California Center for Population Research
University of California - Los Angeles

Death and the Media: Asymmetries in Infectious Disease Reporting During the Health Transition

Dora Costa and Matthew Kahn

PWP-CCPR-2015-007

October 8, 2015

California Center for Population Research
On-Line Working Paper Series

Death and the Media: Asymmetries in Infectious Disease Reporting During the Health Transition

by

Dora L. Costa
UCLA and NBER
costa@econ.ucla.edu

Matthew E. Kahn
UCLA and NBER
mkahn@ioes.ucla.edu

October 8, 2015

We thank Hoyt Bleakley, Ed Glaeser, Rick Hornbeck and seminar participants at the 2015 AEA Meetings, the 2015 NBER Summer Institute, Stanford University, UCLA, Washington University, St. Louis and Yale University for helpful comments. Dora Costa gratefully gratefully acknowledges the support of NIH grant P01 AG10120 and the use of facilities and resources at the California Center for Population Research, UCLA, which is supported in part by NICDH grant R24HD041022.

Death and the Media: Asymmetries in Infectious Disease Reporting During the Health Transition

Abstract

JEL Classifications: I19, L82, N31

In the late 19th Century, cities in Western Europe and the United States suffered from high levels of infectious disease. Over a 40 year period, there was a dramatic decline in infectious disease deaths in cities. As such objective progress in urban quality of life took place, how did the media report this trend? At that time newspapers were the major source of information educating urban households about the risks they faced. By constructing a unique panel data base, we find that news reports were positively associated with government announced typhoid mortality counts and the size of this effect actually grew after the local governments made large investments in public goods intended to reduce typhoid rates. News coverage was more responsive to unexpected increases in death rates than to unexpected decreases in death rates. Together, these facts suggest that consumers find bad news is more useful than good news.

Dora L. Costa
UCLA Department of Economics
9272 Bunche Hall
Los Angeles, CA 90095-1477
and NBER
costa@econ.ucla.edu

Matthew E. Kahn
mkahn@ioes.ucla.edu

How does the media report news? Competitive pressures should drive the media to provide readers with the information they want, be it useful, entertaining, or shocking (e.g. Gentzkow and Shapiro 2008). By turning statistics into stories, the media can make acquiring information less costly for consumers (Dyck, Moss, and Zingales 2013). But a focus on the shocking or editors' desire to run campaigns can lead to biased coverage, i.e. coverage that does not match the statistics. Coverage may also be biased if consumers want to respond immediately to certain types of news, for example, if bad news requires quick action whereas good news does not.

We investigate how the media report news by examining how the media reacted to changes in typhoid death and case rates in the United States between 1890 and 1938. One advantage to examining typhoid is that knowledge of trends was highly valuable. In the early twentieth century, typhoid fever, a water-borne illness for which there was no cure and which killed 10-20% of its victims, ravaged US cities. Individuals thus had a clear self-interest in knowing what typhoid trends were so they could protect themselves against outbreaks, and by the 1890s they knew how to do so. A second advantage to examining typhoid is that because there were sharp declines in typhoid case and death rates after clean water interventions (Alsan and Goldin 2015; Cutler and Miller 2005; Troesken 1999), we can examine new reports under very different mortality regimes. We also can investigate how the media reports progress. For example, five years prior to the filtration and chlorination of Philadelphia's water, weekly deaths from typhoid fever averaged 1.09 per 100,000. Five years after both filtration and chlorination, weekly deaths averaged 0.15 per 100,000. Typhoid rates in Philadelphia, which drew its water from the contaminated river, were unusually high. But even in New York City, where some but not all areas had access to clean water, weekly death rates from typhoid were 0.33 per 100,000 prior to the construction of the New Croton Dam. After the further construction of the Ashokan Reservoir and Catskill Aqueduct, weekly death rates fell to 0.06 per 100,000.[1]

[1] Mortality rates are estimated from the data used in the paper.

We have created a database of weekly counts of articles mentioning typhoid from major US newspapers from 1890 to 1938 to document how news reports responded to weekly death rates in 6 US cities. In this time period newspapers were the major source of information educating urban households about the disease risks they faced. We combine our newspaper data with weekly data for typhoid death and case rates. These data are available for typhoid but not for other diseases.

We find that although news reports were positively associated with mortality and case rates, coverage was biased (i.e. , it was not determined purely by death and case rates). The responsiveness of news reporting to changes in typhoid mortality and case rates differed before the clean water interventions compared to after the clean water interventions. News coverage also was more responsive to unexpected increases in death rates than to unexpected decreases in death rates. Several mechanisms could explain our results. Although we cannot distinguish between them, all of the mechanisms emphasize that what mattered was how useful the information was to consumers. After the clean water interventions, individuals probably cut back on costly self-protection actions. Knowledge of disease outbreaks may thus have been more valuable after the interventions because the returns to self-protective actions were greater. Knowledge of disease outbreaks may also have been more valuable after the intervention if the stigma of a disease transmitted through contaminated fecal matter was greater after the intervention, thus making costly self-protection measures even more valuable. After the interventions, both the mean and variance of death and case rates fell, thus making any information more informative in a classical signal extraction model. Our findings on asymmetries in reporting are consistent either with high gains in survival probabilities to making the correct self-protection choices or with prospect and psychological theories of reference points in which individuals react more to losses than to gains.

Our focus is on how the media reacted to changes in typhoid death rates, but editors' desire to nudge politicians on public health expenditures and readers on private measures of self-protection may lead to "over-reporting." We therefore examine the editorial pages to determine if newspa-

pers were running public health campaigns. Newspaper campaigns in 1894 and 1895 contributed to the public acceptance of diphtheria antitoxin and to public funding for antitoxin (Hammonds 1999 93-117). Case studies suggest that recent media campaigns have reduced smoking, cocaine use by teenagers, HIV infection rates, and deaths from Reye's syndrome (Hornik 2008) but that the media also spread sensationalist misinformation about vaccines (Freed, Katz, and Clark 1996). Several different research designs have yielded clear evidence of the impact of media on behavior. Outcome variables have included voter turnout (Gentzkow 2006; Gentzkow, Shapiro, and Sinkinson 2011), voting outcomes (DellaVigna and Kaplan 2007; cf., Gentzkow, Shapiro, and Sinkinson 2011), the voting of US Congressmen (Dyck, Moss, and Zingales 2013); disaster relief contributions (Eisensee and Stromberg 2007), and corruption (Larreguy, Marshall, and Snyder (2014). We therefore also present suggestive evidence on whether news reports led consumers to take more precautions.[2]

1 Demand and Supply of Public Health News

We assume that consumers demanded typhoid information and that the media will provide readers with the information they want (Gentzkow and Shapiro 2008). In recent times the number of newspaper articles on topics of concern to consumers such as crime, inflation, and disease closely track self-reports of concern in polls and ameliorative actions by consumers (Lowenstein and Mather 1990). Why did consumers demand typhoid information?

Typhoid spread primarily through drinking water contaminated with the wastes of infected individuals. Other modes of transmission were direct contact with a contaminated privy, with the wastes of a typhoid patient, with food prepared by a typhoid carrier, or indirect contact with a contaminated privy through flies. Precautions individuals could take included using individual

[2]We do not have the geographic variation to examine how the media affected municipal expenditures on public health. Our focus is on media responses to changes in death rates.

water filters, bringing water to a roiling boil, pasteurizing milk, thoroughly cooking all vegetables, peeling fruit, disinfecting privies and homes, and sealing privies and homes from flies.

An individual's probability of survival P thus depends on self-protection measures, S, such as filtering or boiling water and on water, W_s, in state s. We are assuming that there is a stochastic component to water quality. Water quality is a random variable, ranging from polluted to clean. Individuals update their subjective assessment of the severity of the water pollution risk based on announced death and case rates reported in news reports. In our empirical work below, we assume that such news reports are urbanites' main source of information concerning the evolving threat of infectious disease.

We also assume that health, H depends on self protection measures and on water in state s, $H(S, W_s)$ and that self-protection and clean water are substitutes. Following Ehrlich and Becker (1972), the consumer's utility in state of world s is weighted by the survival probability

$$P(S, W_s) \times U(H(S, W_s), C) \tag{1}$$

and must satisfy the budget constraint

$$I = p_S S + p_W W_s + p_c C. \tag{2}$$

where I is income and p_S, p_W, and p_c are the prices of self-protection, water, and consumption goods. In the linear case, we can re-write the probability of survival as an index function

$$Y^* = \gamma_1 S + \gamma_2 W_s \tag{3}$$

where an individual survives if Y^* is greater than 0.[3] News stories convey information about recent

[3] We are modeling a household's self protection choice as if its location within a city is not a relevant factor. In a more realistic model, households would select a residential neighborhood and neighborhoods would differ with

changes in water quality and this should affect household self-protection levels.

During the time period we study, many cities made major investments in water treatment and other public goods with the intent of reducing infectious disease. Households are assumed to be aware of the dates of these investment regime shifts. Such local public goods investment should reduce the probability that a water pollution outbreak occurs. Anticipating this fact, such public investments may crowd out private investment in costly self-protection. Because H is concave, the returns to increasing self-protection are greater at lower levels of self-protection. Information about poor water quality is more valuable after the clean water interventions because at lower levels of self-protection, the rates of return to increasing self-protection measures is greater. Such information also may be more valuable after the clean water interventions because individuals can better interpret deviations from trend. A change in mortality rates, whether high or low, represents a sharper deviation from trend in the low level and low variance regime which prevailed after the clean water intervention (the classical signal extraction problem). This phenomenon has been noted in the literature about inflation expectations, where disagreement about the future path of inflation tends to rise both with inflation and with sharp inflation changes (Mankiw, Rise, and Wolfers 2004). Alternatively, if readers want sensationalist stories newspapers will focus on the unusual and thus over-emphasize low-risk causes of death (as found in Frost, Frank, and Maibach 1997).

Our model could be modified to account for the stigma or fear effects of a disease. Because typhoid was transmitted through the fecal-oral route, stigma may have been greater after the intervention when the disease was rarer. For example, there is a cancer premium for the value of a statistical life either because cancer is a dreaded disease or because of the accompanying morbidity (Viscusi, Huber, and Bell 2014).

respect to their disease risk. Real estate rents should be lower in high risk neighborhoods. Once a person has selected a neighborhood, one's infection risk would be a function of overall water quality, neighbor self-protection investments and one's own investments. Self protection by neighbors would be especially important in cases where people live in high density areas.

Our model also implies that "bad" news is more important than "good" news. "Bad" news will lead individuals to revise upwards their probabilities that the water supply is dirty and thus that the media should focus on "bad" rather than "good" news. Because there is a high gain in the probability of survival if individuals make the correct self-protection choices, accurately weighting "bad" events is more important than accurately weighting "good" events. Newspapers thus have an incentive to focus on unexpected disease outbreaks. Such news will sell more copies and thus also raise advertising revenue.

Both prospect theory and psychological theories also imply that individuals react more to an increase in mortality rates than to declines and thus that the media should focus on "bad" rather than "good" news. Kahneman and Tversky (1979) argued that individuals care more about loss in utility than gain. The psychology literature argues that asymmetries arise because of differences in perceptions and because individuals are mildly optimistic. If impressions are based on reference points, the loss is felt more keenly (Helson 1964). If more attention is given to new or novel information, which is extreme information, then negative information is given more weight (Fiske 1980). In examining inflation expectations Carroll (2003) found that not only did the volume of news matter, but also news that represented sharp and negative break from the past.

We therefore hypothesize that

1. An increase (decrease) in typhoid death or case rates will lead to more (less) news reports.

2. An increase (decrease) in typhoid death or case rates will lead to more (less) news reports about typhoid after clean water interventions than before these interventions.

3. An change in typhoid death or case rates will have a bigger impact on news reports when the change is an unexpected increase.

2 Econometric Specifications

As discussed above, we seek to document whether the urban media was responsive to changes in ”objective reality”. Put simply, when the death count increased from infectious disease, did the media cover the story? We then seek to test whether the media's response differs before and after the major local public health interventions. To study this, we estimate count models of how the number of weekly news reports (r) depends on current weekly death or case rates (d), the clean water intervention (I), and the interaction between the clean water intervention and the death rate. In a linear form, we have

$$r = \gamma_1 d + \gamma_2 I + \gamma_3 (d \times I).$$

Our second hypothesis, that the media reacts asymmetrically to increases and decreases in death rates, implies that the the number of weekly news reports depends on the unexpected change in expected death or case rates (D) and on whether this change is an unpleasant surprise,

$$r = \delta_1 D + \delta_2 (D \times (\text{Dummy=1 if bad news}).$$

We adopt a simple forecasting model to determine whether “good” and “bad” news are treated asymmetrically. Urban households having read past newspapers are aware of past broad trends in typhoid death and case rates. The media in supplying news knows that a deviation from past trends, particularly an unexpected increase in deaths, will interest its readers. For example during a time when typhoid death rates are declining, it may not be ”new news” that typhoid death rates are low. In such a setting, new ”bad news” would be if the typhoid death rate in that week is larger than would be expected given the recent time trend. We test whether the media was more responsive to such ”unexpected bad news”.

To operationalize our explanatory variable measuring ”new news”, we assume that deviations from trends are determined in one of two ways.

1. We treat a surprise in death or case rates as a deviation from a two year rolling average of past death or case rates. Thus, in current time period i, the deviation from death rate trend (D) is

$$D_i^0 = d_i - \hat{d}_i \tag{4}$$

   where $\hat{d}_i$ is the predicted death rate (from a two year rolling average) and d_i is the death rate.

2. Alternatively, a surprise in death or case rates is a deviation from a trend determined by de-trending death or case rates, accounting for intercept changes caused by clean water interventions. The death rate is predicted at each date i by running an ARMAX regression (MA(1), AR(1)) using all prior death or case rates (from 2 years of data to all years in the final period),

$$\hat{d}_i = \alpha + \sum_{k=0}^{n} \delta_k I_k$$

   where d is the death rate and I_k is a set of dummy variables indicating that intervention k has occurred.

We specify the relationship between the count of articles and death rates or unexpected deviations in death rates using a negative zero inflated binomial model to account for excess zeros and over-dispersion. Assume that the observed count of articles y_i is the product of two latent variables,

z_i and y_i^*,

$$y_i = z_i y_i^*$$

where z_i is binary variable with values 0 or 1, and y_i^* has a negative binomial distribution. Then,

$$\begin{aligned} \Pr(y_i = 0) &= \Pr(z_i = 0) + \Pr(z_i = 1, y_i^* = 0) \\ &= q_i + (1 - q_i) f(0) \\ \Pr(y_i = k) &= (1 - q_i) f(k), k = 1, 2, ... \end{aligned}$$

where q_i is the probability of no article and $f(\cdot)$ is the negative binomial probability distribution for y_i^*. We model the binary process z_i using a logit model. We perform Vuong tests to determine if the excess number of zeros leads us to prefer a zero-inflated negative binomial model to a standard negative binomial model (a statistically significant statistic suggests yes). If we reject the negative binomial model in favor of the zero-inflated negative binomial model, we will then test whether the dispersion parameter (α) is 0 (or the logarithm of α is negative infinity). A statistically insignificant dispersion parameter suggests that we instead should be using a Poisson model. We estimate our zero-inflated negative binomial regression models with robust standard errors, clustered on the city.

We specify the logit part of the zero-inflated negative binomial regression as

$$\Pr(y = 0) = L(\text{dummy=1 if news event, dummy=1 if holiday week, city and year fixed effects}) \quad (5)$$

We specify the negative binomial part of the zero-inflated binomial regression model in three different ways. In our first specification, Equation 4 below, we examine differential reactions to typhoid death or case rates before and after the clean water interventions. Our specification includes typhoid death or case rates (d), two clean water interventions (I_1 and I_2), and interactions

between deaths rates and the clean water interventions.

$$\Pr(y = k) = F(d, I_1, I_2, I_1 \times d, I_2 \times d, \text{number of total articles}) \tag{6}$$

When possible (i.e. when convergence was not an issue), we also control for city and year fixed effects.

In our other two specifications, Equations 5 and 6 below, of the negative binomial part of the zero-inflated binomial regression model, we examine differential reactions to better and worse than expected typhoid death or case rates. We include either D^0 or D^1, our deviations from expected death or case rates specified in Equations 1 and 2, the interaction between either D^0 or D^1 and a dummy variable indicating whether D^0 or D^1 are positive (and thus death or case rates are greater than expected),

$$\Pr(y = k) = F(D^0 \times (\text{Dummy=1 if } D^0 > 0), \text{number of total articles}) \tag{7}$$

$$\Pr(y = k) = F(D^1 \times (\text{Dummy=1 if } D^1 > 0), \text{number of total articles}). \tag{8}$$

We present average marginals of death or case rates and the interaction of death and case rates for all of our regressions, following the procedures outlined in Lietgöb (2014). Thus when the inverted link function $E(y|x)$, is $\exp(\gamma_1 d + \gamma_2 I_1 + \gamma_3 I_2 + \gamma_4 (d \times I_1) + \gamma_5 (d \times I_2))$, we present estimates of $\frac{\partial E(y|x)}{\partial d}$, $\frac{\partial E(y|x)}{\partial d \partial I_1}$, and $\frac{\partial^3 E(y|x)}{\partial d \partial I_1 \partial I_2}$.[4]

[4] In a non-linear model interaction effects are dependent on covariates and thus differ across individuals and may even have different signs for individuals. We present mean interaction effects. In no cases did we obtain different signs of the interaction effects for individuals.

3 Data

We created a panel data set from newspaper articles and from weekly typhoid death and case rates for New York City, Baltimore, Boston, Chicago, Washington DC, Philadelphia. Weekly deaths and cases for New York City are from our digitization of Emerson and Hughes (1941), which provides continuous data from 1890 until 1938. Weekly deaths and cases for our other cities are from Project Tycho (https://www.tycho.pitt.edu/), which digitized data from the weekly national publication, *Public Health Reports*. These data are incomplete; the number of cases only begins to be published in 1906 and both deaths and cases are more likely to be missing once typhoid deaths have fallen to close to zero. Deaths are available up to 1932. We have used data from the published censuses of population to estimate yearly city populations (adjusted for city annexations of neighboring communities) and thus yearly death and case rates.

We obtained daily counts of the total number of newspaper articles and the number of newspaper articles mentioning typhoid and also typhoid and the city using mechanized searches of *The New York Times*, *The Baltimore Sun*, *The Boston Globe*, *The Chicago Tribune*, *Washington Post* and *The Philadelphia Inquirer*.[5] (For a comparison of mechanized and manual searches see Appendix A.) These were the major "serious" newspapers within each city (their rivals have not been digitized and indexed).[6]

We aggregated our daily counts to the weekly level. Reports include all types of news, including reports from local public health officials, stories of outbreaks, society news, obituaries of well-known individuals, editorials, and appeals to charity.

Cities had to have both digitized and indexed newspapers and good weekly typhoid death data to be included in our panel data set. Our final panel data set has data for New York City for all

[5]The first five newspapers are available from Proquest Historical Newspapers. *The Philadelphia Inquirer* is available from Readex America's Historical Newspapers.

[6]At least in the case of political bias, there is no evidence that the party in power affected the partisan composition of the press in this time period (Gentzkow, Petek, Shapiro, and Sinkinson 2015). Newspapers were becoming more informative (Gentzkow, Glaeser, and Goldin 2006).

weeks for 1890-1938, and, with some weeks missing, for Chicago for 1896-1932, Baltimore for 1900-1932, Boston for 1890-1932, Philadelphia for 1901-1922, and Washington DC for 1890-1932. We also created dummy indicators for a holiday during that specific week and for a major news event that week. What constituted a major news event was a judgement call. Recurring events such as the day after elections and the World Series were labeled major news events, as were outbreaks of war and major war events, natural disasters, New York City ticker tape parades and the events meriting these parades, new world records, and famous trials, murders, and kidnappings.

4 Typhoid Death and Case Rates

The major interventions in each of our cities (see Table 1) take the form either of cleaning up the water supply obtained from the nearby river through chlorination or filtration or of obtaining new, clean sources of water. For each city we could identify two interventions from Cutler and Miller (2005) for Chicago, Baltimore, and Philadelphia and from histories of local water supply systems for New York City, Boston, and Washington, DC. (Although we could identify a third for New York City, the effect of this intervention was negligible.)

Figure 1, which also shows missing data, and Table 2 suggest that the interventions were effective in lowering typhoid mortality and case rates. In the sample as a whole, typhoid death rates per 100,000 were 0.8 prior to any intervention, 0.4 after the first intervention but before the second, and 0.1 after the second intervention. Prior to the first intervention, death rates per 100,000 varied widely across cities with highs of 1.0 and 1.5 in Philadelphia and Washington, DC, respectively and a low of 0.4 in New York City. After both interventions, death rates per 100,000 varied from a high of 0.2 in Baltimore to a low of 0.03 in New York City. Case rates also fell after an intervention and converged across cities. Clean water interventions were statistically significant negative predictors of death rates, controlling for a year trend (see Appendix B). Fatality rates, however, did

not fall sharply. For example, in both New York City and Boston, there were roughly 24 and 21 deaths per 100 cases, respectively, between 1900 and 1909 and 15 and 19 deaths per 100 cases, respectively, after 1919.

5 Results: Death and the Media

Figure 2, which shows smoothed plots of typhoid death rates and of the percentage of typhoid articles, suggests that while on the whole reports of typhoid followed mortality patterns, with more reporting in a high mortality regime than in a low mortality regime, an increase in city death rates led to more news reports in a low than in a high mortality regime. For example, in New York City, the up-tick in typhoid mortality rates in the 1920s is associated with an increase in news reports that is greater than the increase in the early 1890s when typhoid mortality rates spiked up higher. The increases in reporting that are not related to city death rates were often associated with world events such as concern over typhoid epidemics during the Spanish-American War and World War I.

We find that the media are more likely to report changes in typhoid after the clean water interventions than before. Table 3 shows that increases in typhoid death rates increase reporting both pre- and post-intervention but there is a stronger positive effect after both interventions. Prior to any intervention, a half standard deviation increase in post first intervention typhoid death rates (0.141) leads to an increase of at least 0.05 in the differences of the logarithms of the expected count of typhoid articles (=0.141 $\times$ 0.335). After both interventions, this half standard deviation increase of 0.141 in death rates yields an additional increase in the difference of the logarithms of the expected count of typhoid articles of 0.06 (=0.141 $\times$ 0.417) to 0.20 (=0.141 $\times$ 1.422). Although the interaction between death rates and the second intervention is not statistically significant when both the count and zero article part of the negative binomial include both city and fixed effects, the joint effect of both interventions interacted with the death rate becomes statistically significant at

the 10% level when Philadelphia (which has few observations after the intervention) is excluded. The increase in the count of the logarithm of typhoid articles is 0.587 ($\hat{\sigma}$ =0.341).

Media responses to death rates after the two interventions are always statistically significant when we examine local typhoid articles, which excludes Philadelphia for which we have no data (see Table 4. A half standard deviation increase in the death rate leads to an increase of at least 0.05 (0.141 $\times$ 0.366) in the difference in the logarithms of expected articles counts and an additional increase of at least 0.09 (0.141 $\times$ 0.651) after both interventions.

Table 5, which gives the marginal effects, implies that this half standard deviation increase of 0.141 in death rates increases the number of all typhoid articles by up to 0.311 (=0.141 $\times$ 2.206) and by at least 0.211 (=0.141 $\times$ 1.499). The first intervention has a small and statistically insignificant effect on the relationship between death rates and typhoid articles. But, the second intervention increases this relationship by 0.64 (0.141 $\times$ 4.570) to 0.18 (0.141 $\times$ 1.300), a 22 to 6% increase relative to the mean number of articles. In the case of local typhoid articles, this half standard deviation increase in death rates increases the number of local articles by at least 0.090 (=0.141 $\times$ 0.636), an increase of 9% relative to the mean number of articles. The two interventions change this average marginal by at least 0.11 (0.141 $\times$ 0.810), an increase of 10% relative to the mean number of articles.

We find that the media respond more to “bad” than to “good” news. Our specifications for mortality expectations were based on deviations from a 2 year rolling mean and on deviations from the predictions of an ARMAX model. These specifications show that an unexpected increase in death rates (a positive residual) leads to more news reports than an unexpected decrease in death rates (see Tables 6). When mortality rates are higher than expected, a half standard deviation change in the deviation from expected death rates based on a two year rolling mean (0.192 over all time periods) leads to a statistically significant increase of 0.18 (0.192 $\times$ 0.934) in the differences of the logarithms of the number of articles. When mortality rates are lower than expected, the

effect on the differences of the logarithms of the number of articles is statistically insignificant. When we assume that expectations are formed based on an ARMAX prediction, we find that a half standard deviation change in the deviation from expected death rates leads to an increase of 0.14 (0.136 $\times$ 1.005) in the differences of the logarithms of the expected number of articles. An unexpected increase in death rates also leads to more local news.

Table 7, which presents the death rate deviations marginals, shows that an increase of a half standard deviation in the deviation from expected death rates, as estimated by the ARMAX regression, yields an additional 1.17 (=0.136 $\times$ 8.565) total articles and an additional 0.42 local articles, respectively 40 and 38% increases relative to the mean number of articles. A similar calculation using the rolling mean regression yields similar results: increases of 44 and 32% relative to the mean number for total articles and local articles, respectively.

Controlling for case rates, newspapers were more likely to report on typhoid after the clean water interventions, when case rates were lower. We found statistically significant effects of both intervention case rate interactions, even controlling for year fixed effects in both the count and the logit part of the negative binomial (see Tables 8 and 9). After the first intervention, the standard deviation of case rates was 1.682. A half standard deviation increase in case rates increased the logarithm of the count of expected articles after both interventions by at least an additional 0.10 (=0.841 $\times$ 0.114) relative to the pre-intervention period and the number of articles by an additional 0.25 (=0.841 $\times$ 0.294), a 35% increase relative to the mean. The increase in the number of local articles was at least 0.10 (=0.841 $\times$ 0.114), a 12% increase.

We also found that newspapers were more likely to report on typhoid when case rates deviated from expected case rates, but that there was an asymmetry between good news and bad news (see Tables 10 and 11). A half standard deviation decline in the deviation from expected case rates calculated from a rolling mean (0.617 over all time periods) leads to a statistically significant increase of 0.10 (=-0.617 $\times$ -0.163) in the logarithm of the expected total number of articles when

the news is good and a statistically significant increase of 0.18 (=0.617 $\times$ 0.288) when the news is bad. The count of total articles decreases by 0.088 (=-0.617 $\times$ 0.142) when the news is good and increases by an additional 1.76 (=0.617 $\times$ 2.858) when the news is bad. When expected case rates are modeled using an ARMAX specification, the total count of articles increases by 0.04 (-0.329 $\times$ -0.126) when news is good and by 0.93 (=0.329 $\times$ 2.821) when news is bad. We observe the same asymmetry when we examined the number of local articles.

A newspaper article on typhoid was more likely when there was a news event. The effect was statistically significant for all news (see Tables 3, 6, 8, and 10), but not for local news. Because large news events such as wars or natural disasters were associated with typhoid outbreaks (or fear of such outbreaks), we interpret this effect as dominating the displacement of typhoid news from a sensational trial or murder case. In some of our specifications, holidays also had a statistically significant, positive effect on news reports. We would expect a positive effect either if articles about typhoid can be written ahead of time or if appeals to charity are more likely during holidays.

6 Newspapers' Response to Death Rates

Was newspaper reporting on typhoid fever determined by editors' campaigns favoring the adoption of clean water technologies? If yes, we would expect more editorials on typhoid prior to the intervention. This "campaign" effect would therefore counteract our first hypothesis.

Figure 3 shows that only in Chicago was there a large number of editorials prior to the first intervention – the closing of the sewer outfalls of Lake Michigan. There were many editorials on the incompetence of the city for delays in sanitary reforms. The upsurge in editorials in New York City and in Boston prior to the sanitary reforms was associated with worries about typhoid during the Spanish-American War and the building of the Panama Canal. The editorials in Washington DC after the first intervention were largely articles about clean water interventions in other cities. The

editorials in Boston after the second intervention focused on the dangers of typhoid on vacation. There were no editorials in Baltimore about typhoid.

Intensive news reporting, by changing private behavior, could cause reductions in future death rates. We therefore investigate whether the past number of articles about typhoid had an effect on changes in typhoid death rates. We regress the difference in death rates from the past week, multiplied by 100 and where d is the death rate, on the number of past articles, y, controlling for city fixed effects (C), year fixed effects (Y), and week fixed effects (W) and instrumenting for the sum of past articles using the sum of past non-local articles,

$$(d_i - d_{i-1}) \times 100 = \beta_0 + \beta_1 + \sum_{k=i-2}^{i-n} y_k + \beta_2 C + \beta_3 Y + \beta_4 W \tag{9}$$

$$(d_i - d_{i-1}) \times 100 = \beta_0 + \beta_1 + \sum_{k=i-2}^{i-n} y_k + \beta_2 C_i + \beta_3 Y_i + \beta_4 W_i + \sum_{k=i-2}^{i-n} \delta_k d_k. \tag{10}$$

We also run instrumental variables specifications controlling for past lagged death rates, excluding the first lag which is part of the dependent variable.

We find that the number of past news articles reduces the current relative to the past week's death rate, i.e. the coefficient is negative (see Table 12). The effects are statistically significant when we do not control for lagged death rates. However, controlling for lagged death rates, we lose statistical significance both because the sample size falls as we include past lags (recall that data are missing for cities other than New York City) and because there is convergence to the mean in death rates. Regardless of statistical significance, the magnitude of the coefficient on the number of news reports (-0.165 to -0.231) in the two past weeks is substantial relative to the mean of 0.078 for the dependent variable. Although the impact of news reports over more than two weeks is smaller, some of the effects are still large relative to the mean of the independent variable. We interpret our results as suggestive of an impact of news reporting on self-prevention measures.

7 Robustness: Diphtheria and the Media in New York City

We test the robustness of our findings by comparing results for typhoid articles in New York City with diphtheria articles. Diphtheria, an upper respiratory tract infection, is spread through physical contact or breathing the aerosolized secretions of infected individuals. In New York City, diphtheria death rates fell rapidly once antitoxin became widely available in 1895 (it was provided by the City). We would therefore not expect a positive effect of the interaction between typhoid death rates and typhoid interventions on the number of diphtheria articles. However, we would expect a more positive response to death rates after diphtheria antitoxin became widely available.

Tables 13 and 14 show that in New York City the effect of typhoid death rates on all and local typhoid news reports is greater after the second intervention but that there is no effect on diphtheria news reports. Diphtheria death rates, however, have a greater effect on diphtheria news reports once antitoxin become widely available. In New York City, the standard deviation of typhoid deaths after the first clean water intervention was 0.096 and the standard deviation of diphtheria deaths was 0.563 after antitoxin became widely available. A half standard deviation increase in typhoid deaths led to an increase of 0.26 (=0.048 $\times$ 5.374) in the total number of articles and an increase of 0.134 (=0.048 $\times$ 2.788) in the number of local articles, with an additional increase of 0.44 total articles (=0.048 $\times$ 9.230) and 0.25 local articles (=0.048 $\times$ 5.218) after both interventions. These last two increases were, respectively, increases of 33% and 25%. Note that the interaction effects of typhoid interventions with the typhoid death rate had a statistically significant negative effect on the total number of diphtheria articles and a statistically insignificant effect on the number local diphtheria articles. However, after 1895, when diphtheria antitoxin became widely available in the New York City, a half standard deviation increase in the diphtheria death rate increased the number of local diphtheria news reports by an additional 0.05 (=0.282 $\times$ 0.184), a 10% increase.[7]

Our hand-collected data for New York City for 1900=1930 allows us to investigate the use of

[7] We could not estimate a similar specification for all articles because of convergence problems.

alternative dependent variables. These included dummies equal to one if the article was largely about typhoid at the time of writing, if the article was largely about local typhoid at the time of writing, and if the article was a high profile one (either on the title page, more than half a page in length, or with typhoid in the title). In all cases the interaction between the second intervention and the death rate was statistically significant (results not shown).

8 Conclusion

At a time before television, radio and the Internet, newspapers played a central role in disseminating information and thus guiding their readers' choices and viewpoints. The availability of weekly typhoid death and case reports permits us to test how major urban newspapers responded to emerging public health trends during a key time in urban history.

Between 1880 and 1940, today's developed countries experienced a major urban public health transition (see Haines 2001 on the US, Kestzenbaum and Rosenthal 2011 on France, and Brown 2000 on the UK and Germany). How does the media cover the emerging story during a time of rapid progress? Studies of newspapers in the economics literature have focused on political bias (e.g. Gentzkow and Shapiro 2008). We instead have focused on news stories about a disease. We found that news reports were positively associated with typhoid death and case rates but that the size of the effect grew after cities cleaned up their water supplies. In addition, we also found that news coverage was more responsive to bad news (i.e., unexpected increases in death rates) than to good news (unexpected decreases in death rates). The losses to individuals of not knowing bad news are likely to outweigh the losses of not knowing good news. Not knowing good news may lead to too much time and money spent on self-protection. Not knowing bad news may lead to death if not enough time and money is spent on self-protection.

Our findings also have implications for the economic incidence of urban public health im-

provements. If improvements are common knowledge both to incumbent city residents and to non-residents, then standard no arbitrage compensating differentials logic implies that landowners in the cities and neighborhoods that experienced the largest reduction in death would enjoy the windfall of higher prices. But, if outsiders are unaware of the localized quality of life improvements then incumbent renters could gain the windfall. If the media devoted extra coverage to deaths during good times, then the public may perceive a higher risk of infectious disease risk than was actually present, thus leading to rents for incumbents.

References

[1] Alsan, Marcella and Claudia Goldin. 2015. Watersheds in Infant Mortality: The Role of Effective Water and Sewerage Infrastructure, 1880-1915. Unpublished MS, Harvard University. http://scholar.harvard.edu/goldin/publications

[2] Brown, John C. 2000b. Wer bezahlte die hygienische Stadt? Finanzielle Aspekte der sanitären Reformen in England, USA, und Deutschland um 1910 [Who Paid for the Sanitary City? Issues and Evidence Ca. 1910]. In Jörg Vögele and Wolfgang Woelk, eds. Stadt, Krankheit, und Tod. Berlin: Duncker and Humblot: 237-257.

[3] Cutler, David and Grant Miller. 2005. The Role of Public Health Improvements in Health Advances: The Twentieth-Century United States. *Demography*. 42(1): 1-22.

[4] Cutler, David and Grant Miller. 2006. Water, Water Everywhere: Municipal Finance and Water Supply in American Cities. In Glaeser and Goldin, Eds., *Corruption and Reform: Lessons from America's Economic History*.

[5] DellaVigna, Stefano and Ethan Kaplan. 2007. The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics*. 122(3): 1187-234.

[6] Ehrlich, Isaac and Gary Becker. 1972. Market Insurance, Self-Insurance, and Self-Protection. *The Journal of Political Economy*. 80(4): 623-48.

[7] Eisensee, Thomas, and David Strömberg. 2007. News droughts, news floods, and US disaster relief. *The Quarterly Journal of Economics*. 122(2): 693-728.

[8] Emerson, Haven and Harriet E. Hughes. 1941. *Population, Births, Notifiable Diseases, and Deaths, Assembled for New York City, NY: 1866-1938*.

[9] Fiske, Susan T. 1980. Attention and Weight in Person Perception: The Impact of Negative and Extreme Behavior. *Journal of Personality and Social Psychology*. 38(6): 889-906.

[10] Freed, Gary L., Samuel L. Katz, Sarah J. Clark. 1996. Safety of Vacciantions: Miss America, the Media, and Public Health. *Journal of the American Medical Association*. 276(23): 1869-1872.

[11] Frost, Karen, Erica Frank, and Edward Maibach. 1997. *American Journal of Public Health*. 87(5): 842-45.

[12] Gentzkow, Matthew, Edward Glaeser, and Claudia Goldin. 2006. The Rise of the Fourth Estate: How Newspapers Became Informative and Why It Mattered. In Glaeser and Goldin (Eds), *Corruption and Reform: Lessons from America's Economic History*.

[13] Gentzkow, Matthew, Nathan Petek, Jesse M. Shapiro, and Michael Sinkinson. 2015. Do Newspapers Serve the State? Incumbent Party Influence on the US Press, 1869-1928. *Journal of the European Economic Association*. 31(1).

[14] Gentzkow, Matthew and Jesse M. Shapiro. 2008. Competition and Truth in the Market for News. *Journal of Economic Perspectives*. 22(2): 133-54.

[15] Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2011. The Effect of Newspaper Entry and Exit on Electoral Politics. *American Economic Review*. 101(7): 2980-3018.

[16] Haines, Michael. R. 2001. The Urban Mortality Transition in the United States, 1800 to 1940. *Annales De Démographie Historique*. 1: 33-64.

[17] Hammonds, Evelynn Maxine. 1999. *Childhood's Deadly Scourge: The Campaign to Control Diphtheria in New York City, 1880-1930*. Baltimore and London: Johns Hopkins University Press.

[18] Helson, H. 1964. *Adaptation-level Theory*. New York: Harper.

[19] Hornik, Robert C. (Ed.) 2002. *Public Health Communication: Evidence for Behavior Change*. Mahwah, NJ: Lawrence Erlbaum Associaes, Inc.

[20] Kahneman, Daniel and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*. 47(2): 263-92.

[21] Kesztenbaum, Lionel, and Jean-Laurent Rosenthal. 2011. "The health cost of living in a city: The case of France at the end of the 19th century." *Explorations in Economic History*. 48 (2):207-225.

[22] Larreguy, Horacio A., John Marshall, and James M. Snyder, Jr. 2014. Revealing Malfeasance: How Local Media Facilitates Electoral Sanctions of Mayors in Mexico. NBER Working Paper No. 20697.

[23] Leitgöb, Heinz. 2014. Modeling Interactions in Count Data Regression Principles and Implementation in Stata. Slides. German Stata Users Group Meeting. http://www.stata.com/meeting/germany14/abstracts/materials/de14_leitgoeb.pdf

[24] Loud, Ralph W. 1923. The Metropolitan Sewerage Works. *Journal of the Boston Society of Civil Engineers.* 10(8): 325-74.

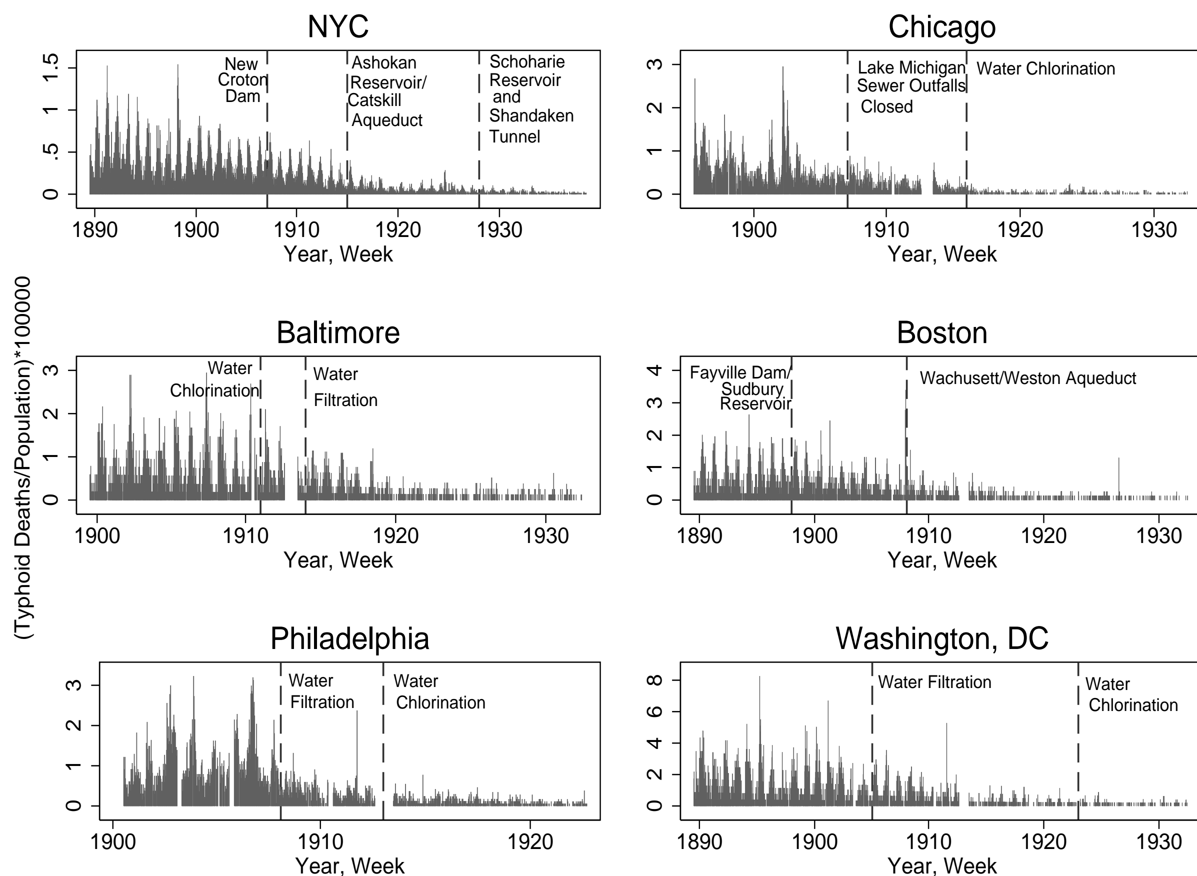
[25] Mankiw, N. Gregory, Ricardo Reis, and Justin Wolfers. 2004. Disagreement about Inflation Expectations. In Gertler and Rogoff (Eds), *NBER Macroeconomics Annual 2003, Volume 18.*

[26] Troesken, Werner. 1999. Typhoid Rates and the Public Acquisition of Private Waterworks, 1880-1920. *Journal of Economic History*. 59(4): 927-948.

[27] Troesken, Werner. 2004. *Water, Race and Disease.* MIT Press, Cambridge, MA.

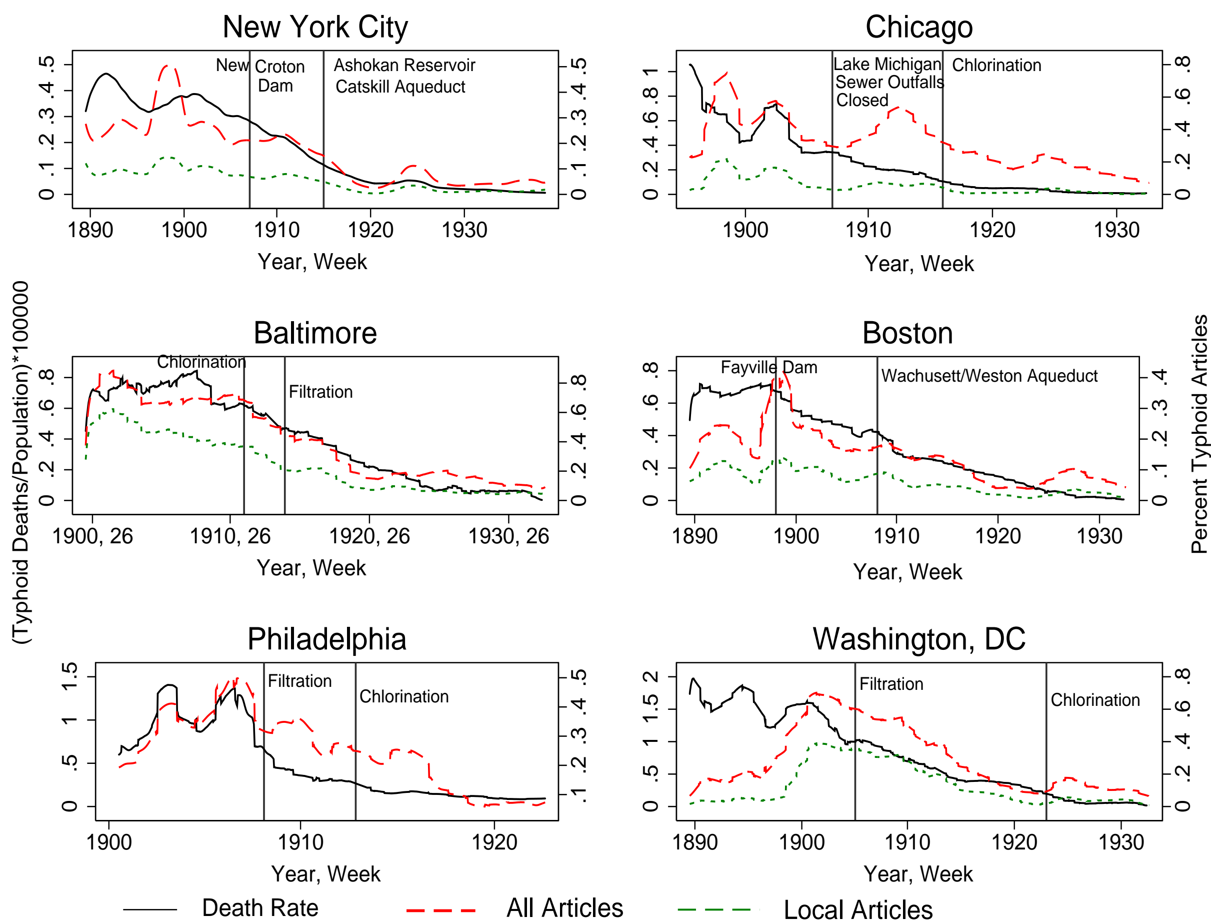
[28] Villarreal, Carlos. 2014. Where the Other Half Lives: Evidence on the Origin and Persistence of Poor Neighborhoods from New York City 1830-2012. Unpublished MS. https://sites.google.com/site/carlosrvillarreal/

Figure 1: Weekly Typhoid Death Rates by City



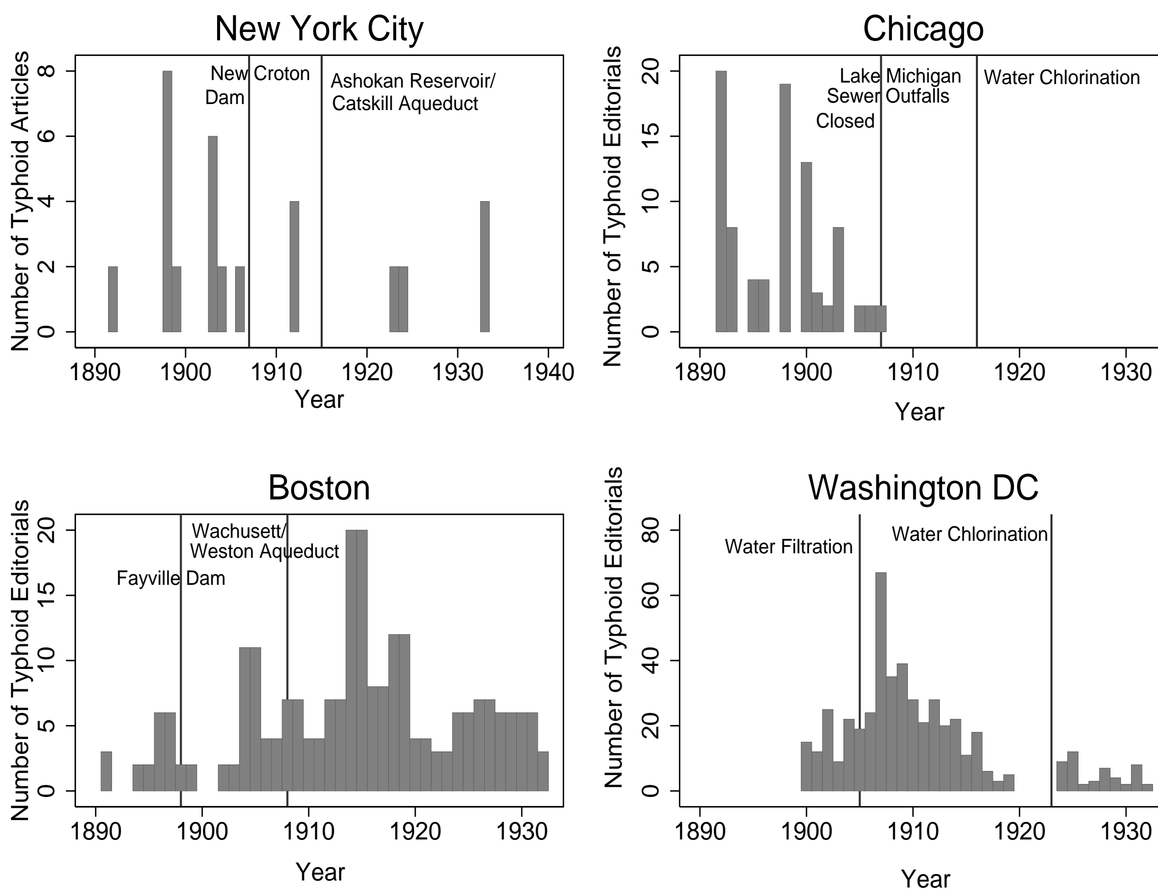
See the text for sources.

Figure 2: Weekly Typhoid Death Rates and Percentage of Typhoid Articles by City



See the text for sources. Death rates and the percentage of articles were smoothed using a lowess estimator.

Figure 3: Number of Typhoid Editorials by City



Editorials could not be identified for Philadelphia. There were no editorials in Baltimore.

Table 1: Clean Water Intervention Dates

City	Intervention	Intervention
Baltimore	1911	1914
	water chlorination	water filtration
Boston	1898	1908
	Fayville Dam/Sudbury Reservoir	Wachusett/Weston Aqueduct
Chicago	1907	1916
	Lake Michigan sewer outfalls closed	water chlorination
New York City	1907	1915
	New Croton Dam	Ashokan Reservoir/Catskill Aqueduct
Philadelphia	1908	1913
	water filtration	water chlorination
Washington DC	1905	1923
	water filtration	water chlorination

Table 2: Mean Death and Case Rates, Before, Between, and After Interventions

	Before First Intervention		Between Interventions		After Second Intervention	
	Death Rate	Case Rate	Death Rate	Case Rate	Death Rate	Case Rate
Baltimore	0.726	4.437	0.490	2.843	0.156	0.878
	(0.525)	(4.921)	(0.311)	(2.410)	(0.198)	(1.282)
Boston	0.687		0.489	3.358	0.110	0.629
	(0.461)		(0.390)	(5.253)	(0.150)	(0.819)
Chicago	0.554	0.471	0.199	1.231	0.026	0.156
	(0.403)	(0.440)	(0.135)	(1.053)	(0.034)	(0.199)
New York City	0.366	1.444	0.185	1.243	0.034	0.266
	(0.231)	(1.112)	(0.118)	(0.972)	(0.040)	(0.292)
Philadelphia	0.971	6.949	0.343	2.141	0.131	0.664
	(0.600)	(5.746)	(0.252)	(1.596)	(0.089)	(0.579)
Washington DC	1.494		0.596	2.662	0.061	0.324
	(1.072)		(0.538)	(3.091)	(0.124)	(0.429)
All Cities	0.765	2.449	0.382	2.067	0.080	0.484
	(0.717)	(3.409)	(0.385)	(2.596)	(0.129)	(0.772)

See text for death and case rate sources. Death and case rates are per 100,000. Standard deviations in parentheses. Case rates are not available for Boston and Washington DC prior to the first intervention.

Table 3: Negative Binomial Regression of Effect of Weekly Typhoid Death Rates on Weekly Newspaper Reports

	(1)		(2)		(3)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of Typhoid Articles						
Death rate	0.335**	0.158	0.386***	0.122	0.350***	0.086
1st intervention	-0.296*	0.151	-0.159	0.098	-0.278***	0.092
2nd intervention	-0.639***	0.067	-0.537***	0.127	-0.121	0.146
1st intervention x death rate	0.107	0.230	0.153	0.166	0.173*	0.092
2nd intervention x death rate	1.316***	0.205	0.617***	0.214	0.244	0.264
Number of total articles	0.014**	0.006	0.006	0.007	0.005	0.006
Constant	1.133***	0.215	1.409***	0.124	0.834***	0.279
City Fixed Effects	N		Y		Y	
Year Fixed Effects	N		N		Y	
Dummy=1 if no typhoid article						
Dummy=1 if big news event	-0.742	0.453	-0.936**	0.434	-0.823**	0.340
Dummy=1 if holiday week	-0.292*	0.166	-0.453	0.461	-0.908***	0.186
Constant	-20.318***	5.659	-6.177	8.369	-30.486	28.637
City Fixed Effects	Y		Y		Y	
Year Fixed Effects	Y		Y		Y	
Both interventions	-0.935***	0.135	-0.697***	0.204	-0.398*	0.214
Both intervention death rate interactions	1.422***	0.217	0.769***	0.198	0.417	0.318
ln(alpha)	-0.798***	0.148	-0.914***	0.229	-1.335***	0.136
Vuong test, z=	9.22***		7.55***		7.771***	
Observations	9,492		9,492		9,492	
Zero observations	2,127		2,127		2,127	

The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid article. Robust standard errors in parentheses, clustered on the city. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 4: Negative Binomial Regression of Effect of Weekly Typhoid Death Rates on Weekly Local Newspaper Reports

	(1)		(2)		(3)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of Typhoid Articles						
Death rate	0.405	0.319	0.373**	0.175	0.366**	0.149
1st intervention	-0.549*	0.303	-0.334***	0.086	-0.274**	0.107
2nd intervention	-0.575***	0.109	-0.774***	0.183	-0.282***	0.072
1st intervention x death rate	0.217	0.361	0.275	0.204	0.244*	0.147
2nd intervention x death rate	0.947***	0.254	0.667***	0.252	0.408***	0.061
Number of total articles	-0.008	0.005	0.003	0.008	-0.000	0.007
Constant	0.778*	0.399	0.328***	0.120	0.507	0.315
City Fixed Effects	N		Y		Y	
Year Fixed Effects	N		N		Y	
Dummy=1 if no local typhoid article						
Dummy=1 if big news event	-0.180	0.236	-0.252	0.339	-0.195	0.290
Dummy=1 if holiday week	0.006	0.164	-0.025	0.289	0.039	0.056
Constant	-15.258***	0.809	-14.551	10.733	-0.766**	0.303
City Fixed Effects	Y		Y		Y	
Year Fixed Effects	Y		Y		Y	
Both interventions	-1.124***	0.405	-1.108***	0.253	-0.556***	0.104
Both intervention death rate interactions	1.164***	0.333	0.942***	0.105	0.651***	0.174
ln(alpha)	-0.494***	0.094	-0.797***	0.278	-0.997***	0.373
Vuong test, z=	19.81***		13.08***		13.50***	
Observations	8,596		8,596		8,596	
Zero observations	3,792		3,792		3,792	

The regressions exclude Philadelphia because we could not mechanically identify local articles. The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid article. Robust standard errors in parentheses, clustered on the city. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Average Death Rate and Death Rate Interaction Marginals from Negative Binomial Regression of Effect of Typhoid Death Rates on All and Local Newspaper Reports

	$E(y\|x)$	$\frac{\partial E(y\|x)}{\partial d}$	$\frac{\partial^2 E(y\|x)}{\partial d \partial I_1}$	$\frac{\partial^3 E(y\|x)}{\partial d \partial I_1 \partial I_2}$
Dependent variable:				
A. Total number of articles				
Specification 1	3.056	2.206***	-0.238	4.570***
		(0.457)	(0.798)	(0.920)
Specification 2	3.034	1.877***	0.300	2.482***
		(0.388)	(0.562)	(0.681)
Specification 3	3.033	1.499***	0.242	1.300
		(0.421)	(0.427)	(1.008)
B. Number of local articles				
Specification 1	1.208	0.804*	-0.079	1.542***
		(0.316)	(0.646)	(0.628)
Specification 2	1.160	0.710***	0.199	1.228***
		(0.195)	(0.345)	(0.168)
Specification 3	1.177	0.636***	0.203	0.810***
		(0.165)	(0.236)	(0.224)

Marginals are for the regressions in Tables 3 and 4; d is the death rate and I_1 and I_2 are the first and second interventions, respectively. Robust standard errors in parentheses, clustered on the city. *** p<0.01, ** p<0.05, * p<0.1. Marginals were computed following the procedure outlined in Leitgöb (2014). The mean number of total and local typhoid articles is 2.917 ($\hat{\sigma}$ =3.489) and 1.097 ($\hat{\sigma}$ =2.019), respectively. The number of observations for panels A and B is 9,492 and 8,596, respectively.

Table 6: Negative Binomial Regression of Effect of Deviations in Expected Weekly Typhoid Death Rates on Weekly Newspaper Reports

	All Articles				Local Articles			
	Rolling Mean		ARMAX		Rolling Mean		ARMAX	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of Typhoid Articles								
Death rate - predicted	-0.304	0.350	-0.817**	0.344	-0.047	0.299	-0.820*	0.421
(Death rate - predicted) x (Dummy=1 if positive)	1.238**	0.613	1.822**	0.750	0.944	0.710	1.775*	0.947
Number of total articles	-0.001	0.010	-0.000	0.010	-0.031***	0.008	-0.029***	0.008
Constant	1.050***	0.271	1.039***	0.271	0.817***	0.253	0.758***	0.232
City Fixed Effects	N		N		N		N	
Year Fixed Effects	N		N		N		N	
Dummy=1 if no typhoid article								
Dummy=1 if big news event	-0.512***	0.119	-0.498***	0.143	-0.258	0.199	-0.253	0.201
Dummy=1 if holiday week	-0.531***	0.153	-0.497***	0.146	-0.092	0.114	-0.060	0.125
Constant	-4.383*	2.656	-4.597	3.585	-5.332	22.799	0.294	1.202
City Fixed Effects	Y		Y		Y		Y	
Year Fixed Effects	Y		Y		Y		Y	
Total effect of pos(Death rate-predicted)	0.934***	0.306	1.005***	0.411	0.896**	0.445	0.955*	0.527
ln(alpha)	-0.556***	0.129	-0.502***	0.124	-0.346***	0.107	-5.522	27.953
Vuong test, z=	9.61***		9.52***		21.62***		21.44***	
Observations	9,114		9,114		8,269		8,269	
Zero observations	2,040		2,040		4,619		4,619	

The local article regressions exclude Philadelphia. The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid article. Robust standard errors in parentheses, clustered on the city. The rolling mean uses a 2 year moving window. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: Average Death Rate Deviation and Death Rate Deviation Interaction Marginals from Negative Binomial Regression of Effect of Typhoid Death Rate Deviations on All and Local Newspaper Reports

	$E(y \mid x)$	$\frac{\partial E(y \mid x)}{\partial D}$	$\frac{\partial^2 E(y \mid x)}{\partial D \partial (D>0)}$
Dependent variable:			
A. Total number of articles			
Rolling mean	3.300	0.882	6.634*
		(1.099)	(4.178)
ARMAX	3.214	-0.091**	8.565**
		(1.199)	(4.637)
B. Number of local articles			
Rolling mean	1.235	0.511	1.835
		(0.458)	(1.738)
ARMAX	1.204	-0.041*	3.121***
		(0.563)	(2.200)

Marginals are for the regressions in Table 6; D is the deviation from the death rate. Robust standard errors in parentheses, clustered on the city. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Marginals were computed following the procedure outlined in Leitgöb (2014). The mean number of total and local typhoid articles is 2.917 ($\hat{\sigma}$ =3.489) and 1.097 ($\hat{\sigma}$ =2.019), respectively. The number of observations for panels A and B is 9,114 and 8,269, respectively.

Table 8: Negative Binomial Regression of Effect of Weekly Typhoid Case Rates on Weekly Newspaper Reports

	(1)		(2)		(1)		(2)	
	All Articles				Local Articles			
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of typhoid articles								
Case rate	0.093***	0.022	0.078***	0.019	0.141**	0.057	0.105***	0.034
1st intervention	-0.378***	0.109	-0.285*	0.161	-0.303	0.257	-0.297	0.204
2nd intervention	-0.605***	0.105	-0.128	0.165	-0.660***	0.081	-0.217	0.175
1st intervention x case rate	0.005	0.026	0.010	0.014	-0.026	0.055	0.001	0.026
2nd intervention x case rate	0.213***	0.063	0.104***	0.034	0.171***	0.016	0.129***	0.017
Number of total articles	0.013***	0.003	0.011*	0.005	-0.003	0.004	0.016***	0.006
Constant	1.173***	0.103	0.711*	0.363	0.512	0.388	-0.215	0.370
City Fixed Effects	N		Y		N		Y	
Year Fixed Effects	N		Y		N		Y	
Dummy=1 if no typhoid article								
Dummy=1 if big news event	-0.675***	0.227	-0.993***	0.180	-0.326	0.280	-0.504	0.333
Dummy=1 if holiday week	-0.378**	0.156	-1.001***	0.202	-0.170	0.245	-0.134	0.158
Constant	-5.415***	1.452	-19.704***	1.521	-2.903**	1.380	-0.520***	0.094
City Fixed Effects	Y		Y		Y		Y	
Year Fixed Effects	Y		Y		Y		Y	
Both interventions	-0.984***	0.066	-4.129	0.256	-0.963***	0.326	-0.514*	0.294
Both intervention case rate interactions	0.218***	0.066	0.114***	0.029	0.145***	0.051	0.130***	0.013
ln(alpha)	-1.097***	0.121	-1.523***	0.126	-0.971***	0.177	-1.250***	0.177
Vuong test, z=	8.61***		6.45***		17.32***		11.07***	
Observations	8,214		8,214		7,467		7,467	
Zero observations	2,078		2,078		4,438		4,438	

The local article regressions exclude Philadelphia. The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid article. Robust standard errors in parentheses, clustered on the city. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 9: Average Case Rate and Case Rate Interaction Marginals from Negative Binomial Regression of Effect of Typhoid Death Rates on All and Local Newspaper Reports

	$E(y\|x)$	$\frac{\partial E(y\|x)}{\partial c}$	$\frac{\partial^2 E(y\|x)}{\partial c \partial I_1}$	$\frac{\partial^3 E(y\|x)}{\partial c \partial I_1 \partial I_2}$
Dependent variable:				
A. Total number of articles				
Specification 1	2.636	0.473***	-0.161	0.614***
		(0.092)	(0.089)	(0.218)
Specification 2	2.587	0.325***	-0.060	0.294***
		(0.064)	(0.058)	(0.087)
B. Number of local articles				
Specification 1	1.014	0.178***	-0.101**	0.088**
		(0.040)	(0.075)	(0.182)
Specification 2	0.941	0.137***	-0.040*	0.114***
		(0.026)	(0.030)	(0.032)

Marginals are for the regressions in Tables 8 and 4; c is the death rate and I_1 and I_2 are the first and second interventions, respectively. Robust standard errors in parentheses, clustered on the city. *** p<0.01, ** p<0.05, * p<0.1. Marginals were computed following the procedure outlined in Leitgöb (2014). The mean number of total and local typhoid articles is 2.437 ($\hat{\sigma} = 2.847$) and 0.820 ($\hat{\sigma} = 1.483$), respectively. The number of observations for panels A and B is 8,214 and 7,467, respectively.

Table 10: Negative Binomial Regression of Effect of Deviations in Expected Typhoid Case Rates on Newspaper Reports

	All Articles				Local Articles			
	Rolling Mean		ARMAX		Rolling Mean		ARMAX	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of Typhoid Articles								
Case rate - predicted	-0.163**	0.072	-0.342***	0.067	-0.121	0.087	-0.390***	0.037
(Case rate - predicted) x (Dummy=1 if positive)	0.450***	0.086	0.739***	0.108	0.391***	0.072	0.792***	0.081
Number of total articles	0.004	0.008	0.003	0.008	-0.015***	0.002	-0.014***	0.002
Constant	0.769***	0.279	0.801***	0.277	0.294*	0.160	0.284**	0.143
City Fixed Effects	N		N		N		N	
Year Fixed Effects	N		N		N		N	
Dummy=1 if no typhoid article								
Dummy=1 if big news event	-0.476***	0.114	-0.430***	0.121	-0.309*	0.187	-0.276	0.177
Dummy=1 if holiday week	-0.275**	0.135	-0.280*	0.144	-0.084	0.102	-0.075	0.098
Constant	-4.024***	1.024	-4.063***	1.256	-1.702***	0.434	-1.591***	0.374
City Fixed Effects	Y		Y		Y		Y	
Year Fixed Effects	Y		Y		Y		Y	
Total effect of pos(Case rate-predicted)	0.288***	0.054	0.397***	0.051	0.270***	0.044	0.402***	0.050
ln(alpha)	-0.782***	0.064	-0.716***	0.071	-0.683***	0.080	-0.656***	0.070
Vuong test, z=	9.81***		10.24***		19.18***	19.99***		
Observations	7,706		7,706		7,062		7,062	
Zero observations	2,004		2,004		4,268		4,268	

The local article regressions exclude Philadelphia. The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid article. Robust standard errors in parentheses, clustered on the city. The rolling mean uses a 2 year moving window. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 11: Average Case Rate Deviation and Case Rate Deviation Interaction Marginals from Negative Binomial Regression of Effect of Typhoid Case Rate Deviations on All and Local Newspaper Reports

	$E(y\|x)$	$\frac{\partial E(y\|x)}{\partial D}$	$\frac{\partial^2 E(y\|x)}{\partial D \partial (D>0)}$
Dependent variable:			
A. Total number of articles			
Rolling mean	2.905	0.142**	2.858***
		(0.183)	(0.778)
ARMAX	2.576	-0.126***	2.821***
		(0.156)	(0.533)
B. Number of local articles			
Rolling mean	0.979	0.067	0.815***
		(0.065)	(0.181)
ARMAX	0.896	-0.050***	1.184***
		(0.038)	(0.175)

Marginals are for the regressions in Table 10; D is the deviation from the case rate. Robust standard errors in parentheses, clustered on the city. *** p<0.01, ** p<0.05, * p<0.1. Marginals were computed following the procedure outlined in Leitgöb (2014). The mean number of total and local typhoid articles is 2.436 ($\hat{\sigma}$ =2.847) and 1.483 ($\hat{\sigma}$ =0.820, respectively. The number of observations for panels A and B is 7,706 and 7,062, respectively.

Table 12: Effect of Past News Articles on Weekly Change in Death Rates Multiplied by 100, IV Results

	Independent Variable=Number of News Articles in Past N Weeks, N=					
	1	2	3	4	5	6
Main regression						
	-0.231	-0.231*	-0.100*	-0.104**	-0.085**	-0.090**
	(0.207)	(0.129)	(0.054)	(0.048)	(0.038)	(0.042)
Observations	7,659	7,659	7,659	7,659	7,659	7,659
R^2	0.011	0.011	0.011	0.011	0.011	0.011
Restricted to data with non-missing death lag values						
		-0.222*	-0.015	-0.107	-0.053**	-0.107*
		(0.130)	(0.000)	(0.072)	(0.026)	(0.064)
Observations		7,043	6,598	6,262	5,997	5,779
R^2		0.013	0.012	0.013	0.012	0.012
Regression includes lagged death values						
		-0.165	0.013	-0.086	-0.017	-0.081
		(0.114)	(0.030)	(0.075)	(0.017)	(0.060)
Includes lagged death values of		2	2-3	2-4	2-5	2-6
Observations		7,043	6,598	6,262	5,997	5,779
R^2		0.014	0.014	0.014	0.017	0.017

The dependent variable is the difference in death rates between this week's and last's week's death rates multiplied by 100. The mean of the dependent variable is 0.078. All regressions include year, week, and city fixed effects. In all cases the IV is the past number of non-local typhoid articles, e.g. when the independent variable is the number of typhoid articles in the past 4 weeks, the IV is the number of non-local typhoid articles in the past 4 weeks. Philadephia is excluded from the regressions because we cannot identify local typhoid articles. The main regression is Equation 9 in the text. The regression that includes lagged death values is Equation 10 in the text. We also run Equation 9 in the text restricting the data to observations with non-missing lagged death values. Robust standard errors in parentheses, clustered on the city. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 13: Negative Binomial Regression of Effect of Weekly Typhoid and Diphtheria Death Rates on Weekly Typhoid and Diphtheria Newspaper Reports

	All Articles				Local Articles					
	Typhoid		Diphtheria		Typhoid		Diphtheria		Diphtheria	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Number of articles										
Typhoid death rate	0.806***	0.092	0.221*	0.122	1.507***	0.173	-0.009	0.189		
1st typhoid intervention	-0.080	0.083	-0.395***	0.139	-0.092	0.128	-0.589**	0.252		
2nd typhoid intervention	-0.563***	0.110	-0.160	0.152	-0.540***	0.166	-0.122	0.287		
1st typhoid intervention x death rate	0.430	0.276	-0.466	0.504	0.548	0.412	0.096	0.763		
2nd typhoid intervention x death rate	2.572***	0.844	-1.752	1.201	4.208***	1.562	-0.640	1.883		
Diphtheria death rate									-0.057	0.087
Diphtheria intervention									-0.887***	0.287
Diptheria intervention x death rate									0.599***	0.121
Number of total articles	0.006**	0.003	0.016***	0.003	0.007	0.005	0.018***	0.006	0.019***	0.004
Constant	1.072***	0.065	0.446***	0.081	-0.202*	0.110	-0.308**	0.150	-0.223	0.265
Dummy=1 if no article										
Dummy=1 if big news event	-1.113**	0.464	-0.269	0.274	-1.044**	0.464	-0.106	0.304	-0.141	0.294
Dummy=1 if holiday week	-0.757*	0.400	-0.255	0.228	-0.848**	0.414	-0.397*	0.241	-0.394*	0.230
Constant	-3.669	3.460	-27.033	45,614.242	-1.915	1.670	-1.249	0.819	-1.314	0.858
Year Fixed Effects	Y		Y		Y		Y		Y	
Both interventions	-0.643***	0.106	-0.556***	0.122	-0.632***	0.171	0.711***	0.230		
Both intervention death rate interactions	3.002***	0.810	-2.219**	1.108	4.756***	1.532	-0.544	1.754		
ln(alpha)	-1.507***	0.102	-1.513***	0.134	-0.990***	0.124	-1.805***	0.422	-2.139***	0.590
Vuong test, z=	5.74***		8.51***		5.98***		7.82***		8.08***	
Observations	2,548		2,548		2,548		2,548		2,548	
Zero observations	497		1,607		1,271		1,695		1,695	

The dependent variable for the count part of the negative zero-inflated binomial is number of typhoid or diphtheria articles. The dependent variable for the logit part of the model is a dummy equal to 1 if there was no typhoid or diphtheria article. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 14: Average Case Rate and Case Rate Interaction Marginals from Negative Binomial Regression of Effect of Typhoid and Diphtheria Death Rates on All and Local Newspaper Reports

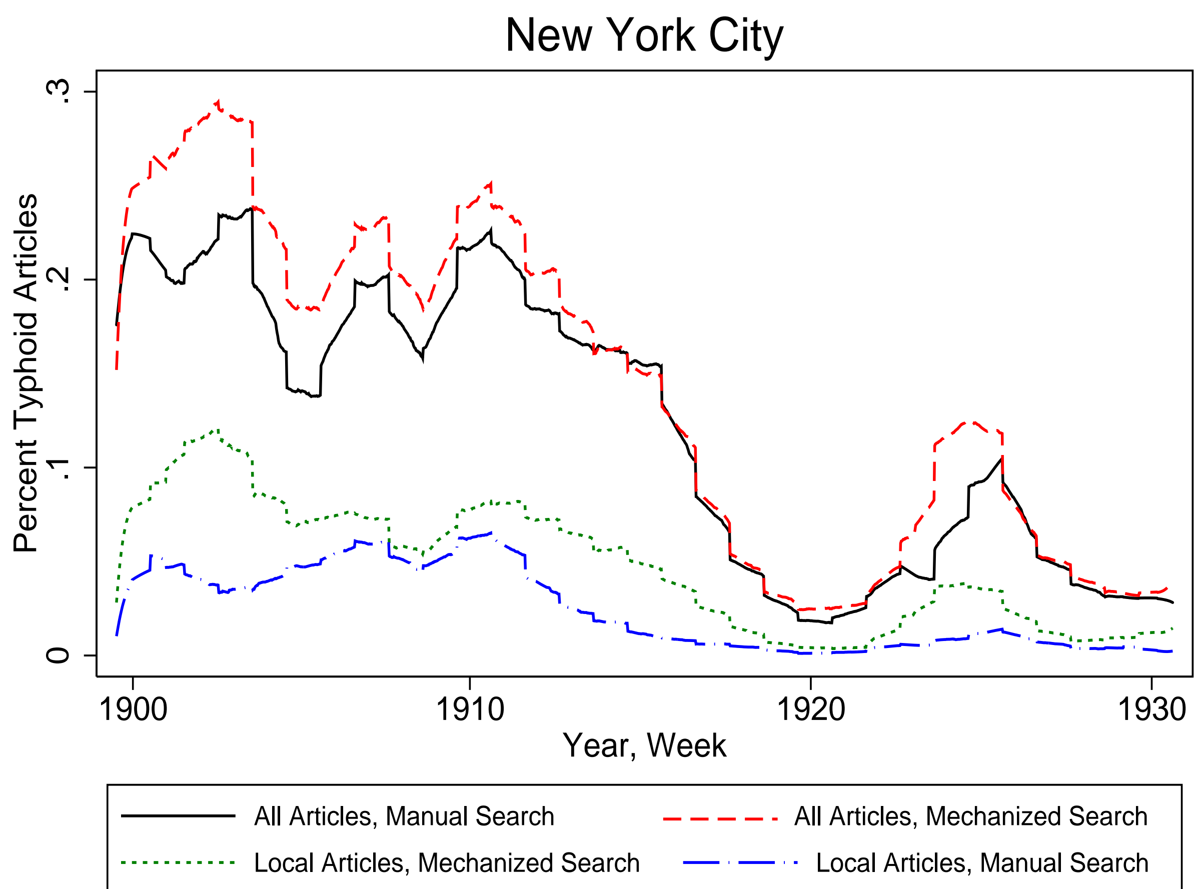
	$E(y\|x)$	$\frac{\partial E(y\|x)}{\partial d}$	$\frac{\partial^2 E(y\|x)}{\partial d \partial I_1}$	$\frac{\partial^3 E(y\|x)}{\partial d \partial I_1 \partial I_2}$
Dependent variable:				
A. Total number of articles				
Typhoid articles on typhoid deaths	2.984	5.374***	1.189	9.230***
		(1.019)	(0.809)	(2.543)
Diphtheria articles on typhoid deaths	1.457	-1.166	-0.225	-3.273**
		(0.864)	(0.936)	(1.640)
B. Number of local articles				
Typhoid articles on typhoid deaths	1.005	2.788***	0.597	5.218***
		(0.544)	(0.530)	(1.780)
Diphtheria articles on typhoid deaths	0.506	0.122	0.120	-0.285
		(0.484)	(0.496)	(0.918)
Diphtheria articles on diphtheria deaths	0.508	0.232**	0.184**	
		(0.043)	(0.053)	

Marginals are for the regressions in Tables 13; d is the death rate and I_1 and I_2 are the first and second interventions, respectively. We could not achieve convergence for the regression of weekly diptheria articles on total diptheria deaths. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Marginals were computed following the procedure outlined in Leitgöb (2014). The mean number of total and local typhoid articles is 2.996 and 1.008, respectively. The mean number of total and local diphtheria articles is 1.456 and 0.505, respectively. The number of observations for panels A and B is 2,548.

Appendix A

Figure A1 compares our mechanized searches of *The New York Times* for typhoid fever with manual searches for 1900-1932. Except for some learning by doing by data inputters in the early manual searches, peaks and troughs in the percentage of articles mentioning typhoid coincide across the two samples. However, the mechanized searches for mentions of local typhoid fever yield a consistently greater percentage than the manual searches. The mechanized searches also capture deaths or illnesses of New Yorkers outside of the city.

Figure A1: Comparison of Manual and Mechanized Searches, *The New York Times* 1900-1932



Source: *The New York Times*.

Appendix B

Table B1 shows that in all cities the first intervention was a statistically significant predict of death rates, controlling for a year trend. Controlling for the effects of the first intervention, in all cities except for Washington DC the second intervention was a statistically significant predictor of death rates and the combined effect of both interventions was statistically significant in all cities except for Washington DC. The inclusion of lagged death rates in the specification also yielded statistically significant effects of the interventions (results not shown).

We also investigated whether the expansion of sewage systems in the Boston metropolitan area had any effects on typhoid rates in Boston. We included in our regression specification three dummy variables each equal to one, respectively, after 1892 (when the Charles River System was completed), after 1897 (when the Neponset Valley System was completed), and after 1904 (when the High Level System was completed). We did not find a statistically significant effeect of any of the sewage dummies. We lost significance on the first clean water intervention (probably because after 1897 is too close to after 1898) but not on the second clean water intervention. Sewage construction continued after 1904. For example, high level sewers were constructed for Newton, Brighton, and Brookline in 1907, 1908, and 1909, respectively but these interventions are too close to the second clean water intervention (1908). The Mystic Sewer was built in 1912, 1913, and 1914 but we did not obtain any robust results for this intervention.[8]

[8]For information on sewage systems see Alsan and Goldin (2015) and Loud (1923).

Table B1: OLS Regression of Effect of Clean Water Interventions on Typhoid Death Rates

	NYC	Chicago	Baltimore	Boston	Philadelphis	DC
1st intervention	-0.116***	-0.158***	-0.080*	-0.075**	-0.614***	-0.257***
	(0.010)	(0.020)	(0.042)	(0.030)	(0.056)	(0.084)
2nd intervention	-0.073***	0.088***	-0.112***	-0.127***	-0.197***	0.050
	(0.009)	(0.021)	(0.039)	(0.030)	(0.046)	(0.060)
Combined interventions	-0.189***	-0.070**	-0.192**	-0.202***	-0.811***	-0.207
	(0.013)	(0.035)	(0.043)	(0.037)	(0.090)	(0.132)
Observations	2,548	1,538	1,387	1,588	904	1,558
R^2	0.517	0.503	0.391	0.396	0.476	0.428

All regressions include a linear time trend and constant term. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1