Events	(3) Political	(4) Political	(5) Violence	(6) Violence
Internal R. Ethnic.						
Spatial autocorr.	0.34***	0.22***	0.42***	0.29***	0.25***	0.16***
	(0.04)	(0.03)	(0.09)	(0.06)	(0.03)	(0.02)
Dependent t-1		0.60***		0.59***		0.59***
		(0.03)		(0.06)		(0.05)
Capital	7.90	8.01	-0.08	-0.10	-2.21	1.98
	(24.76)	(12.01)	(3.81)	(1.94)	(19.89)	(9.43)
Access	1.92	0.59	0.58	0.32	1.35	0.40
	(2.22)	(1.24)	(1.31)	(0.66)	(1.38)	(0.90)
Outside	-2.36	-0.86	-0.20	0.64+	-1.83+	-1.46
	(1.78)	(1.74)	(0.89)	(0.36)	(1.01)	(1.46)
Distance	-15.05**	-9.17***	-0.48	-0.51	-12.73*	-8.54**
	(7.07)	(3.00)	(0.98)	(0.43)	(7.15)	(3.38)
R-squared	0.05	0.46	0.10	0.43	0.00	0.44
N	594	594	594	594	594	594
External R. Ethnic.						
Spatial autocorr.	0.02	0.01	-0.04***	-0.01**	0.04+	0.03*
	(0.02)	(0.01)	(0.01)	(0.00)	(0.02)	(0.01)
Dependent t-1		0.67***		0.91***		0.61***
		(0.03)		(0.04)		(0.04)
Capital	-10.38	-3.47	4.61*	0.84	-19.54	-8.39
	(20.09)	(8.08)	(1.96)	(0.80)	(17.61)	(7.85)
Access	2.51	1.26	0.08	-0.03	1.91	1.00
	(2.57)	(1.43)	(1.55)	(0.46)	(1.43)	(0.94)
Outside	-1.29	-0.48	-0.61	0.24	-1.10	-0.94
	(1.87)	(1.38)	(0.75)	(0.35)	(1.45)	(1.34)
Distance	-15.21**	-9.22***	-0.19	-0.27	-13.18*	-8.85***
	(7.12)	(3.06)	(1.05)	(0.52)	(7.15)	(3.37)
R-squared	0.01	0.52	0.10	0.73	0.00	0.42
N	622	622	622	622	622	622

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 9a.

Durbin model: contiguity

	(1) N. Events	(2) Riots	(3) Violence
Contiguity- X			
Lambda	0.40***	0.48***	0.43***
	(0.05)	(0.08)	(0.07)
Population	1.21*	0.89*	-0.05
	(0.61)	(0.39)	(0.26)
Pop. density	3.22	0.15	2.51
	(5.89)	(1.57)	(4.43)
Nightlights	0.55	0.44**	-0.09
	(0.37)	(0.15)	(0.29)
Rainfall deviations	-0.83	-0.05	-0.70
	(1.32)	(0.20)	(1.18)
Capital	-5.17	-0.24	-9.37
	(15.96)	(3.39)	(12.47)
Access	2.72	1.07	0.76
	(2.00)	(1.19)	(1.07)
Outside	-3.10	-0.72	-2.18
	(2.85)	(1.29)	(1.81)
Distance	-0.03	-0.00	-0.02
	(0.02)	(0.00)	(0.02)
Wx			
Population	-0.09	-0.19	0.11
	(0.37)	(0.22)	(0.23)
Density	11.48	-0.58	11.33
	(19.11)	(4.88)	(17.16)
Nightlights	-0.34	0.13	-0.45
	(0.62)	(0.34)	(0.42)
Rainfall deviations	1.08	0.06	0.96
	(1.48)	(0.21)	(1.31)
Capital	-18.87	0.44	-9.90
	(41.99)	(8.02)	(36.03)
Access	-1.60	-1.83	0.65
	(3.48)	(1.94)	(2.22)
Outside	9.78	4.81	2.38
	(6.99)	(3.44)	(4.55)
Distance	-0.01	0.01	-0.02
	(0.03)	(0.01)	(0.03)
R-squared	0.02	0.07	0.00
N	764	764	764

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 9b.

Durbin model: residual contiguity

	(1) N. Events	(2) Riots	(3) Violence
Resid. Contiguity- X			
Lambda	0.25***	0.26***	0.26***
	(0.05)	(0.05)	(0.05)
Population	1.19*	0.58**	0.24
	(0.54)	(0.19)	(0.40)
Pop. density	6.03	-4.78+	13.13
	(22.06)	(2.46)	(19.35)
Nightlights	0.22	0.66*	-0.65
	(0.68)	(0.33)	(0.60)
Rainfall deviations	-0.27	0.10	-0.35
	(0.98)	(0.12)	(0.90)
Capital	-8.15	2.48	-16.07
	(21.60)	(2.73)	(17.89)
Access	1.65	0.60	0.07
	(2.31)	(1.50)	(1.10)
Outside	-0.96	-1.07	-0.19
	(2.75)	(1.68)	(1.92)
Distance	-0.04	0.00	-0.04
	(0.03)	(0.00)	(0.03)
Wx			
Population	-0.61	-0.16	-0.29
	(0.43)	(0.17)	(0.33)
Pop. density	19.33	-3.76	22.70
	(37.66)	(2.99)	(36.09)
Nightlights	0.07	0.23	-0.26
	(1.00)	(0.52)	(0.68)
Rainfall deviations	0.66	-0.14	0.84
	(1.11)	(0.14)	(1.00)
Capital	28.20	-1.04	33.99
	(41.71)	(5.25)	(34.57)
Access	0.94	-0.06	1.12
	(2.05)	(0.77)	(1.57)
Outside	2.11	0.78	0.66
	(2.38)	(1.54)	(1.35)
Distance	0.01	0.00	0.01
	(0.03)	(0.00)	(0.02)
R-squared	0.02	0.09	0.00
N	592	592	592

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 9c.

Durbin model: ethnicity

	(1) N. Events	(2) Riots	(3) Violence
Ethnicity- X			
Lambda	0.40***	0.61***	0.39***
	(0.05)	(0.10)	(0.05)
Population	0.91	0.80+	-0.22
	(0.68)	(0.43)	(0.32)
Pop. density	10.91	2.36	6.88
	(9.39)	(2.00)	(7.90)
Nightlights	0.16	0.24+	-0.19
	(0.47)	(0.13)	(0.38)
Rainfall deviations	0.52	-0.08	0.54
	(0.47)	(0.12)	(0.38)
Capital	-5.54	1.20	-12.25
	(17.69)	(3.06)	(15.19)
Access	3.03	0.66	1.49
	(2.10)	(1.16)	(1.15)
Outside	-2.48	0.09	-2.63
	(2.58)	(0.56)	(1.94)
Distance	-0.03	-0.00	-0.02
	(0.02)	(0.00)	(0.02)
Wx			
Population	1.06	-0.12	1.05+
	(0.67)	(0.21)	(0.55)
Pop. density	-13.20*	-5.76*	-4.30
	(6.19)	(2.65)	(4.45)
Nightlights	0.52	0.31	-0.25
	(0.80)	(0.58)	(0.28)
Rainfall deviations	-0.75	0.08	-0.68
	(0.61)	(0.15)	(0.49)
Capital	-69.30	-3.77	-61.54
	(75.75)	(10.46)	(69.33)
Access	3.81	3.43*	1.57
	(4.44)	(1.56)	(2.41)
Outside	30.76	1.24	29.39
	(22.24)	(5.62)	(19.00)
Distance	-0.02	0.01	-0.03
	(0.03)	(0.00)	(0.03)
R-squared	0.03	0.09	0.00
N	699	699	699

$^{+}p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

TABLE 9d.

Durbin model: residual ethnicity

	(1) N. Events	(2) Riots	(3) Violence
Resid. Ethnicity- X			
Lambda	0.29***	0.52***	0.24***
	(0.04)	(0.10)	(0.04)
Population	0.99	0.82+	0.05
	(0.69)	(0.44)	(0.34)
Pop. density	10.88	2.37	7.43
	(9.45)	(1.97)	(8.11)
Nightlights	0.17	0.22	-0.12
	(0.46)	(0.14)	(0.38)
Rainfall deviations	0.54	-0.04	0.51+
	(0.37)	(0.11)	(0.30)
Capital	-13.34	0.61	-18.48
	(19.78)	(2.85)	(17.44)
Access	2.85	0.62	1.60
	(2.21)	(1.25)	(1.23)
Outside	-3.36	-0.27	-2.81
	(2.43)	(0.55)	(1.80)
Distance	-0.03	0.00	-0.03
	(0.03)	(0.00)	(0.02)
Wx			
Population	1.29+	-0.17	1.33*
	(0.73)	(0.23)	(0.62)
Pop. density	-13.19**	-5.05**	-6.04+
	(4.88)	(1.87)	(3.30)
Nightlights	0.77	0.57	-0.16
	(0.74)	(0.53)	(0.28)
Rainfall deviations	-0.92+	0.02	-0.72+
	(0.50)	(0.14)	(0.37)
Capital	49.17	-0.99	49.89
	(86.46)	(7.39)	(79.69)
Access	3.16	2.54	1.80
	(4.00)	(1.85)	(2.34)
Outside	12.71	-5.28	19.39
	(14.82)	(4.12)	(12.85)
Distance	-0.02	-0.00	-0.03
	(0.03)	(0.01)	(0.02)
R-squared	0.04	0.09	0.00
N	694	694	694

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$