State Borders and Public Health

John M. Chambers

Princeton University

v11.2 — January 18, 2024

Abstract

I contribute to the debates on optimal jurisdiction size and federalism through a quasi-natural experiment in the context of US food safety. State borders create fissures in information networks and authority hierarchies. Accordingly, I use the varying extent to which Census-designated metro area populations straddle state borders as the nature-assigned variable; and I show that these fissures prolong foodborne illness outbreaks only in the case of chemically-induced outbreaks—the swiftest category of outbreak. I conclude that policy areas requiring a quick response will tend to benefit from larger jurisdictions but that the benefits can largely be derived by centralizing information without centralizing control.

1 Introduction

Coordination, which consists of both central communication and central control, is a critical component of a wide number of bureaucratic operations, especially where time is of the essence. Loss of communication on the battlefield can damn a commander's strategy by obstructing the inflow of status reports and the outflow of adaptive orders based on those reports. Loss of coordination among aircraft through air traffic control and radio communications could easily result in catastrophic collisions. And loss of coordination among police units via radio can substantially ease the escape of a mobile suspect.

Of course, on the other hand, maintaining central communication and control can undermine the effectiveness of local responses. If a unit has to send status reports up to its superiors and secure their approval before taking any actions in response to the information reported, then the entire operation may be substantially slowed. Further, as debates surrounding federalism and local home rule emphasize, centralized control often cuts against tailoring to local preferences.

CDC Disclaimer: The findings and conclusions in this work are those of the author and do not necessarily represent the official position of the United States Centers for Disease Control and Prevention.

In this chapter, I bring empirical evidence to bear on this debate, showing the benefits of coordination in the case of foodborne illness supervision. The varying extent to which US metro areas straddle state borders provides a quasi-natural experiment that tests whether and how much the control of foodborne illness outbreaks is enhanced when all parts of an area share the same state public health infrastructure. The results show that outbreaks specifically of chemically-induced foodborne illness—those for which time is most emphatically of the essence—are shorter when a metro area's public health authorities do not have to coordinate across state lines as frequently.

1.1 Place in the Literature

The literature on centralization and decentralization, whether the concern is subnational units or nations, is largely about the trade-offs. Oates (1999) lauds smaller units' ability to maximize utility by tailoring policy to local preferences and costs, but he recognizes that smaller units may be unable to control the costs and benefits of a given policy simultaneously. Landry (2008) takes on the dichotomy between a decentralized economic development scheme that leverages local expertise and a centralized system of control that secures national priorities, arguing that the Chinese Communist Party was able to get both through its promotions system. However, most immediate for this chapter is Alesina and Spolaore's (2003) work seeking out the optimal country size by balancing size advantages and preference heterogeneity. Though their analysis primarily concerns country size, they explicitly acknowledge its ready analogy to subnational units. Here, I further develop and qualify their posited size advantages with an emphasis on transaction costs.

Coordinating across state boundaries to solve foodborne illness outbreaks clearly incurs transaction costs as public health workers are forced to work with multiple states' informational and authority structures. The CDC reduces the *informational* transaction costs substantially by serving as a national public health data clearinghouse with each state's consensual and ongoing participation. As I show in this chapter, the setup is generally effective, demonstrating that centralized authority is not always necessary to secure the benefits of centralization. Thus, in a similar spirit to Landry, the advantages of size and those from customized local policy are not necessarily mutually exclusive.

However, this conclusion is not without qualification. Despite the CDC's work, chemically-induced foodborne illness outbreaks are clearly prolonged when they occur in metropolitan areas that span multiple states. Such outbreaks begin and end swiftly, leaving little time for successful intervention; and the transaction costs or frictions that are not remedied by the current system of coordination for interstate outbreak control are sufficient to leave a measurable and deleterious effect. Most broadly, this indicates that, for policy areas in which speed is crucial for an effective response, larger and more centralized jurisdictions will be advantageous.

Importantly, there could be at least two particular mechanism for this tendency, each with very different implications from the other. First, the informational infrastructure might not be sufficiently consolidated. States still maintain their own public health databases apart from the CDC's system, and the associated lag between data entry and data consolidation could be to blame for reduced performance. To the extent that this is the case, the performance deficit could be eliminated without compromising local autonomy and innovation.

On the other hand, the reduced performance could spring from an insufficiently consolidated authority structure. The CDC is typically required to have a state's invitation before it can intervene and actively coordinate a multistate response, and that process may sometimes delay the response so long that it can no longer be effective. To the extent that this is the case, there is an upper limit on the capacity for consensual transaction cost reductions, and the benefits from central authority and local autonomy cannot be fully and simultaneously secured.

1.2 Foodborne Illness Detection & Control

The contamination responsible for an outbreak of foodborne illness can occur at any point in the food supply chain: farm, slaughterhouse, packaging, distribution, refrigerator, kitchen, and takeout box. After an outbreak begins, those affected might all stay in the same town; but some could commute back to homes in nearby counties and states; and others might fly a few thousand miles away to return from a vacation. Assuming these people do actually go see a physician for their illness at some point or otherwise report it to the public health authorities, the evidence they leave for that illness could be concentrated in a single household, or it might be scattered across the globe. In addition, the farther up the food supply chain the contamination occurs, the more widely distributed could be the contamination, and the more complicated would be the investigation because of the added layers of private business information the government would have to acquire and wade through to reach the source.

All of this makes the detection and control of foodborne illness incredibly coordination-intensive. Each county and city may have its own public health office; and each state may have its own information-sharing and analysis infrastructure. The local offices are usually the ones to collect meal and symptom histories from affected people through case interviews, and they are the first to consolidate and analyze the data from those interviews.[1] Loosely speaking, when they can't identify a clear cause or lack jurisdiction over the

[1] See the model questionnaire at `http://cifor.us/downloads/clearinghouse/NHGQ_v2_OMB0920_0997.pdf`.

outbreak's suspected origin or reach, they may seek assistance from the state; and, if the state encounters similar problems, it may, in turn, request help from the CDC.[2]

While this system is a natural outgrowth of the US's federated system, it may create problems for multistate cities or metro areas, as I hypothesize here. Take New York City for example. If a restaurant near New York Penn Station serves improperly stored poultry, then the affected patrons could plausibly go home to and only report their illnesses in New Jersey. In this case, New Jersey would initiate the investigation, but it would have to rely on New York to inspect the implicated restaurant. That extra bureaucratic step—which is unnecessary in single-state metro areas—could easily delay control efforts, leaving bad chicken in circulation for longer and extending the outbreak.

More generally, the more states there are involved in an outbreak, and the more evenly distributed the affected population is among them, the more difficult could be the investigation. A more even population distribution means that each state has fewer case reports to compare against one another, making it all the more difficult to identify a single cause. If an *outbreak's* affected population is often split among its metro area's states, then metro areas that straddle state lines will have greater difficulty controlling foodborne illness than will metros areas that lie entirely within a single state.

1.3 Measurement & Key Variables

Operationalizing this intuition requires a definition of the US's metro areas and a measure of their dispersion among the states. For metro areas, I use the U.S. Office of Management and Budget's delineation of Core-Based Statistical Areas (CBSAs) from February 2013. Each designated metro has an urban population core of at least 10,000, and its constituent counties have substantial commuting relationships with one another.[3] The February 2013 vintage is the most appropriate option because it lies towards the middle of the study period and is the first delineation to make use of the 2010 Census, which informs several of the key variables.

To measure dispersion among the states within a metro, I adapt the Herfindahl Index of industry concentration to population share by state. If state j's *proportion* (not percentage share) of the population of metro i is s_j, then i's Normalized Herfindahl Index H_i is

$$H_i \equiv \sum_j s_j^2$$

[2]See `https://www.cdc.gov/fdoss/faq.html` and information on SEDRIC later.
[3]For further details, see subsection A.1.

The adapted index ranges from 0 to 1, with 0 indicating greatest dispersion and 1 indicating greatest concentration; of course, the observed range in the data is more limited. The lower bound is 0.4, held by the Washington-Arlington-Alexandria metro, whose population is spread reasonably evenly across the nation's capital, Virginia, and Maryland. The upper bound is 1, held by a variety of metros including Miami-Fort Lauderdale-West Palm Beach, whose population is contained entirely within the state of Florida.

I use outbreak length as my outcome variable. The CDC measures this as the number of days from the first date of exposure of any primary case to the last, inclusive.[4] A primary case is an individual with illness resulting from direct exposure to the responsible chemical agent via the initially implicated food.[5] On the theoretical side, this interval reflects whether and to what extent public health authorities are able to control the root cause of an outbreak; either the food goes on intoxicating people until it runs out, or its spree is cut short via successful investigation and intervention by public health authorities. On the practical side, outbreak length is reported only for primary cases;[6] so I am forced to restrict my attention and analyses accordingly.

1.4 Sample & Scope

Broadly, for the main analyses, the sample is the set of outbreaks of chemically induced foodborne illness in the CDC's National Outbreak Reporting System (NORS).[7] Each of these outbreaks is voluntarily reported to the system by one or more local, state, and/or federal entities; and, while such reporting potentially biases the dataset so that better-functioning metros report more and longer outbreaks than their worse-functioning siblings, if anything, this should work against my hypotheses: If greater consolidation indeed enhances public health performance, it could also make a metro's outbreak lengths appear artificially inflated.

From this set, I restrict to those which can be definitively assigned to a metro via the county (or counties) in which toxin exposure occurred. If an outbreak's exposures occurred in more than one metro, I treat each metro-outbreak pair as a separate outbreak; though the pairs might share the same ultimate cause, they each provide a test of a separate local public health infrastructure since substantial portions of the investigation and control phases are necessarily local.[8] About 4% of outbreaks are duplicated one or more times as a result of this rule. This leaves 272 outbreaks (or metro-outbreak pairs), and their dates range from 1998 through 2020.[9]

[4]Centers for Disease Control and Prevention 2017, p. 15.

[5]Centers for Disease Control and Prevention 2017, pp. 15, 18. By contrast, a secondary case is a non-primary case whose exposure is often the result of person-to-person contact with a primary case.

[6]Centers for Disease Control and Prevention 2017, p. 13.

[7]Centers for Disease Control and Prevention 2021.

[8]Unfortunately, the data's aggregated nature does not enable me to assign a unique length for each metro's portion of an outbreak.

[9]When the dataset is expanded to include other etiologies, it extends back to 1997.

Chemical	# of Outbreaks
Scombroid toxin	87
Ciguatoxin	80
Histamine	18
Mycotoxins	12
Plant/Herbal toxins	5
Heavy metals	4
Pesticides	3
Cleaning agents	3
Paralytic shellfish poison	3
Puffer fish tetrodotoxin	1
Other	56

Table 1.1: Chemical Outbreaks by Specific Etiology.

As Table 1.1 shows, these outbreaks are typically caused by biotoxins like scrombroid toxin and ciguatoxin that appear naturally in fish like tuna, mackerel, and shellfish.[10] More familiar chemicals like cleaning agents and pesticides can also contaminate food and induce illness, but outbreaks resulting from these are comparatively rare.

While NORS also records outbreaks from other etiologies like viruses and bacteria, I focus primarily on chemical outbreaks here because of their comparatively shorter exposure periods. From start to finish, a typical chemical outbreak does all of its poisoning in a little over 3 days. In that short time frame, if public health authorities are to control the outbreak at all, they must move swiftly; even minor delays due to jurisdictional barriers could combine to extend the outbreak substantially as a proportion of the typical length. For example, state A might identify a variety of leads on an outbreak's cause and ask state B to investigate; but state B's officers might not investigate until the next business day. That short 8-hour delay would waste a whopping 10% of the typical outbreak's exposure period—10% of the public health authority's window for action; and investigation is just one of several points in the course of the epidemiological life cycle at which state borders could interfere.

In contrast, the other major etiologies all have longer average exposure periods—ranging from 4 to 13 days. Such lengths mean that public health authorities still have time to make a difference in outbreak length even if their investigation and control measures suffer delays from border friction. If that same 8-hour delay in coordination between states A and B occurs with one of these etiologies, only 3% to 8% of the typical window for action is lost.[11]

[10] On ciguatoxin, see National Organization for Rare Disorders (NORD) 1986 and B. Seymour, Andreosso, and J. Seymour 2015. On scrombroid toxin, see U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition 2017 and Stratta and Badino 2012.

[11] For more on the chemical focus, see Appendix B.

1.5 Methodology

My primary mode of analysis is OLS; but, because of the limited sample size, I employ a variety of additional techniques to maximize the information I extract from each observation. My first step is to use the lasso for variable selection, with the tuning parameter λ chosen via 10-fold cross-validation to avoid under- and overfitting;[12] I then estimate the lasso's preferred model using OLS. This approach ensures that the statistical significance of the included variables is not artificially deflated or masked by the inclusion of uninformative variables.

However, when faced with discrete or post-processed continuous variables, the lasso is not guaranteed to have a single optimal variable set.[13] Both of these variable types appear in my models; so I have to mitigate this potential nonuniqueness. My first response is to run the lasso many times and accept the model that it selects in the plurality of cases.[14] This essentially assumes that, with a sufficiently large sample of lasso runs, the plurality model will converge to the lasso's optimal model.

Unfortunately, with this method, the lasso sometimes chooses a model that includes only the constant/intercept term. In these cases, I start with a small model—typically a univariate model—and progressively expand the set of variables I offer to the lasso until its chosen model becomes implausibly small. Naturally, this approach also allows me to check the robustness of my findings to alternative specifications.

There are some cases in which I need to introduce another variable in order to test an alternative hypothesis even when the lasso is already "saturated." For these cases, I turn to partial F-tests to determine whether the additional variable improves the fit sufficiently to merit inclusion.

Concerning missingness, while I generally use listwise deletion and drop any observations that have missing values for any of the variables in a given regression, there is one case in which that approach excessively shrinks the sample size. In this case, I instead use missingness indicators. In particular, for each additional variable, I create an indicator variable for whether a value is missing in that variable, and I set the missing values in the original variable to 0. This allows me to retain observations that have missing values in those variables and still generate estimates for those variables. The missingness indicators essentially give the variance associated with a variable's missing values somewhere to pool.

To enhance interpretability and unless otherwise noted, I standardize all variables that I can before fitting models or presenting results. As a consequence, intercept coefficients are sometimes 0. Notably,

[12] See James et al. 2013, §§6.2.2-6.2.3.

[13] See Tibshirani 2013.

[14] Note that the variation in which model the lasso selects is a consequence both of the lasso's optimization algorithm having different starting points and of the randomness with which the sample is allocated among the cross-validation folds.

I do not standardize indicator and factor variables since doing so tends to obscure rather than enhance interpretability.

Finally, I rely on the Bayesian Information Criterion (BIC) throughout for the purposes of model comparison and selection.[15] I take a baseline BIC from a simple bivariate model in the next section. Then, for each subsequent model, I report a *deltaBIC*, which is the difference between the BIC for the new model and the BIC for the bivariate model when both models are run using the same observations. A positive deltaBIC implies that the new model is a worse choice than the bivariate model while a negative deltaBIC implies that the new model is a better choice.

[15] See Neath and Cavanaugh 2012.

2 Main Analyses

I start the analysis with a simple model and focus solely on outbreaks of chemical toxicity. The unit of analysis is a single outbreak; the lone explanatory variable is the associated metro area's concentration (cbsaHerfindahlNorm); and the outcome variable is outbreak length. As Table 2.1 and Figure 2.1 show, multijurisdictional control indeed appears to impede the response to a chemical outbreak: the less consolidated the metro area, the longer the outbreak.

	DV: logOnePlusExposureLengthDays
	Simple Model
(Intercept)	0.000
	(1.000)
cbsaHerfindahlNorm	-0.203 **
	(0.002)
N	223
R2	0.041
logLik	-311.231
BIC	638.684

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 2.1: Metro concentration shortens outbreaks.

To assess the robustness of my findings, I expand out to the model most frequently chosen by the lasso. In addition to the concentration measure, this model includes cbsaHeatingDegreeDaysAnnualAvg (a measure of climate) and monthFirstIll11 (a measure of seasonality).

Concerning the first, climate is an established determinant of disease prevalence.[16] Different species from the largest whale to the smallest bacterium thrive in different temperature ranges, moisture and light levels, and wind conditions—among a litany of other factors. Further, whether a species is a pathogen, a pathogen carrier, or just a supportive member of a pathogen's ecosystem, variation in its prevalence can contribute to a disease's prevalence. In other words, climatic variation contributes both directly and indirectly to a disease's success.

[16] For an example of this phenomenon, see Mordecai et al. 2019.

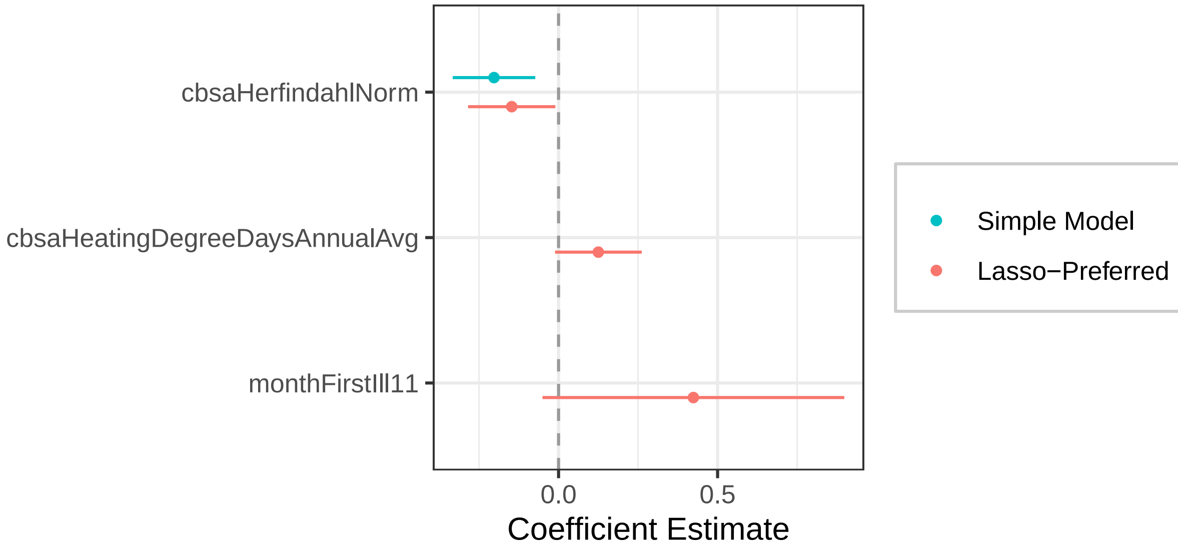

Figure 2.1: Effect plot with 95% confidence intervals. In both of the primary models, a 1-standard-deviation increase in metro concentration (cbsaHerfindahlNorm) is associated with a reduction in outbreak length; and, in the lasso's preferred model, the magnitude of that effect is comparable to that associated with a 1-standard-deviation increase in a metro's average heat burden (cbsaHeatingDegreeDaysAnnualAvg), but the former actually attains statistical significance.

To help account for this, I include cbsaHeatingDegreeDaysAnnualAvg—a measure of how hot a metro area tends to be. I also offer the lasso the cooling degree day equivalent cbsaCoolingDegreeDaysAnnualAvg, but it regularly excludes that variable in the various specifications.

To build some intuition for what heating degree days actually capture—and, in turn, to see how they might matter for chemical outbreaks as well as what they might miss—it is useful to summarize their computation. To compute heating degree days for a given weather station on a single date i, one starts with the station's mean temperature M_i for that date, which is just the average of the day's high and low temperatures—U_i and L_i, respectively:

$$M_i \equiv \frac{U_i + L_i}{2}$$

Then, to compute heating degree days D_i for date i, one subtracts off the magnitude of the reference temperature (65 °F here) and sets the minimum result to 0:

$$D_i \equiv \max(M_i - 65, 0)$$

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

Intuitively, this captures only the extent to which the day's mean temperature *exceeds* the reference temperature. Cooling degree days capture the other side.

Next, to compute the station's heating degree days for year j, one sums across all dates in that year. So, if J is the set of all dates in year j, then

$$A_j \equiv \sum_{i \in J} D_i$$

Finally, to compute the average annual heating degree days C for the station, one averages across all years in the data period. So, if K is the set of years in that period, then

$$C \equiv \frac{\sum_{j \in K} A_j}{|K|}$$

Now, while this provides a measure for a given weather station, there remains the issue of what value to choose for a metro. A metro may be quite large and have tens of weather stations within its borders, so taking a value from just one of those stations or generating some composite could both be reasonable.

My source for temperature information is NOAA's 30-Year U.S. Climate Normals for 1991-2020.[17] The dataset supplies summary and average weather information for thousands of weather stations throughout the United States. Though NOAA offers earlier vintages and a 15-year version, my chosen vintage and coverage length best reflect the climate during the outbreaks in the sample; in addition, the 30-year version better controls for years with unusual weather than does the 15-year version.

Data completeness varies substantially from station to station, and NOAA adapts World Meteorological Organization standards to classify station records according to completeness. A record is considered "standard" when it has at least 24 years of data in the 30-year period. Records falling below that threshold are considered "representative" when they have at least 10 years of data in the 30-year period and can reasonably have the remaining years imputed from nearby weather stations.[18] For the purpose of generating the degree day measures, I consider only those stations whose records fall into one of these two categories. NOAA reports average annual heating degree days for each such station.

Returning now to the issue of metro value, for each metro, I select the weather station that's closest to the metro's population centroid and take its average annual heating degree days directly as the metro's value. Intuitively, this captures the weather that the averagely-located person in the metro experiences.

[17] NOAA 2021a.
[18] See NOAA 2021b, p. 4.

Month First Ill	Average Outbreak Length
1	2
2	1.75
3	1.9
4	1.53
5	2.07
6	1.9
7	2.27
8	2.45
9	2.30
10	10.4
11	4.11
12	7.76

Table 2.2: Chemical outbreaks are longest in the final three months of the year.

To calculate metro population centroid, I use a population-weighted average of the U.S. Census's population centroids for the metro's component counties. To approximate the geodetic distances between each metro population centroid and weather station, I take the shortest distance between those two points along the surface of the NAD83 ellipsoid—an approximation of the Earth's shape and location in space that is optimized for use in North America and largely standard for U.S. government mapping.[19]

Turning now to the other variable I add to reach the lasso-preferred specification, monthFirstIll11 is an indicator for whether an outbreak's first case to experience symptoms saw initial onset of those symptoms during November. Intuitively, this is a seasonality indicator—the only one of the twelve month indicators reliably to survive the lasso's sieve.

Just as influenza's prevalence varies throughout the year, chemical outbreak length also has a seasonality. As Table 2.2 demonstrates, chemical outbreaks typically last the longest in the final three months of the calendar year. Variation in the life cycles of the algae and fish that contribute to the most common chemical etiologies may be responsible for some of this seasonality. Human "wintering" in the tropical vacation destinations where these organisms thrive could also play a part, lengthening outbreaks and the investigations thereof by introducing people who are unfamiliar with the common symptoms of these illnesses and less familiar with the public health response that they demand. Whatever the case, the seasonality of length is clear; and, after controlling for metro climate, November is the month that makes the most statistically influential difference.

Admittedly, October is the clear outlier for outbreak length; but it does not convince the lasso. And, since these seasonality indicators are ultimately just nuisance parameters, I do not pursue the reasons for the lasso's choices any further here.

[19] For details on these calculations, see subsection A.2.

	DV: logOnePlusExposureLengthDays
	Lasso-Preferred Model
(Intercept)	-0.034
	(0.615)
cbsaHeatingDegreeDaysAnnualAvg	0.125
	(0.072)
monthFirstIll11	0.424
	(0.080)
cbsaHerfindahlNorm	-0.147 *
	(0.036)
N	223
R2	0.068
logLik	-308.068
deltaBIC	4.487

*** p < 0.001; ** p < 0.01; * p < 0.05. p values in parentheses.

Table 2.3: Metro concentration is still associated with shorter outbreaks when I include the lasso's preferred variables.

As Figure 2.1 and Table 2.3 demonstrate, when I add these two primary controls to complete the lasso-preferred model, metro concentration's effect survives intact. Its significance is slightly attenuated, but that's to be expected with the added variables using up a couple of degrees of freedom.

What is perhaps more surprising is that the other two variables do not attain significance. As I already established, aside from metro concentration, the lasso most frequently prefers these two over all the others I offer it. In addition, these have fairly strong theoretical explanations undergirding them. Even if the metro concentration variable itself were not significant, we would expect at least one of these variables to be.

All this speaks to the limitations inherent in the data. With only 223 outbreaks distributed among 75 metro areas, the *effective* degrees of freedom may not be sufficient to demonstrate the existence even of powerful and known effects. This should condition our interpretation of all the results here. Most importantly, data limitations may be part of the reason the lasso rejects all variables when I offer it too many—though, as I have already discussed, my inclusion of dichotomous variables is likely the primary contributor.

2.1 Further Variables & Effect Stability

To build further confidence that the result on metro concentration is not a spurious byproduct of omitted variable bias, I move from the Simple Model to the Lasso-Preferred Model by progressively expanding the set of variables offered to the lasso. Along the way, I explain how each yet-to-be-discussed and ultimately unselected variable could plausibly be expected to affect outbreak length—and, in turn, why it demands consideration from the lasso if omitted variable bias is reasonably to be avoided. As Table 2.4 demonstrates, at every stage of the expansion, the metro concentration variable is selected, and its coefficient is negative and statistically significant. Not only effect is stable and reasonably robust to alternative specifications; it is also favored against a variety of other plausibly influential variables for its explanatory value.

	DV: logOnePlusExposureLengthDays			
	Population	Climate	Season	SEDRIC
cbsaHerfindahlNorm	-0.203 **	-0.163 *	-0.147 *	-0.147 *
	(0.002)	(0.020)	(0.036)	(0.036)
N	223	223	223	223
R2	0.041	0.055	0.068	0.068
logLik	-311.231	-309.584	-308.068	-308.068
deltaBIC	0.000	7.520	4.487	4.487
Lasso Variable Selection				
logCBSAPopDensityMilesSq	Offered	Offered	Offered	Offered
logCBSAPopulation	Offered	Offered	Offered	Offered
cbsaHeatingDegreeDaysAnnualAvg		Selected	Selected	Selected
cbsaCoolingDegreeDaysAnnualAvg		Selected	Offered	Offered
cbsaAnnAvgsStationElevation		Offered	Offered	Offered
monthFirstIll			Selected	Selected
postSEDRIC				Offered
cbsaHerfindahlNorm	Selected	Selected	Selected	Selected

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 2.4: Metro concentration's shortening effect is robust to alternate specifications.

The first stage of expansion adds population variables to the offered set: raw population (logCBSAPopulation) and population density (logCBSAPopDensityMilesSq). I construct metro population by summing the 2010 Census population totals for each metro's component counties according to their definitions as of February 2013, and I construct metro population density by normalizing metro population by the 2010 Census's land areas for the each metro's component counties. I then log each value to reduce the influence of any outliers.

Concerning the data for these variables, I use the 2010 Census in particular simply because it lies in the middle of our sample, and I use the February 2013 metro definitions because they are the first vintage to make use of the 2010 Census data. Land areas are taken from the 2020 Census Gazetteer.[20]

Theoretically, population could easily be a determinant of outbreak length. On one hand, larger populations might allow for better-staffed and better-equipped public health departments and secure any benefits to scale that derive therefrom. Indeed, metro concentration might very well enhance any such effect since having the same volume of expanded resources all under a single state's control would likely allow for further specialization and reduced role redundancy.

On the other hand, larger populations could complicate matters by inviting more complex and extensive outbreaks. Added restaurants securing supplies from a larger variety of commercial food distributors and additional commuters who might purchase meals anywhere along their journeys both add readily available nodes for any potential outbreak.

Similarly, population density could *easily* play a part. On one hand, greater density could reduce a public health investigator's travel time, speed up their investigations and implementation of control measures, and ultimately shorten outbreaks. On the other hand, greater density could enhance an outbreak's growth speed, complexity, and length.

Despite these plausible modes of influence, somewhat surprisingly, the lasso roundly rejects both of these variables in all of the expansions. Whatever effects they might have, they are not sufficient to displace metro concentration or even to merit inclusion.

The second stage of expansion adds our climatological variables, including a metro's heating degree days (cbsaHeatingDegreeDaysAnnualAvg), cooling degree days (cbsaCoolingDegreeDaysAnnualAvg), and elevation (cbsaAnnAvgsStationElevation). As I have already discussed, heating degree days capture how much a metro's daily mean temperature exceeds 65 °F over the course of an average year. Cooling degree days are analogous, instead capturing the extent to which a metro's daily mean temperature falls *below* 65 °F over the course of an average year. Lastly, my elevation measure is simply the elevation of the weather station from which I source a metro's degree days values.

[20]US Census Bureau 2022.

Theoretically, cooling degree days could influence outbreak length for the same reasons that heating degree days do: temperature affects the activities and diets of humans as well as the behavior and survival of the animals, plants, and other organisms that may facilitate humans' chemical intoxication. Most importantly, cooler metro areas and the areas from which they can cheaply source food are likely inhospitable for the tropical organisms from which people most frequently contract chemical foodborne illness.

Since I control for temperature, elevation's potential influence stems primarily from its role in determining air density and, in turn, oxygen availability. At higher elevations, hypoxia can have an impact on humans by impairing cognition and altering the immune system, potentially making outbreaks more difficult for public health workers to solve or rendering more individuals susceptible to harmful effects at lower toxin doses.[21] The effects on other forms of life are presumably similarly severe, contributing like temperature to whether the organisms that facilitate chemical agent generation and distribution even survive.

As before, despite these variables' theoretical potential, the lasso typically rejects everything but heating degree days. Elevation is rejected every time it's offered while cooling degree days manages to secure inclusion only in the second stage expansion. Even when the latter is included, it does not explain away metro concentration's effect.

My third stage of expansion adds variables for seasonality: a full set of indicator variables (monthFirstIllX) for the month in which symptoms were first experienced by any case in an outbreak. As I have already established, there is good theoretical reason to suspect that seasonal effects exist for outbreaks of chemical foodborne illness; and splitting the year up into months instead of seasons or weeks affords the lasso greater flexibility to pinpoint those effects without spreading the 223 outbreaks across so many time periods that OLS lacks sufficient variance to estimate *any* of their effects with reasonable precision.

As Table 2.2 already showed, outbreak length is especially pronounced in the final three months of the year, but the lasso only reliably selects the November indicator. It is possible that what appears on first blush as a broader seasonal effect is better captured by one of the other selected variables. Alternatively, there may just be too little data accurately to estimate the complete seasonal effect.

Whatever the case, in this third expansion, the concentration measure once again survives selection, and the effect remains negative and statistically significant; and, whatever the seasonality effect, it is insufficient to displace the Herfindahl.

In the final expansion, I add an indicator for whether an outbreak occurred after the implementation of SEDRIC. Sensitive to the costs of a fractured public health information network, the CDC debuted

[21] Virués-Ortega et al. 2004; Manella et al. 2022.

SEDRIC to consolidate and analyze the nation's outbreak-related data;[22] however, its rollout was far from an instantaneous event. The pilot launched in August 2010, and the system was pitched to the wider U.S. public health community at a conference on September 20, 2011;[23] but knowing about a platform and actually adopting it are two very different things.

Like most programs run by the CDC, SEDRIC depends on voluntary participation by states and localities; and it is unclear exactly when and to what extent each of them incorporated SEDRIC into its workflows. Simple user base size offers some insight. By 2015, the user base had grown to 225 people, with at least one in each state.[24] By late 2016, that number had swelled to 450;[25] and it seems to have remained there ever since.[26] Unfortunately, no public numbers exist for the years before 2015.

SEDRIC training sessions at the nation's public health conferences are another heuristic for the nation's progressive uptake of the system. Such sessions were offered at least as early as 2013 and continued to be offered at least through 2019.[27] The fact that these sessions happened when they did likely indicates a combination of several things: a lack of complete program uptake among the nation's public health units, a continued desire among some of them to pursue further adoption, and a desire among the system's administrators to cultivate further interest. Of course, some of the sustained demand could be a simple result of job turnover within public health departments, which could attenuate the measure's reflectivity of uptake. In addition, the apparent cessation of training sessions in and after 2020 could be more a result of COVID forcing the use of alternative training mediums than any exhaustion of demand for such training.

This ambiguity in both the timing and location of launch and adoption does not lend itself to sharp—and certainly not to jurisdiction-specific—measures of SEDRIC's presence. In this case, a general indicator for whether an outbreak occurred before or after the system's rollout seems most sensible.

As for what date to use for SEDRIC's beginning, the general launch is superior to the pilot launch because it runs a lower risk of masking SEDRIC's effect. Though the system was in operation with the pilot, its jurisdictional reach was presumably limited; so its likelihood of influence in any given outbreak during the period was probably low. In other words, most outbreak investigations that occurred during the pilot likely *didn't* benefit from SEDRIC; so including them in the "post-SEDRIC" group would unfairly dilute the system's effect.

[22]See Centers for Disease Control and Prevention 2020b.

[23]Williams and Nguyen 2011.

[24]Food Safety Modernization Act Surveillance Working Group, Board of Scientific Counselors, Office of Infectious Diseases 2015

[25]Centers for Disease Control and Prevention 2016b.

[26]Centers for Disease Control and Prevention 2020b.

[27]See Association of Public Health Laboratories 2013, Association of Public Health Laboratories 2014a, Association of Public Health Laboratories 2014b, U.S. Centers for Disease Control and Prevention 2015, Association of Public Health Laboratories 2016, Reamer 2016, and U.S. Centers for Disease Control and Prevention 2019.

Even after the general launch, uptake took years; so using the general launch still dilutes SEDRIC's effect. However, it is at least likely that most states were onboarded shortly after the general launch while most almost certainly were *not* during the pilot.

Unfortunately, beyond a year, the general launch's date is not publicly reported; but the creation date on the CDC's PowerPoint presentation describing SEDRIC's core features is March 17, 2011.[28] The earliest iterations of this presentation were probably used to introduce the system to representatives of the states' public health departments; so its original creation likely predates any meaningful system usage. I therefore adopt that date as my post-SEDRIC cutoff.

When the lasso is offered the SEDRIC indicator, it roundly rejects it, choosing instead to retain the model it chose in the last expansion. Thus, all estimates including that for the metro concentration variable remain unchanged. Though SEDRIC was designed to counter the frictions that state borders introduce, this analysis suggests that the system has not been able wholly to eliminate those frictions. Part of the reason likely lies in the fact that SEDRIC can only address *informational* frictions—not *control* frictions; creating a national information-sharing platform doesn't establish a commander-in-chief for the nation's public health infrastructure. Another plausible reason is a lack of universal uptake; with only 450 users when there are 3,290 public health departments at the local level alone, SEDRIC simply may not have been fully integrated into daily workflows.[29]

Naturally, this evaluative approach has its limitations. The indicator variable essentially compares a bunch of non-SEDRIC outbreaks with a combination of both SEDRIC *and* non-SEDRIC outbreaks. To the extent that it exists, SEDRIC's influence will be masked in this setup; so its effectiveness may very well be stronger than this analysis suggests. Unfortunately, the publicly-available data just does not allow for a higher-precision approach.

Finally in this section, I turn to incubation period—the period from exposure to an agent to the presentation of symptoms. It's plausible that longer incubation periods would lead to longer outbreaks. When a toxin avoids inducing symptoms for longer, it gives itself more of an opportunity to intoxicate more people (in the non-alcoholic sense) before the public health authorities can learn of its activities and intervene.

To take a simple example, if a restaurant has a batch of tuna that contains ciguatoxin, it will likely serve meals out of that batch either until the batch runs out or until one of the people who consumed the tuna comes down with symptoms of ciguatera poisoning. If those symptoms present relatively quickly,

[28] Centers for Disease Control and Prevention 2020c.
[29] National Association of County and City Health Officials 2022.

then the toxin-bearing tuna batch can be identified and thrown out before it is exhausted, shortening the outbreak.

In addition, an especially long incubation period that occurs in the last person intoxicated in an outbreak could lengthen that outbreak simply by delaying symptom onset. While not an especially interesting mechanism, it would certainly affect the data.

Now, concerning which measure of incubation period to use, the longest, the shortest, and the median are the available options; and each can shed light on some part of the outbreak investigation.[30] On the leading side, the first incubation period to end marks the first opportunity a public health department has to learn of and to begin responding to an outbreak; but the first to end is not necessarily the shortest. The median will generally be the best estimate, but the others provide the outer limits.

Towards the middle of an outbreak investigation is when the case reports have started arriving in larger numbers, enabling the authorities to narrow down the causes. The median indicates how long the authorities will have to wait to hear about most cases; but, notably, those waiting periods are generally spread out along with the exposures that began them.

Finally, once an outbreak investigation is concluded, the control measures fully implemented, and further intoxication halted, the last few exposures in the outbreak will still prolong that outbreak's observed length. As with the first period to end, the last period to end will typically be of median length, but the shortest and longest periods provide its outer limits. Accordingly, I use use all three.

Unfortunately, incubation period is unavailable in 23% of the sample; so including it comes at the cost of a lot of statistical power. In addition, when the SEDRIC model's set of offered variables is expanded to include the incubation period measures, the lasso selects only the intercept. So, to extract at least *some* information on whether incubation period can explain away the metro concentration effect, I pit the two against one another in Table 2.5. Here, none of the three incubation period measures are statistically significant; and, while attenuated, the metro concentration effect remains negative and statistically significant.

[30]Note that roughly 4% of outbreaks are the result of multiple etiologies, and the outbreak median is reported for *all* associated etiologies—not just chemicals.

	DV: logOnePlusExposureLengthDays
	Simple Incubation Model
(Intercept)	1.785 ***
	(0.000)
logOnePlusIncubationMedianDays	0.106
	(0.761)
logOnePlusIncubationLongDays	0.404
	(0.119)
logOnePlusIncubationShortDays	-0.793
	(0.272)
cbsaHerfindahlNorm	-0.818 *
	(0.022)
N	172
R2	0.060
logLik	-178.780
deltaBIC	10.257

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 2.5: The metro concentration effect survives a head-to-head contest with incubation period.

3 Major Alternative Hypotheses

While I have addressed a number of implicit alternative hypotheses up to this point via control variables, I turn now to alternative hypotheses which demand special attention: that either port traffic or politics is driving the effects I detect.

3.1 Ports

Many of the more fractured cities in the United States are also port cities, which is to be expected.[31] Because rivers ease transportation into the continent, port cities often straddle them; but, because rivers impair governance by inhibiting travel *within* the continent, many of them became state borders. This overlap means that metro concentration could be a red herring—a mere heuristic for a city's port status and volume.

Port cities, almost by definition, could be more likely to have longer outbreaks. From one direction, the port-specific traffic coming into the city from both land and sea could introduce a comparatively greater variety and volume of intoxicated foods from a larger number of sources than a non-port city might encounter; the variety in particular could make initial detection more difficult, allowing an outbreak more time to spread.

In the other direction, port-specific traffic leaving the city could extend an outbreak by making it more difficult to track down all the intoxicated food associated with that outbreak before it's consumed. To take a simple example, say a trucker picks up her load from a shipping terminal, buys a tuna sandwich from a convenience store to eat later down the road, and starts on her way. If the convenience store learns later that day that the tuna in that sandwich is intoxicated, it and the local health department could have an especially difficult time notifying the trucker. She's probably not part of local hearsay networks or at least not likely to get updates from that network for a few days; she's unlikely to stop by the convenience store again before she eats the sandwich; and, even if the outbreak is significant enough for an advisory to be included in the local evening news, she's likely to be well clear of the local media market when it airs.

To evaluate these possibilities and to determine whether they can explain away the metro concentration effect, I need a measure of port activity. The U.S. Department of Transportation's Bureau of Transportation Statistics tracks several metrics for the nation's ports that capture that concept: import tons, total tons, and vessel calls.[32] Import tons is the number of tons of cargo that came in through the port from a foreign source; total tons is the number of tons of either foreign or domestic cargo that either came

[31] See the balance regression in subsection C.2 for statistical evidence on this point.
[32] Tonnage is reported in short tons, which are 2,000 lbs. each.

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

in or left through the port;[33] and vessel calls counts the number of times any waterborne vessel visited the port. For all three, I use the annual total, averaged across all years for which a port has data; while per-year annual values might be superior, such values are not reported for all ports in all years.

Each metric captures a different mechanism of port activity's potential influence on outbreak length. Import tons and vessel calls both capture waterborne and international activity, but imports emphasizes the volume of goods while vessel calls emphasizes the volume of people; after all, each additional vessel visit, no matter the vessel's size, will almost always bring a fresh crew to the port. On the other hand, total tons captures both land and sea activity, domestic and foreign; just as foodborne illness can come in from another country and over water, it can also come in from a nearby city and on the back of a semitruck.

Importantly, these metrics are reported for ports and port organizations, which cannot always be uniquely associated with a metro. For example, the Port of Virginia has locations in both the Richmond and the Norfolk metros; and, because the data is reported for the organization and not for each of its locations, it is impossible properly to split the traffic between the metros. Unfortunately, complicating matters further, the dataset does not report the location of each port's terminals.

To remedy these issues, first, I determine each port's terminal locations by referencing its website. Second, I assign a port to a metro if and only if its terminal locations are at least partially contained within metro and not within any others. So, for cases like the Port of Virginia, I do not assign the port to *any* metro. For metros that contain more than one port, I simply sum the values for the component ports. The prime example here is the Los Angeles metro, which contains both the Port of Los Angeles and the Port of Long Beach.

Unfortunately, the non-assignment approach leaves data missing for a number of metros. In addition, while the DOT dataset captures the nation's largest seaports—and most of its smaller ones—it may not be comprehensive. I do not assume that a metro had no port activity simply because its associated port is missing from the dataset; instead, I treat all such metros as having missing values for the port activity variables.

Since I use listwise deletion, this missingness could seriously hamper my statistical power by eliminating much of my dataset. To forestall that elimination, for each port volume variable, I add an indicator for whether the value in that variable is missing, and I change all of those missing values to 0. This gives

[33]Tonnage can be broken down according to whether it is foreign or domestic and by whether the port is the beginning or the end of the cargo's waterborne journey. Imports and exports have a foreign origin or destination, respectively; and cargo with a domestic origin or destination is all reported under the unified category "domestic cargo."

the variance associated with those missing values somewhere to pool such that it does not interfere with the other coefficients' estimates, and it keeps me from having to drop observations.

Next, because of the potentially substantial collinearity among the port activity measures, I carry only one of the three forward to the broader regression with metro concentration. To decide among them, for each variable, I first run a regression that includes only the variable (transformed according to the missingness procedure) and its associated missingness indicator. I then compare common measures of how much information each variable provides.

Table 3.1 reports these regressions and their associated information measures. Since all three regressions use the same sample, their information measures can be usefully compared. Among the three, vessel calls maximizes both R^2 and log-likelihood while minimizing deltaBIC, thus uniformly meriting selection. Note that, in this case, the deltaBIC serves primarily to gauge the measures' performance specifically in the Simple Model's sample rather than in the larger sample the models in this table use.

Briefly to dismiss a tempting but erroneous alternative means of inference, the comparative size and precision of the variables' coefficients cannot be trusted properly to identify the most appropriate variable. The (non-indicator) variables here are not normalized by their standard deviations, and their units are small since vessel calls can exceed 9,000 while total tons can exceed 145 million. However, in this case, comparative size does not mislead: vessel calls is the preferred variable according to the BIC.

As Table 3.2 shows, adding vessel calls to either the Lasso-Preferred Model or the Simple Model does not explain away metro concentration's effect. In both cases, the effect remains negative and statistically significant; and, even if that were not the case, the deltaBIC prefers the Simple Model over these variations.

Finally, because of the comparatively sparse amount of data for the port variables, I run several partial F-tests to determine whether any combination of the vessel calls variables is still useful to include despite their statistical insignificance on a per-variable basis.[34] In each test, I compare a full model including those variables against a base or nested model without them; the resulting p value evaluates the null hypothesis that the full model does *not* fit the data significantly better than does the base model. As Table 3.3 demonstrates, for base models on a spectrum from the Simple Model to the Lasso-Preferred Model, the null is not rejected. The port activity variables do not unseat the metro concentration effect; and, weakly speaking, adding them does not provide an enhanced explanation of the data.

[34] See Navarro 2022, §16.5.

Data sparsity is unlikely to be the cause of the port activity variables' ineffectiveness. For all models that are party to one of the F-tests, port activity is available for 112 outbreaks. That should be enough observations for a port effect to show itself, but no effect *does* show up. Given this, metro concentration's putative effect seems all the more likely to be genuine. The reduction in outbreak length that more concentrated metros enjoy is not attributable to reduced port activity but to metro concentration itself. More unified control and less friction in information exchange indeed shorten outbreaks.

	DV: logOnePlusExposureLengthDays		
	Import Tons	Total Tons	Vessel Calls
(Intercept)	0.96264 ***	0.93402 ***	0.91770 ***
	(0.00000)	(0.00000)	(0.00000)
cbsaPortImportShortTonsAnnualAvg	0.00000		
	(0.66980)		
cbsaPortImportVolumeIsMissing	0.11954		
	(0.23154)		
cbsaPortTotalShortTonsAnnualAvg		0.00000	
		(0.46193)	
cbsaPortTotalShortTonsIsMissing		0.14816	
		(0.17905)	
cbsaPortTotalVesselCallsAnnualAvg			0.00006
			(0.09306)
cbsaPortTotalVesselCallsIsMissing			0.16448
			(0.08704)
N	233	233	233
R2	0.00631	0.00787	0.01767
logLik	-236.64094	-236.45851	-235.30139
deltaBIC	13.37123	12.98293	10.51848

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.1: Among the three measures of volume, a CBSA's number of vessel calls is the measure most predictive of outbreak length.

	DV: logOnePlusExposureLengthDays	
	Preferred + Ports	Herf + Vessel Calls
(Intercept)	1.65350 ***	1.81068 ***
	(0.00001)	(0.00000)
cbsaHeatingDegreeDaysAnnualAvg	0.00003	
	(0.19983)	
monthFirstIllNovember	0.30266	
	(0.07649)	
cbsaHerfindahlNorm	-0.78713 *	-0.90969 **
	(0.02842)	(0.00703)
cbsaPortTotalVesselCallsAnnualAvg	-0.00001	0.00001
	(0.70374)	(0.73634)
cbsaPortTotalVesselCallsIsMissing	0.06242	0.13874
	(0.54289)	(0.13897)
N	223	223
R2	0.07149	0.05105
logLik	-211.99208	-214.42029
deltaBIC	14.46896	8.51102

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.2: Port throughput cannot explain away the metro concentration effect. The Herfindahl's effect survives when I add the vessel call variables to the Lasso-Preferred Model and when I put it in a head-to-head contest with those variables. The lasso (not shown) rejects the port variables in favor of the Lasso-Preferred Model.

	(1)	(2)	(3)
Partial F-test p Value	0.323	0.698	0.667
Base Model Variable(s)			
cbsaHerfindahlNorm	Included	Included	Included
cbsaHeatingDegreeDaysAnnualAvg		Included	Included
monthFirstIllNovember			Included

Table 3.3: The partial F-tests indicate that the vessel calls variables should *not* be added to the Lasso-Preferred Model. Even when the base/nested model has fewer variables, the vessel calls variables still do not merit inclusion.

3.2 Partisan Politics

In a project like this one wherein federalism and public health are involved, the potential role of partisanship simply cannot be ignored. Party differences can hinder cooperation; partisanship itself—both among elites and among voters—can lend itself to more or less aggressive public health measures.[35] In this section, I evaluate each of these alternative hypotheses—both for their own sake and to determine whether they can explain away the metro concentration effect.

3.2.1 Partisan Friction among Elites

Like any group identity, partisanship can enhance cooperation among the in-group and hinder cooperation between the in-group and the out-group. In this context, that fact could be especially important for state governors. Like presidents, state governors appoint individuals to control arms of the bureaucracy;[36] so their partisan affiliations get transferred down into that bureaucracy. In addition, even when governors do not directly constitute the bureaucracy, they wield budgetary and administrative influence that could be used to persuade an insulated bureaucracy to behave as if they were the governor's copartisans.

Because investigations of interstate outbreaks are typically coordinated among state agencies as well as the CDC, any partisan differences among the leaders of those state agencies could hinder cooperation and lengthen outbreaks; and, since governor partisanship can be effectively transferred down, partisan differences among governors could also lengthen outbreaks.

To test if this is the case, I add an indicator variable for whether, at the time an outbreak's first case experienced symptoms, the governors with jurisdiction over any part of the outbreak's metro had different national party affiliations from one another. In the case of Washington, D.C., I treat the district's mayor as its governor-equivalent.

To assemble the dataset of state governors and their party affiliations, I draw initial data from the National Governor Association's (NGA) database of former governors and compare it with Wikipedia's rosters of governors for each state and state equivalent.[37,38] For cases in which the two conflict, I resolve the issue by appeal to historical newspaper articles, Ballotpedia, and/or official state sources.

[35] Indeed, even intraparty factions sometimes disagree with their parties on such matters; but I am unable to test for the influence of that effect with this dataset.

[36] In South Carolina, for example, the governor appoints the members of the board which, in turn, selects the director of the state's public health agency (South Carolina 2012).

[37] Though not peer-reviewed, the data from Wikipedia proved overwhelmingly more up-to-date, detailed, and correct than any other available source. The NGA database sometimes includes year but not date; its years are sometimes wrong; and it sometimes wholly omits governors. By comparison, Wikipedia's most grievous error was recording a term start date that was off by a day.

Note that the only available alternative to either of these sources is Kaplan 2020, which is itself drawn entirely from the NGA.

[38] National Governors Association 2022.

Local Party	National Party
Alaskan Independence Party	Republican
A Connecticut Party	Independent
Democratic-Farmer-Labor Party	Democratic
Minnesota Independence Party	Republican

Table 3.4: Local & regional parties' national party equivalents.

I treat each governor as having had control for the entire day on their first day in office, regardless of the time at which they actually assumed office.[39] For cases in which a sitting governor changes party affiliation in the middle of their term, I use their avowed party at the time of the outbreak.

To ensure reasonable comparability between states, where possible, I use party platform to select a corresponding national party (i.e. Democrat or Republican) for smaller, more regional parties. If the smaller party's platform does not sufficiently resemble that of one of the national parties, or if its platform either draws freely from those of multiple national parties or does not resemble that of any national party, I classify the associated governor(s) as an independent.

For example, one Alaska governor was a member of the Alaska Independence Party. The party's 2020 platform endorses, among other things, strict adherence to the U.S. Constitution; defense of state, individual, and property right; eliminating the word "privilege" from Alaskan statutes; elimination of all property and income taxes; privatization of government services; parents' control over their children's education; traditional family; and banning abortion, euthanasia, and infanticide.[40] Since these track the national Republican coalition's positions fairly closely, and because the platform's remaining positions conform to neither Republican nor Democratic positions, I classify that governor as a Republican at the national level. Table 3.4 catalogs the full set of equivalencies.[41]

Measured in this manner, partisan discord among elites neither affects outbreak length nor unseats metro concentration's effect. Table 3.5 shows two analyses supporting this position. The first column is a bivariate regression with only metro concentration and state governor discord; this setup gives discord the opportunity to unseat metro concentration while reserving as much variance as a possible to allow for a more precise estimation of both coefficients. While metro concentration's effect remains negative and significant, discord's effect is statistically insignificant.

[39] A regular inauguration, a sitting governor's resignation (generally in favor of higher office), and a sitting governor's death all occasion a change of power at least once.

[40] Alaska Independence Party 2020

[41] The classifications for Puerto Rico's governors do not affect the analysis since none of the territory's CBSAs straddle state lines. However, for the sake of having a complete dataset, I classify them according to whether they caucus with the Democratic Governors Association or the Republican Governors Association. Platform-based classifications are impracticable in this case because the island's partisan politics are essentially orthogonal to those of the rest of the country. The primary fissure concerns the future of Puerto Rico's sovereign status as a state, territory, freely associated state, or independent nation; and the island's major parties do not select positions on peripheral issues in a clearly and consistently Republican or Democratic fashion.

The second column is the lasso-selected model when I add state governor discord to its set of available variables. The lasso rejects discord and retains its preferred model, wherein metro concentration's effect is both negative and significant. In addition to being statistically insignificant when forcibly included, the presence of partisan discord among governors is not helpful enough in explaining the data to merit inclusion in the model.

	DV: logOnePlusExposureLengthDays	
	Herf + Discord	Lasso
(Intercept)	0.050	-0.034
	(0.493)	(0.615)
cbsaStateGovDiscord	-0.447	
	(0.123)	
cbsaHerfindahlNorm	-0.301 **	-0.147 *
	(0.001)	(0.036)
cbsaHeatingDegreeDaysAnnualAvg		0.125
		(0.072)
monthFirstIll11		0.424
		(0.080)
N	223	223
R2	0.052	0.068
logLik	-310.022	-308.068
deltaBIC	2.987	4.487

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.5: When the state governors with jurisdiction over some part of a given metro are not all members of the same political party, that difference has no statistically significant impact on outbreak length, and it does not explain away the influence of metro concentration.

3.2.2 Governor Partisanship

Another way in which partisan politics might influence outbreak length is through sheer policy position. Over the course of the coronavirus pandemic, Republicans developed a reputation for being opposed to public health measures. President Trump, a Republican, was quoted many times saying, "We can't let the cure be worse than the problem itself," essentially arguing that some public health measures were not worth the economic costs.[42] Republican-led Georgia pursued a similar line, encouraging private businesses and performance venues to remain open by largely shielding them from liability for COVID contracted on their premises.[43] Republican governors and attorneys general vigorously opposed the Biden administration's vaccine mandates.[44] And several Republican-led states banned mask mandates for the general public.[45]

The themes of privacy, freedom to contract, small government, and economic expedience could have similar implications in the realm of foodborne illness outbreaks. Foodborne illness prevention and control involve the centralized collection of private health information and the forcible regulation and inspection of businesses and their records. Republicans might not want their personal or business information logged into yet another government database. They might find food safety regulations to be an invasive compromise of the private freedom to contract—something for the market to decide. And they might contend that some regulations cost more than they're worth—that the "cure [is] worse than the problem itself." If so, then outbreaks could be longer in metro areas that have a section subject to a Republican governor's jurisdiction.

In reciprocal fashion, over the course of the pandemic, Democrats developed a reputation for being in favor of more stringent public health measures. As I already noted, President Biden, a Democrat, attempted to mandate COVID vaccination for workers on a very large scale. Democrat-led California pursued requiring K-12 school children to be vaccinated against COVID, though it ultimately relented.[46] And, as of February 25, 2022, only Democrat-led states still required masking for the general public.[47]

The themes of protection, safety, and big government could easily transfer from COVID to food safety. Less concerned about privacy and economic cost when a compelling government interest like health is at stake, Democrats might be willing to wield public health investigation and control powers more aggressively. In addition, they might fund their public health infrastructure better than Republicans do. If so, then outbreaks could be shorter in metro areas that have a section subject to a Democratic governor's jurisdiction.

[42] See Samuels and Klar 2020.
[43] See Fisher & Phillips LLP 2021.
[44] See DeMillo and Mulvihill 2021.
[45] See Avila et al. 2022.
[46] See Aguilera 2022.
[47] See Avila et al. 2022.

To test these hypotheses—and to ensure that the absence of these effects is not driving my result on metro concentration—I construct two indicator variables: one for whether a Democratic governor has jurisdiction over any part of an outbreak's metro and one for whether a Republican governor has jurisdiction over any of those parts. Importantly, these are not simply inverted versions of one another; instead, they mark the state at two ends of a spectrum from total Republican control to total Democratic control (excluding third parties).[48]

Table 3.6 shows examples of how these variables might be valued. For the first outbreak, the only governor with jurisdiction, Linda Lingle, was a Republican; so only atLeast1Rep's value is 1. For the second outbreak, a mix of Republican and Democratic governors held jurisdiction. In August 2010, Ed Rendell (D-PA), Chris Christie (R-NJ), and David Paterson (D-NY) all controlled some part of the New York metro area; so both atLeast1Rep's and atLeast1Dem's values were 1. For the final outbreak, only Democratic governors Jay Nixon (D-MO) and Mark Parkinson (D-KS) held jurisdiction over some part of the Kansas City metro in August 2009; so only atLeast1Dem's value was 1.

Outbreak Date	**Metro Area**	**Repub. Governors' 2-Party Share of Metro Pop.**	**atLeast1Dem**	**atLeast1Rep**
Apr. 10, 2009	Urban Honolulu, HI	100%	0	1
Aug. 6, 2010	New York-Newark-Jersey City, NY-NJ-PA	33%	1	1
Aug. 11, 2009	Kansas City, MO-KS	0%	1	0

Table 3.6: The "at least 1" variables are not mutually exclusive. Some jurisdictions have governors only from one party at the time of an outbreak; others have a mix of both Democratic *and* Republican governors when an outbreak occurs.

As Table 3.7 demonstrates, metro concentration's effect remains both negative and statistically significant in the presence of each of these variables; the partisanship of the governors and state administrations holding jurisdiction over some part of the metro does not explain it away. As for the hypotheses on partisan influence, the presence of Democratic governors clearly does not have a statistically significant impact; but, on first blush, the presence of a Republican governor *does* seem to reduce outbreak length in a statistically significant way. It turns out that this latter result is a consequence of a single, rather unusual outlier.

This outlier outbreak occurred in the Minneapolis-St. Paul-Bloomington metro area from October through December 2020 during Minnesota and Wisconsin's first major COVID wave, which spanned essentially the same period.[49] Minnesota public health in particular has a stellar reputation,[50] but even the best

[48]Third-party governor control does not come up among chemical outbreaks but does among some of the other etiologies.
[49]"Wisconsin Coronavirus Map and Case Count" 2020 and The New York Times 2020b.
[50]Morelli 2018.

	DV: logOnePlusExposureLengthDays	
	Dem. Gov. Present	Rep. Gov. Present
cbsaHerfindahlNorm	-0.165 *	-0.211 **
	(0.022)	(0.001)
cbsaAtLeast1DemGov	0.197	
	(0.182)	
cbsaAtLeast1RepGov		-0.315 *
		(0.033)
N	223	223
R2	0.049	0.061
logLik	-310.325	-308.918
deltaBIC	3.595	0.781

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.7: The partisan affiliations of the governors holding jurisdiction over a city do not explain away the metro concentration effect. While Republican presence appears to shorten outbreaks, this is the result of an unusual outlier.

departments will be hard-pressed to perform their normal duties well when a shock as large as COVID's first wave strikes and demands an all-hands-on-deck response. In addition, at the time, the effects of George Floyd's murder may still have been working their way through Minnesota's government—including its public health department;[51] while they were (understandably) focused on reallocating resources to identify and to right racial injustice, their performance of normal operations including foodborne illness control might have suffered.

As Table 3.8 shows, when I exclude this outlier from the dataset, Republican governor presence loses significance; Democratic governor presence remains insignificant; and metro concentration remains significant in both models.

Now, one might be concerned that these conditions of exclusion might include other outbreaks. On the COVID side, it turns out that the Minneapolis outbreak is the only one whose length is likely to have been affected by COVID. Only three other outbreaks in the dataset occur after the CDC first advised the nation's clinicians to be on the lookout for COVID (January 8, 2020)—the earliest point at which the virus

[51] Minnesota Department of Health 2023. Note that the City of Minneapolis's public health department does not perform foodborne illness outbreak investigations.

	DV: logOnePlusExposureLengthDays	
	Dem. Gov. Present	Rep. Gov. Present
cbsaHerfindahlNorm	-0.184 *	-0.218 **
	(0.011)	(0.001)
cbsaAtLeast1DemGov	0.143	
	(0.335)	
cbsaAtLeast1RepGov		-0.253
		(0.088)
N	222	222
R2	0.049	0.057
logLik	-308.950	-307.943
deltaBIC	3.595	0.781

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.8: When the post-COVID, post-Floyd Minneapolis outlier is dropped, Republican presence is no longer significant; but all other results remain the same.

could be expected seriously to impair food safety operations.[52] The first takes place in Florida in January 2020—well before the virus is known to have reached the state. The New York Times 2020a. The other two take place in Puerto Rico in 2020, and the territory did not see its first major COVID spike until well into 2021.[53] Still, as the first column of Table 3.9 shows, if I simply exclude all outbreaks that occur after COVID, metro concentration remains significant while the presence of a Republican governor remains insignificant.

On the George Floyd condition, the Minneapolis outbreak is the *only* chemical outbreak to occur after George Floyd's death on May 25, 2020—the earliest date at which the event could be expected to affect public health operations.[54] Thus, the models in Table 3.8 that omit this outlier already show that the metro concentration findings are robust to the exclusion of post-Floyd outbreaks. However, in case there is something unusual about Minneapolis or Minnesota more generally, Table 3.9 includes a column demonstrating that the findings are also robust to the exclusion of outbreaks from those jurisdictions. I only include one column for both since the same outbreaks are excluded by each restriction.

[52] Centers for Disease Control and Prevention 2020a.
[53] The New York Times 2020c.
[54] Hill et al. 2020.

	DV: logOnePlusExposureLengthDays	
	Pre-COVID	No Minneapolis/Minnesota
cbsaHerfindahlNorm	-0.926 **	-0.934 **
	(0.001)	(0.001)
cbsaAtLeast1RepGov	-0.154	-0.152
	(0.096)	(0.102)
N	220	218
R2	0.056	0.057
logLik	-200.939	-199.874
deltaBIC	0.781	0.781

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.9: When I exclude post-COVID outbreaks (using US dates), outbreaks from the Minneapolis metro, or outbreaks from any metro that's partly in Minnesota, metro concentration remains significant while the effect of a Republican governor's presence goes away.

To provide further evidence concerning the potential influence of Republican governors, I divide the full sample, including the outlier, into outbreaks that occur in single-state metros and those that occur in multi-state metros (think San Francisco and New York, respectively); the outlier is in a multi-state metro. If Republican governor presence is significant among the former, then the effect can be attributed more cleanly to party *itself*; there are no state borders in the mix to inhibit information transfer and coordination, and there cannot be partisan strife among state administrations where only one is involved. As the first column of Table 3.10 shows, Republican governor presence is *not* significant among such outbreaks; thus, whatever partisan effect there might be, contra the hypotheses, it is not simply that Republicans run their public health operations better or worse than their Democratic counterparts. Unlike with COVID control measures, food safety could be fairly equally demanded by people of both parties, with the governors responding accordingly.

In view of this, if Republican presence is significant among outbreaks in multi-state metros, then a more nuanced partisan mechanism could be at work. As Table 3.10's second and third columns show, Republican governor presence *is* significant among outbreaks in multi-state metros—but only when I include the outlier.

In all likelihood, this effect is just a result of noise. However, if not, it could be that adding a Republican to a group of Democrats creates a sense of competition among the parties; but the null result

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

on the presence of a Democratic governor in previous models suggests that any such "mixed-party" dynamic only emerges when the ratio of Republican control to Democratic control is low. Perhaps the Democratic administrations in a mostly Democrat-controlled metro feel they can pull the few laissez-faire Republican administrations along, but Democratic administrations in metros with wider Republican control think the task is too difficult.

Whatever the case, returning to the original purpose of these analyses, among outbreaks in multi-state metros, metro concentration has a negative and statistically significant effect as long as the outlier is excluded. While the qualification might seem like cause for concern, recall that metro concentration is significant in the full sample—including the outlier. The multi-state-only sample includes only 18% of the available observations, and such substantial sample reductions can easily rob variables of their statistical power.

In summary, metro concentration's effect is clearly not the result of omitting governor party; and, contra the hypotheses, the mere presence of a governor from one party or the other does not have a statistically significant impact on outbreak length.

	DV: logOnePlusExposureLengthDays		
	Single-State	Multi-State	Multi-State w/o Outlier
cbsaHerfindahlNorm		-0.960	-1.563 *
		(0.194)	(0.016)
cbsaAtLeast1RepGov	-0.127	-0.623 *	-0.307
	(0.196)	(0.042)	(0.237)
N	183	40	39
R2	0.009	0.136	0.180
logLik	-164.181	-45.253	-36.626
deltaBIC	0.781	0.781	0.781

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.10: Republican governorship is only significant among outbreaks in multi-state metros, showing that simple party is not responsible for the effect; and the effect goes away when I exclude the outlier. While metro concentration loses significance in the multi-state outbreaks model, this should not be interpreted as evidence against city concentration's effect; that model is estimated on only 18% of my sample.

3.2.3 Metro Partisanship

Beyond state partisanship, the partisanship of the metro's population could also play a role in determining the length of chemical outbreaks. For example, individual Democrats might be more inclined to report their illness because of a stronger belief in public health and a greater comfort with regulation and big government.

On the other hand, skepticism of government is present in segments of both parties; and that skepticism could hamper personal engagement with public health, lengthening outbreaks. For Republicans, libertarians are the most vocal skeptics, though the business and rural segments are often similarly disenamored of government. For Democrats, the Black community—long a pillar of the Democratic coalition—is largely skeptical of police; and that skepticism could transfer to other arms of government. When the public health officer calls to collect a case's dietary history, people in these groups might be less likely to respond, complicating the officer's job of pinpointing an outbreak's cause and, as a result, extending the outbreak.

Skepticism of healthcare is also present in both parties and could reduce engagement with public health. On the Republican side, rural communities are often skeptical of medicine—perhaps because of an enduring sense from past generations that medicine tends to do more harm than good.[55] On the Democratic side, the Black community harbors a similar skepticism thanks in part to the deliberate mistreatment its members received in programs like the Tuskegee Syphilis Study. For both groups, an aversion to medicine—as well as a sheer lack of access—could prevent them from going to see a medical professional about their symptoms; and, while physicians and medical laboratories are generally supposed to inform the public health bureaucracy of any potential indications of chemical outbreaks, they can't do that if they never even see the patient.

Given the ambivalence of these forces, it is not clear in which direction the coefficient should point; but the potential importance of a metro population's partisan affiliation in determining public health outcomes is apparent. To measure a metro's partisanship, I simply take the Republican presidential candidate's 2-party vote share in the last presidential election to occur before an outbreak began. Presidential elections typically enjoy greater turnout than other elections, making them a reasonable heuristic for the general population's partisan leanings. Further, restricting to the two-party vote share reduces any artificial deflation caused by third-party candidates, though it does not remedy any cases in which the third parties draw asymmetrically from the two major parties.

[55] According to Wootton 2006, it was not until 1865 that medicine's effectiveness assumed an upward inflection with Lister's application of the germ theory of putrefaction to surgery; and that paradigm shift took several decades to fully penetrate the profession and make medicine a net positive for patients, extending lives instead of shortening them.

As the first column of Table 3.11 shows, when I include overall metro partisanship (cbsaRep2PartyVoteShare) in a regression with metro concentration, only the latter is statistically significant. Perhaps expressly *because* of the ambivalence of the pertinent forces, a metro's overall partisanship is associated with neither an increase nor a decrease in chemical outbreak length. In addition, metro partisanship's inclusion does not disrupt metro concentration's effect; the political entities that matter are the state borders.

	DV: logOnePlusExposureLengthDays	
	Rep Vote	Trump Vote
(Intercept)	1.728 ***	1.776 ***
	(0.000)	(0.000)
cbsaHerfindahlNorm	-0.944 **	-0.931 **
	(0.002)	(0.002)
cbsaRep2PartyVoteShare	0.004	
	(0.311)	
cbsa2PartyVoteShareTrump2020		0.003
		(0.427)
N	218	218
R2	0.046	0.044
logLik	-211.490	-211.692
deltaBIC	4.339	4.743

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.11: The partisanship of a metro's population—either in general or in 2020 specifically—does not explain away the metro concentration effect, and it does not have a statistically significant influence on outbreak length.

Of course, the salience of different issues varies over time; and public health in particular is typically not a major consideration in presidential contests. However, the coronavirus pandemic put public health front and center for the 2020 election. The population had suffered the costs of both public health action and public health inaction. On the action side, it had been hit with lockdowns, school closures, and ubiquitous masking; and, on the inaction side, it had endured substantial uncertainty and fear concerning the threat the virus actually posed, seen the healthcare infrastructure overwhelmed, and lost hundreds of thousands of lives.[56] It had also suffered severe economic strife and delayed medical treatment, but these could be attributed to both action and inaction.

From March through October of 2020, voters had ample time to learn how they felt about various public health measures; and the November election was largely an opportunity for them to say what the public health response to the pandemic ought to be going forward, with Republicans offering a less aggressive approach and Democrats a more aggressive one. To the extent that voters seized on this opportunity, the 2020 election results are a heuristic for how they felt about public health in general. Further, while areas certainly change over time, their political cultures are often sticky; so these "public health" election results are likely informative not just for 2020 but for the duration of the outbreaks dataset.

Proceeding on this assumption, in the second column of Table 3.11, I substitute a metro's 2020 Republican presidential two-party vote share for its most recent two-party vote share; but the results are essentially unchanged. Metro concentration's effect remains negative and statistically significant while the effect of Trump's 2020 two-party vote share is not statistically significant.

[56] See Our World in Data 2022.

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

3.2.4 Metro's County Average Partisanship

Aside from state partisan control and individual partisan inclinations, local government control could also play a significant role in determining the length of chemical outbreaks. In a number of states, public health is overseen by both state *and* local officials. For example, Princeton Municipality in New Jersey has its own health department; but its members collaborate with the state health department when necessary.[57]

For metro areas with segments that lie in states with this structure, the competence, resources, and aggressiveness of a constituent county's or municipality's public health department could easily influence the metro area's outbreak lengths. First, a local jurisdiction with a poorly equipped public health department likely takes longer to detect, diagnose the cause(s) of, and control outbreaks that start within its borders. Even if well-equipped nearby jurisdictions quickly identify those outbreaks and their likely source, these jurisdictions lack the authority to act on that information directly and must rely on the poorly equipped jurisdiction or wait for the relevant state(s) to step in. Second, for outbreaks that start beyond its borders but spill in, a poorly equipped jurisdiction likely collects less and lower-quality information more slowly from affected individuals; this reduces the information base available for solving those outbreaks, slows hypothesis development, and leads to bad hypotheses which waste investigative resources.

Given all this, if a local jurisdiction's partisan sentiment influences how well it equips its public health department, then that sentiment could also indirectly influence the jurisdiction's outbreak lengths. By comparison with a Democratic county council, a Republican county council might emphasize law enforcement and road maintenance over public health, and they might also wish to maintain smaller budgets in general to keep sales and property taxes low. If so, then their outbreaks might tend to be longer. Then, aggregating across the metro area, the more of the metro's counties are under Republican control, the longer would be its typical outbreak.

To evaluate this hypothesis, I build a heuristic measure of the proportion of a metro's counties under Republican control at the time an outbreak began. First, for each county in the outbreak's metro area, I generate an indicator for whether the Republican candidate received more votes than the Democratic candidate in the most recent presidential election to occur before an outbreak started. Second, I average those indicators to get the proportion of the metro's counties likely under Republican control.

This measure assumes that if a party secures a majority of the two-party vote share in the presidential election, it also controls the county; and, obviously, this might not always be the case. First, the local parties might diverge from the national parties, which would make split ticket voting more likely.

[57]See Municipality of Princeton n.d. and State of New Jersey 2022.

For example, New England Republicans are often more moderate than the national Republican Party, so a Mainer might vote for a Republican county commissioner and a Democratic presidential candidate. Second, the county could contain municipalities that control their own public health policy and whose voters' preferences diverge substantially from those of their county neighbors; after all, urban areas tend to be more Democratic while rural areas tend to be more Republican. In these cases, municipalities might have better public health infrastructure than the rest of the county. Third, the county's council districts could be gerrymandered to preserve control for a majority that no longer exists. Finally, the county's elections might not co-occur with presidential elections; and non-presidential electorates are often more Republican than their presidential counterparts. In such cases, the measure might underestimate a county's likelihood of Republican control.

Despite these limitations, this measure wields several important advantages. First, absent a national dataset of actual local party control, it provides a strong guess. Second, it affords uniformity by forcing all localities to choose between the *same* two candidates at the *same* time in any given period. Third, as I show in Appendix D, at least in the case of Oregon in 2020, the measure predicts control of county legislatures well; and, while data is more limited for elected county chief administrators, the measure also performs well in that limited dataset.

As Table 3.12 shows, metro concentration's effect survives when I include this measure of the proportion of a metro's county governments that are under Republican control. This holds whether I compute the measure using the most recent presidential election preceding an outbreak (left column) or the 2020 presidential election (right column), during which the population was better primed by the coronavirus pandemic to consider matters of public health. Of secondary interest, Republican control itself does not appear to have any influence over outbreak length, though the 2020 measure comes considerably closer to statistical significance than does its "most recent election" counterpart.

Of course, simple county control might not tell the whole story; the *strength* of partisan control within each county's government could also have a bearing on outbreak length. When the public health bureaucracy has fewer minority party officials around to challenge its handling of food safety, it has less informal pressure to moderate. In addition, counties that are more uniformly of one party could be more likely to have officials from the majority party's more extreme—or at least exotic—fringes; fully Republican areas could be more likely to have libertarians while fully Democratic areas could be more likely to have socialists; and their zeal could potentially exacerbate any effect of party on outbreak length.

	DV: logOnePlusExposureLengthDays	
	Prop Cty Rep Majority	Prop Cty Rep Majority (2020)
(Intercept)	-0.000	-0.000
	(1.000)	(1.000)
cbsaHerfindahlNorm	-0.197 **	-0.192 **
	(0.004)	(0.004)
cbsaPropCountiesWithRepMajority	0.068	
	(0.311)	
cbsaPropCountiesWithRepMajority2020		0.108
		(0.109)
N	218	218
R2	0.046	0.053
logLik	-303.730	-302.941
deltaBIC	4.343	2.767

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.12: The proportion of a metro area's county's with a Republican majority—both in the most recent presidential election before an outbreak and in the 2020 presidential election—does not influence outbreak length and does not disrupt metro concentration's effect.

For the similar reasons, the relative size of the county's compliant population could also influence outbreak length. I test this for the metro as a whole in Table 3.11), but the distribution of partisan populations within the metro could also have an influence. To take a simplified example, if a metro has a large, 100% Democratic urban county and seven small, 100% Republican rural counties, even though the metro as a whole might be overwhelmingly Democratic, its geographically large and homogeneous Republican pockets could still be given to substantially longer outbreaks than if that Republican population were more evenly distributed throughout the metro. If Democrats are indeed more likely to engage the public health bureaucracy more quickly, then having a more even distribution of them throughout the metro should lead to comparatively shorter outbreaks since the bureaucracy would be notified of an outbreak sooner.

To test all of these hypotheses, I create a more population-centric than elite-centric measure of a metro's average county partisanship by averaging Republican 2-party vote share across the metro's counties—again, using both an outbreak's most recent presidential election and the 2020 presidential election. As Table 3.13 shows, neither of these versions of county partisanship has a statistically significant influence on

	DV: logOnePlusExposureLengthDays	
	Rep Vote (Cty Avg)	Trump Vote (Cty Avg)
(Intercept)	-0.000	-0.000
	(1.000)	(1.000)
cbsaHerfindahlNorm	-0.199 **	-0.197 **
	(0.003)	(0.004)
cbsaCountyAvgRep2PartyVoteShare	0.094	
	(0.159)	
cbsaCountyAvg2PartyVoteShareTrump2020		0.090
		(0.177)
N	218	218
R2	0.050	0.049
logLik	-303.241	-303.326
deltaBIC	3.365	3.536

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 3.13: The county average Republican two-party vote share within a metro area—both in the most recent presidential election before an outbreak and in the 2020 presidential election—does not influence outbreak length; and metro concentration's effect survives the inclusion of either variable intact.

outbreak length. However, as has been the case time and again, metro concentration's effect survives intact and unspoiled when I account for these potentially omitted variables.

4 Other Etiologies

While metro concentration has a robust shortening effect on chemical outbreaks, the same cannot be said of outbreaks with other etiologies. As Table 4.1 and Table 4.2 show, among each of the other major CDC-designated etiological categories, metro concentration has no statistically significant effect on outbreak length; indeed, for all but one of these, the coefficient's p value exceeds 0.70. This difference in effect is likely a result of differences in each category's average outbreak length.

As I show in Table B.1, chemical outbreaks easily have the shortest exposure period (the number of days from first exposure to last exposure, inclusive) among the major etiological categories, with all intoxication concluded in just over 3 days. In contrast, average exposure periods for the other categories are a little over 4 days at the very shortest. While that one extra day might not seem like much, at this scale, *hours* are enough to count.

For the public health apparatus, the dichotomy is comparable to that between driving on a typical Interstate highway and driving in a Formula One race. In the former, most drivers are trying to get to their destination reasonably quickly; but being behind or ahead by a few minutes rarely makes a meaningful difference, so driving's intricate technicalities usually aren't a conscious consideration. In contrast, in Formula One, a few minutes can mean the difference between finishing first or last; a minor inefficiency in any technical area from part selection and engine tuning to a driver's gear shifting and braking can mean utter defeat, so all of these are *heavily* optimized.

In the same way, against etiologies besides chemical toxins, metros whose public health bureaucracies are fractured and therefore slowed by state lines are perfectly up to the task. To be effective, they must still respond with speed; but they won't seriously suffer from the delays that state borders engender. In contrast, against chemical toxins, such metros suffer measurably from their border frictions. Where time is of the essence, centralization can make its biggest difference.

	DV: logOnePlusExposureLengthDays	
	Viruses	Bacteria
(Intercept)	0.91222 ***	1.30693 ***
	(0.00000)	(0.00000)
cbsaHerfindahlNorm	0.07794	0.04272
	(0.16985)	(0.72172)
N	4501	3143
R2	0.00042	0.00004
logLik	-3922.92905	-4657.80881
BIC	7871.09427	9339.77642

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 4.1: Metro concentration does not have a statistically significant effect on the length of viral or bacterial outbreaks.

	DV: logOnePlusExposureLengthDays	
	Parasites	Hepatitis
(Intercept)	1.81632 **	1.92677
	(0.00593)	(0.23221)
cbsaHerfindahlNorm	0.17672	0.56985
	(0.80287)	(0.73098)
N	117	38
R2	0.00054	0.00332
logLik	-184.07791	-58.29982
BIC	382.44235	127.51240

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table 4.2: Metro concentration does not have a statistically significant influence on length of parasite or hepatitis outbreaks

5 Conclusion

As evidence from outbreaks of chemically-induced foodborne illness shows, bureaucratic centralization can provide a substantial advantage for policy areas wherein a rapid response is critical. These outbreaks' characteristically brief duration provides a very narrow window for effective intervention; and, by comparison with their more fractured siblings, metro areas whose public health bureaucracies are better-consolidated within fewer states enjoy a clear advantage in combating chemical outbreaks. Their more centralized command structure enables investigation and control measures to be carried out with greater dispatch since they are less delayed by the need to develop consensus among the responsible agencies or to go through the extra layer of coordination that is the CDC. Further, their more consolidated information-sharing systems allow for the quicker identification of ongoing outbreaks, the responsible chemical agent, the food(s) carrying that agent, and the outbreak's reach—all of which expedite outbreak control.

I also find that SEDRIC, an information-sharing tool which was apparently created to overcome decentralization's limitations, was not effective in shortening chemical outbreaks. Centralized information may not be able to deliver the benefits of centralization unless it is paired with centralized control. To be sure, the system may simply not yet have seen sufficient uptake in the nation's public health departments to deliver measurable benefits; but, to the extent that this is the case, it indicates that centralized information may itself not be possible without centralized control.

These findings endorse policy decisions like the U.S. Constitution's endowing the presidency with central *control* of the nation's armed forces. As with chemical outbreaks, in war, expeditious information transfer and decision-making are critical; and centralized control delivers both. However, qualifying Publius's arguments in *Federalist* No. 70 for a singular, unified, and energetic executive, a rapid response is not mission-critical in all policy areas. On questions of centralization and decentralization, rapidity's criticality must be a salient consideration.

References

Address Confidentiality Program (2022).

Aguilera, Elizabeth (Apr. 15, 2022). "California Backs Away from COVID Vaccine Mandates for Kids". In: *CalMatters. Coronavirus.* URL: http://calmatters.org/health/coronavirus/2022/04/kids-covid-vaccine-mandate-california/ (visited on 10/24/2022).

Alaska Independence Party (Oct. 17, 2020). *Alaskan Independence Party Platform.* URL: https://www.akip.org/platform.html (visited on 06/15/2022).

Alesina, Alberto and Enrico Spolaore (2003). *The Size of Nations.* Cambridge, Mass: MIT Press. 261 pp. ISBN: 978-0-262-01204-1.

Applequist, Scott (Feb. 10, 2022). *Normals Lat/Lon Projection.* Letter.

Association of Oregon Counties (Nov. 28, 2022). *Oregon County Structures Chart.* URL: https://drive.google.com/file/d/1rVk4kPNEO6sjPudc1wHoNHcU1ZHFcHvX/view?usp=embed_facebook (visited on 02/09/2023).

Association of Public Health Laboratories (Nov. 18, 2013). *InFORM 2013 Conference.* APHL. URL: http://www.aphl.org (visited on 04/03/2022).

— (Nov. 19, 2014a). *Joint South Central/Southeast Regional PulseNet & OutbreakNet Meeting Registration Form.*

— (Dec. 5, 2014b). *PulseNet & OutbreakNet Regional Meeting: Joint Mountain/West Registration Form.*

— (Oct. 14, 2016). *PulseNet/OutbreakNet Regional Meeting: Southeast Region Meeting Agenda (Draft).* Association of Public Health Laboratories. URL: https://www.aphl.org/conferences/annualmeeting/Documents/Southeast-Draft-Agenda-10-14.pdf#search=ALL%28%22SEDRIC%22%29 (visited on 04/03/2022).

Avila, Yuriria et al. (Feb. 25, 2022). "See Mask Mandates and Guidance in Each State". In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2021/us/cdc-mask-guidance-states.html (visited on 10/24/2022).

Cancellation of Registration (2022).

Centers for Disease Control and Prevention (Mar. 15, 2016a). *PulseNet: 20 Years of Making Food Safer to Eat.* URL: https://www.cdc.gov/pulsenet/pdf/pulsenet-20-years_4_pg_final_508.pdf (visited on 04/21/2022).

— (Dec. 19, 2016b). *SEDRIC: System for Enteric Disease Response, Investigation, and Coordination.* The Wayback Machine. URL: https://web.archive.org/web/20161219231937/https://www.cdc.gov/foodsafety/outbreaks/pdfs/sedric-factsheet.pdf (visited on 04/03/2022).

Centers for Disease Control and Prevention (Sept. 28, 2017). *National Outbreak Reporting System (NORS) Guidance*. URL: https://www.cdc.gov/nors/downloads/guidance.pdf (visited on 04/05/2022).

— (May 2, 2019). *Pulsed-Field Gel Electrophoresis (PFGE) | PulseNet Methods| PulseNet | CDC*. URL: https://www.cdc.gov/pulsenet/pathogens/pfge.html (visited on 04/20/2022).

— (Feb. 18, 2020a). *Second Travel-related Case of 2019 Novel Coronavirus Detected in United States*. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/media/releases/2020/p0124-second-travel-coronavirus.html (visited on 03/03/2023).

— (May 4, 2020b). *SEDRIC: System for Enteric Disease Response, Investigation, and Coordination*. URL: https://www.cdc.gov/foodsafety/outbreaks/pdfs/sedric-factsheet.pdf (visited on 04/03/2022).

— (Jan. 22, 2020c). *SEDRIC: System for Enteric Diseases, Response, Investigation, and Coordination*. URL: https://www.cdc.gov/foodsafety/outbreaks/ppt/SEDRIC-Overview.pptx.

— (Dec. 29, 2021). *National Outbreak Reporting System*. Atlanta, GA: US Department of Health and Human Services, CDC.

Conditions Where Disclosure of Elector's Residence Address Required (2022).

Congressional Quarterly (June 8, 2022). *CQ Voting and Elections Collection*. URL: https://library.cqpress.com/elections/download-data.php (visited on 06/09/2022).

Crombie, Noelle (June 1, 2021). "Clackamas County Commissioner Mark Shull Compares Vaccine Passports to Jim Crow Laws". In: *The Oregonian/Oregonian Live. Clackamas County*. URL: https://www.oregonlive.com/clackamascounty/2021/06/clackamas-county-commissioner-mark-shull-compares-vaccine-passports-to-jim-crow-laws.html (visited on 02/10/2023).

DeMillo, Andrew and Geoff Mulvihill (Nov. 4, 2021). *GOP State Officials Push Back on Employer Vaccine Mandate*. AP NEWS. URL: https://apnews.com/article/covid-states-suing-employer-vaccine-mandates-f4300ed015ed716a1054949cf199a25a (visited on 10/24/2022).

Democracy for America (n.d.). *Democracy for America : Heather Buch - Lane County Commissioner of Oregon*. URL: https://www.democracyforamerica.com/site/our_candidates/644-heather-buch-for-lane-county-commissioner-oregon (visited on 12/04/2022).

Democratic Party of Lane County (n.d.). *DPLC Day of Action for Heather Buch*. Democratic Party of Lane County. URL: https://www.dplc.org/calendar/2022/4/25/dplc-day-of-action-for-heather-buch (visited on 02/10/2023).

Enamorado, Ted, Benjamin Fifield, and Kosuke Imai (2019). "Using a Probabilistic Model to Assist Merging of Large-scale Administrative Records". In: *American Political Science Review* 113.2, pp. 353–371. DOI: 10.1017/S0003055418000783.

Fellegi, Ivan P. and Alan B. Sunter (1969). "A Theory for Record Linkage". In: *Journal of the American Statistical Association* 64.328, pp. 1183–1210. ISSN: 0162-1459. DOI: 10.2307/2286061. JSTOR: 2286061. URL: http://www.jstor.org/stable/2286061 (visited on 12/21/2022).

Fisher & Phillips LLP (May 18, 2021). *Georgia Extends Limited COVID-19 Liability Protections*. Fisher & Phillips LLP. URL: https://www.fisherphillips.com/news-insights/georgia-extends-limited-covid-19-liability-protections.html (visited on 10/24/2022).

Food Safety Modernization Act Surveillance Working Group, Board of Scientific Counselors, Office of Infectious Diseases (Apr. 7, 2015). *2014 Annual Report to the Secretary, Department of Health and Human Services*. URL: https://www.cdc.gov/ddid/docs/BSC_OID_FSMA_Surv_WG_2014_Annual_Report.pdf (visited on 04/03/2022).

Gibson, Mark (Nov. 11, 2021). "County Seeks Engagement on MCMC/Kramer Field Proposal". In: *Columbia Gorge News*. URL: https://www.columbiagorgenews.com/county-seeks-engagement-on-mcmc-kramer-field-proposal/article_831341fc-4345-11ec-939b-6be1f873d4dd.html (visited on 02/10/2023).

Gorge Country Media (May 5, 2021). *Wasco County COVID-19 Update 05 05 2021*. Gorge Country Media. URL: https://gorgenewscenter.com/2021/05/05/wasco-county-covid-19-update-05-05-2021/ (visited on 02/10/2023).

Hill, Evan et al. (June 1, 2020). "How George Floyd Was Killed in Police Custody". In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/2020/05/31/us/george-floyd-investigation.html (visited on 02/27/2023).

Inter-university Consortium for Political and Social Research (n.d.). *United States Historical Election Returns Series*. Inter-university Consortium for Political and Social Research. URL: https://www.icpsr.umich.edu/web/ICPSR/series/00059 (visited on 12/22/2022).

James, Gareth et al. (2013). *An Introduction to Statistical Learning*. Vol. 103. Springer Texts in Statistics. New York, NY: Springer New York. ISBN: 978-1-4614-7138-7. DOI: 10.1007/978-1-4614-7138-7. URL: http://link.springer.com/10.1007/978-1-4614-7138-7 (visited on 07/23/2021).

Jaquiss, Nigel (Jan. 7, 2019). "Clackamas County, Where Metro Used to Be a Dirty Word, Considers Hiring Top Metro Manager". In: *Willamette Week*. URL: https://www.wweek.com/news/2019/01/07/clackamas-county-where-metro-used-to-be-a-dirty-word-considers-hiring-top-metro-manager/ (visited on 02/10/2023).

— (Nov. 24, 2022). "Clackamas County Republicans Move Right". In: *Willamette Week*. URL: https://www.wweek.com/news/2022/11/24/clackamas-county-republicans-move-right/ (visited on 02/10/2023).

Kaplan, Jacob (July 1, 2020). *United States Governors 1775-2020*. Inter-university Consortium for Political and Social Research (ICPSR). DOI: 10.3886/E102000V2. URL: https://www.openicpsr.org/openicpsr/project/102000/version/V2/view (visited on 06/10/2022).

Klokan Technologies (2022). *WGS 84 / Pseudo-Mercator - Spherical Mercator, Google Maps, OpenStreetMap, Bing, ArcGIS, ESRI - EPSG:3857*. URL: http://epsg.io (visited on 09/19/2022).

L2, Inc. (Jan. 9, 2022). *L2 DataMapping*. Bothell, Washington, USA: L2, Inc. URL: https://www.l2datamapping.com/ (visited on 02/11/2023).

Landry, Pierre F. (2008). *Decentralized Authoritarianism in China: The Communist Party's Control of Local Elites in the Post-Mao Era*. 1st ed. Cambridge University Press. ISBN: 978-0-521-88235-4. DOI: 10.1017/CBO9780511510243. URL: https://www.cambridge.org/core/product/identifier/9780511510243/type/book (visited on 12/24/2023).

Manella, Gal et al. (Aug. 16, 2022). "The Human Blood Transcriptome Exhibits Time-of-Day-Dependent Response to Hypoxia: Lessons from the Highest City in the World". In: *Cell Reports* 40.7. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2022.111213. pmid: 35977481. URL: https://www.cell.com/cell-reports/abstract/S2211-1247(22)01030-0 (visited on 09/26/2022).

Mesh, Aaron (Oct. 30, 2012). "The King of Clackistan". In: *Willamette Week*. URL: https://www.wweek.com/portland/article-19841-the-king-of-clackistan.html (visited on 02/10/2023).

Minnesota Department of Health (Jan. 31, 2023). *Reporting Suspected Foodborne and Waterborne Illness: Foodborne and Waterborne Illness Complaints - MN Dept. of Health*. URL: https://www.health.state.mn.us/diseases/foodborne/reporting.html (visited on 03/03/2023).

MIT Election Data Science Lab (2018). *County Presidential Election Returns 2000-2020*. Version V9. DOI: 10.7910/DVN/VOQCHQ. URL: https://doi.org/10.7910/DVN/VOQCHQ.

Mordecai, Erin A. et al. (2019). "Thermal Biology of Mosquito-Borne Disease". In: *Ecology Letters* 22.10, pp. 1690–1708. ISSN: 1461-0248. DOI: 10.1111/ele.13335. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13335 (visited on 09/18/2022).

Morelli, Jeff (Aug. 16, 2018). *Personal Correspondence*. Letter.

Municipality of Princeton (n.d.). *Health Department | Princeton, NJ*. URL: https://www.princetonnj.gov/270/Health-Department (visited on 10/29/2022).

National Association of County and City Health Officials (2022). *Directory of Local Health Departments*. URL: https://www.naccho.org/membership/lhd-directory?searchType=standard&lhd-search=health&lhd-state=#card-filter (visited on 10/17/2022).

National Conference of State Legislatures (Sept. 28, 2022). *Access To and Use Of Voter Registration Lists*. URL: https://www.ncsl.org/elections-and-campaigns/access-to-and-use-of-voter-registration-lists (visited on 02/11/2023).

National Governors Association (2022). *Former Governors*. National Governors Association. URL: https://www.nga.org/former-governors/ (visited on 10/20/2022).

National Organization for Rare Disorders (NORD) (1986). *Ciguatera Fish Poisoning*. In: *Rare Disease Database*. URL: https://rarediseases.org/rare-diseases/ciguatera-fish-poisoning/ (visited on 03/15/2022).

Navarro, Danielle (July 14, 2022). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners*. Statistics LibreTexts. URL: https://stats.libretexts.org/Bookshelves/Applied_Statistics/Book%3A_Learning_Statistics_with_R_-_A_tutorial_for_Psychology_Students_and_other_Beginners_(Navarro).

Neath, Andrew A. and Joseph E. Cavanaugh (2012). "The Bayesian Information Criterion: Background, Derivation, and Applications". In: *WIREs Computational Statistics* 4.2, pp. 199–203. ISSN: 1939-0068. DOI: 10.1002/wics.199. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/wics.199 (visited on 09/29/2022).

NOAA (Apr. 26, 2021a). *1991-2020 U.S. Climate Normals Annual Variable Documentation*. URL: https://www.ncei.noaa.gov/data/normals-annualseasonal/1991-2020/doc/Normals_ANN_Documentation_1991-2020.pdf (visited on 02/08/2022).

— (Apr. 26, 2021b). *Normals Calculation Methodology 2020*. URL: https://www.ncei.noaa.gov/data/normals-annualseasonal/1991-2020/doc/Normals_Calculation_Methodology_2020.pdf (visited on 02/08/2022).

— (n.d.). *NADCON - North American Datum Conversion Utility - Tools - National Geodetic Survey*. URL: https://www.ngs.noaa.gov/TOOLS/Nadcon/Nadcon.shtml (visited on 09/19/2022).

Northup, Nancy (1987). "Local Nonpartisan Elections, Political Parties and the First Amendment". In: *Columbia Law Review* 87.8, pp. 1677–1701. ISSN: 0010-1958. DOI: 10.2307/1122744. JSTOR: 1122744. URL: http://www.jstor.org/stable/1122744 (visited on 12/21/2022).

Notice of Deaths to Secretary of State and County Clerk; Effect of Notice (2022).

Oates, Wallace E. (1999). "An Essay on Fiscal Federalism". In: *Journal of Economic Literature* 37.3, pp. 1120–1149. ISSN: 0022-0515. JSTOR: 2564874. URL: https://www.jstor.org/stable/2564874 (visited on 12/27/2023).

Our World in Data (Oct. 9, 2022). *Excess Mortality: Cumulative Number of Deaths from All Causes Compared to Projection Based on Previous Years*. Our World in Data. URL: https://ourworldindata.org/grapher/cumulative-excess-deaths-covid (visited on 10/28/2022).

Poole, Keith T. and Howard Rosenthal (2007). *Ideology & Congress*. 2nd rev. ed. New Brunswick: Transaction Publishers. 344 pp. ISBN: 978-1-4128-0608-4.

Reamer, Terry (Nov. 6, 2016). *Joint Mountain/West Regional PulseNet & OutbreakNet Meeting Registration Form*.

Residence Address Disclosure Exemptions (2022). URL: https://secure.sos.state.or.us/oard/view.action?ruleNumber=165-005-0130.

Rogoway, Mike (Sept. 1, 2022). "Google Set to Win New Tax Breaks in the Dalles, but Uncertainty Lingers over an Earlier Deal". In: *The Oregonian/The Oregonian Live*. URL: https://www.oregonlive.com/silicon-forest/2021/10/google-set-to-win-new-tax-breaks-in-the-dalles-but-uncertainty-lingers-over-an-earlier-deal.html (visited on 02/10/2023).

Samuels, Brett and Rebecca Klar (Mar. 23, 2020). "Trump: 'We Can't Let the Cure Be Worse than the Problem Itself'". In: *The Hill*. URL: https://thehill.com/homenews/administration/488965-trump-hints-at-changes-to-restrictions-we-cant-let-the-cure-be-worse/ (visited on 10/24/2022).

Seymour, Benjamin, Athena Andreosso, and Jamie Seymour (2015). "Cardiovascular Toxicity from Marine Envenomation". In: *Heart and Toxins*. Elsevier, pp. 203–223. ISBN: 978-0-12-416595-3. DOI: 10.1016/B978-0-12-416595-3.00007-4. URL: https://linkinghub.elsevier.com/retrieve/pii/B9780124165953000074 (visited on 02/06/2023).

South Carolina (2012). URL: https://www.scstatehouse.gov/code/t44c001.php (visited on 10/20/2022).

State of New Jersey (Oct. 29, 2022). *Department of Health for the State of New Jersey | Homepage*. URL: https://www.nj.gov/health/ (visited on 10/29/2022).

State of Oregon (Jan. 8, 2023). *Oregon Counties*. URL: https://www.oregon.gov/pages/counties.aspx (visited on 02/09/2023).

State of Oregon Secretary of State (June 3, 2021). *Blue Book - County Government*. URL: https://web.archive.org/web/20210603103913/https://sos.oregon.gov/blue-book/Pages/local-county.aspx (visited on 02/09/2023).

Stratta, Piero and Guido Badino (Apr. 3, 2012). "Scombroid Poisoning". In: *CMAJ : Canadian Medical Association Journal* 184.6, p. 674. ISSN: 0820-3946. DOI: 10.1503/cmaj.111031. pmid: 22231690. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3314039/ (visited on 03/15/2022).

Stringer, Grant (May 14, 2022). "Conservative Control of Clackamas County Commission Hangs in Balance with Fired-up, Big Money Races for 2 Seats". In: *The Oregonian/Oregon Live. Politics*. URL: https:

//www.oregonlive.com/politics/2022/05/conservative-control-of-clackamas-county-commission-hangs-in-balance-with-fired-up-big-money-races-for-2-seats.html (visited on 02/10/2023).

The Internet Archive (2023a). *Wayback Machine*. URL: https://archive.org/web/ (visited on 02/11/2023).

— (Jan. 8, 2023b). *Wayback Machine*. URL: https://web.archive.org/ (visited on 02/09/2023).

The New York Times (Apr. 1, 2020a). "Florida Coronavirus Map and Case Count". In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2021/us/florida-covid-cases.html (visited on 03/03/2023).

— (Apr. 1, 2020b). "Minnesota Coronavirus Map and Case Count". In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2021/us/minnesota-covid-cases.html (visited on 02/27/2023).

— (Apr. 1, 2020c). "Puerto Rico Coronavirus Map and Case Count". In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2021/us/puerto-rico-covid-cases.html (visited on 02/27/2023).

Tibshirani, Ryan J. (Jan. 1, 2013). "The Lasso Problem and Uniqueness". In: *Electronic Journal of Statistics* 7 (none). ISSN: 1935-7524. DOI: 10.1214/13-EJS815. URL: https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-7/issue-none/The-lasso-problem-and-uniqueness/10.1214/13-EJS815.full (visited on 04/29/2022).

Tootie Smith (n.d.). In: *Ballotpedia*. URL: https://ballotpedia.org/Tootie_Smith (visited on 12/04/2022).

U.S. Census Bureau (Oct. 2009). *TIGER/Line Shapefiles 2009: Technical Documentation*.

— (Mar. 2011). *Centers of Population Computation for the United States, 1950-2010*. URL: https://www2.census.gov/programs-surveys/decennial/2010/program-management/4-release/press-kit/center-of-population/centers-of-population-computation-documentation.pdf (visited on 09/19/2022).

U.S. Centers for Disease Control and Prevention (July 20, 2015). *InFORM 2015 Conference Agenda Draft 7.20.15*. URL: https://www.aphl.org/conferences/Documents/InFORM%202015%20Conference%20Agenda%20DRAFT%207.20.15.pdf#search=ALL%28%22SEDRIC%22%29 (visited on 04/03/2022).

— (Feb. 13, 2019). *BioNumerics WGS Workshop for PulseNet Participants Draft Agenda*. URL: https://www.aphl.org/Materials/May19_BN7_Agenda_draft.pdf#search=ALL%28%22SEDRIC%22%29 (visited on 04/03/2022).

U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition (Oct. 30, 2017). *Scombrotoxin Poisoning and Decomposition*. U.S. Food and Drug Administration. URL: https://www.fda.

gov/food/seafood-guidance-documents-regulatory-information/scombrotoxin-poisoning-and-decomposition (visited on 03/15/2022).

US Census Bureau (2022). *Gazetteer File Record Layouts*. Census.gov. URL: https://www.census.gov/programs-surveys/geography/technical-documentation/records-layout/gaz-record-layouts.html (visited on 02/03/2022).

Van Sickle, Jan (n.d.). *Modern Geocentric Datum*. GEOG 862: GPS and GNSS for Geospatial Professionals. URL: https://www.e-education.psu.edu/geog862/node/1801 (visited on 09/23/2022).

Virués-Ortega, Javier et al. (Dec. 2004). "Neuropsychological Functioning Associated with High-Altitude Exposure". In: *Neuropsychology Review* 14.4, pp. 197–224. ISSN: 1040-7308. DOI: 10.1007/s11065-004-8159-4. pmid: 15796116.

Williams, Ian and Thai-An Nguyen (Sept. 20, 2011). "Foodborne Outbreak Response Using the Palantir Enterprise Platform". 7th Annual OutbreakNet Conference (Long Beach, CA).

"Wisconsin Coronavirus Map and Case Count" (Apr. 1, 2020). In: *The New York Times. U.S.* ISSN: 0362-4331. URL: https://www.nytimes.com/interactive/2021/us/wisconsin-covid-cases.html (visited on 02/27/2023).

Wootton, David (2006). *Bad Medicine: Doctors Doing Harm since Hippocrates*. Oxford, United Kingdom: Oxford University Press, Incorporated. ISBN: 978-0-19-151672-6. URL: http://ebookcentral.proquest.com/lib/princeton/detail.action?docID=431394 (visited on 10/28/2022).

Appendices

Appendix A Variable Definitions and Generation

A.1 Core-Based Statistical Areas (CBSAs)

Core-Based Statistical Areas (CBSAs) are approximate delineations of metropolitan and micropolitan areas; and they are defined by the U.S. Office of Management and Budget for the U.S. states, the District of Columbia, and Puerto Rico. Historically, the larger units among these have been known variously as "standard metropolitan areas" (SMAs), "standard metropolitan statistical areas" (SMSAs), and "metropolitan statistical areas" (MSAs). Each CBSA is composed of one or more "core counties" and may also include any number of "outlying counties," merging a population core with the outlying areas which have a substantial commuting relationship with that core. More technically, each core county must have either 1) 50+% of its population living in Census-designated urban areas each of which has a population of at least 10,000 or 2) 5,000+ residents out of a single urban area that has a population of at least 10,000.[58] Each associated outlying county must have either 1) 25+% of the workers *living* within its borders commuting for employment to the CBSA's central county or counties or 2) 25+% of those *working* within its borders commuting from the CBSA's central county or counties.

A.2 Distances between Metro Population Centroids & Weather Stations

To calculate the distance between a given metro population centroid and a given weather station, I assume that the coordinates for both are referenced to the NAD83 geodetic datum, and I measure the distance between the two along the NAD83 ellipsoid. This method may yield errors as large as 25 miles because of variation in the datums used to report the geographic coordinates. The county population centroids are indeed referenced to NAD83 and, as a consequence, are not a source of error.[59] However, weather station coordinates are referenced not just to NAD83 (47% of stations) but also to WGS84 (8%), NAD27 (1%), and OLD HAWAIIAN (<1%); the datums for the remaining 44% are unknown but are likely Spherical Mercator.[60]

[58] *Federal Register*, 2010 Standards, p. 37,250. Note that urban areas may span county lines.

[59] The Census computed the component county population centroids using its TIGER database, which records locations in NAD83 U.S. Census Bureau 2011; U.S. Census Bureau 2009.

[60] Most of the stations with unknown projections are volunteer-operated CoCoRaHS stations; and their volunteer operators are encouraged to find their coordinates using Google Maps, which uses Spherical Mercator (Applequist 2022; Klokan Technologies 2022).

Calculations errors for distances between coordinates referenced to NAD83 and any of the listed datums *other* than Spherical Mercator likely do not exceed 1 km. NAD27 and OLD HAWAIIAN differ from NAD83 by a maximum of a few hundred meters;[61] and, depending on their respective versions, WGS84 differs from NAD83 by a meter or two.[62]

Unfortunately, Spherical Mercator may differ from WGS84 (and, by extension, from NAD83) by as much as 25 miles. Unlike the other datums, it assumes the Earth is spherical rather than ellipsoidal; and, while that simplifies computations, it can introduce nontrivial error. Fortunately, according to NOAA scientists, this magnitude of noise is acceptable for climatological purposes.[63] As a rule, areas within 25 miles of one another share the same climate; so I can reasonably ignore any errors introduced by differences in reference datum.

[61] NOAA n.d.
[62] Van Sickle n.d.
[63] See Applequist 2022.

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

Appendix B Why Chemical Toxicity?

There are several interrelated reasons that we should expect control of chemical outbreaks to leverage and demonstrate the value of a consolidated bureaucracy. The first several all center on the comparatively greater speed associated with chemical outbreaks while the last is the simple inability to sequence chemicals.

General Etiology	Exposure Period (days)	Incubation Shortest (days)	Incubation Median (days)	Incubation Longest (days)
Chemicals	3.32	0.09	0.25	0.59
Viruses	4.17	0.66	1.36	2.39
Bacteria	9.12	0.66	1.49	3.49
Parasites	12.9	3.33	8.05	14.5
Hepatitis	23.3	21.5	28.2	37.2

Table B.1: Average Temporal Metrics by General Etiology in Days. Chemical outbreaks have a shorter exposure period and substantially shorter incubation periods than other types of outbreaks.

For starters, as we see in Table B.1, a chemical outbreak's typical exposure period—the time from first exposure to last exposure—is shorter than that for the other etiologies. All of a toxin's work exposing people is done in about 3 days while the others etiologies take 4 days—or substantially more. Naturally, these lengths are conditioned by public health authorities' attempts to curtail them; but that simple fact does not render them uninformative for our purposes. Ultimately, both more and less concentrated metro areas are working on that shorter time scale; and its shortness minimizes the role that slower-moving interstate infrastructure can play.

Take the CDC's National Outbreak Reporting System (NORS) for example. After an outbreak is discovered and initially investigated, it is often reported in NORS. That report allows every state and locality in the nation to be on the lookout for further expansion of that outbreak—and more expeditiously to control it. But performing the initial investigation and reporting its results both take time, and getting up-to-date on the most recent NORS reports takes time; so, with short exposure periods like those typical for chemical outbreaks, NORS is unlikely to play a preventive role. In contrast, when the exposure period typically lasts for more than a week—as with bacteria, parasites, and hepatitis—there is ample time for NORS to assist directly in an outbreak's control.

Pushing in the opposite direction, a factor that enhances a chemical outbreak's controllability is its incredibly short incubation period—certainly in relation to its exposure period. A typical case in a chemical outbreak presents symptoms in a mere 6 hours after exposure; and, even for the cases with the longest incubation periods, symptoms present within 15 hours. Such short incubation periods enable cases better to recall details of where and what they ate, which helps public health officials more quickly and effectively

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260

identify the cause and control the outbreak. Further, assuming an outbreak and its first case are typical, the local public health department could be made aware of the outbreak within 6 hours of its beginning and still have 3 full days to chase down would-be cases. Of course, without the swift response of which consolidated bureaucracies are more capable, those 3 days may well be wasted in multi-state metros.

In contrast, the other etiologies tend to have longer incubation periods, giving people more time to forget what and where they ate—and giving the health department a larger set of potential foods to blame. With noisier information and more suspects, health departments' investigations are more difficult, and control is delayed. With a median incubation period around 1.5 days, viruses and bacteria don't cause too much strain; but, with median incubation periods around 1 and 4 weeks, respectively, parasites and hepatitis present a true nightmare for investigation and control. Unless cases happen to keep a food journal, health departments may not even be be able to identify the cause of these outbreaks—much less control them.

Beyond their longer raw median incubation periods, parasites and hepatitis have substantial median incubation periods as a proportion of their exposure periods. A parasitic outbreak's typical incubation period covers more than half of its exposure period—meaning that, by comparison with chemical outbreaks, much more of the potential damage of exposure is often already done by the time the health department has the chance to learn of a parasitic outbreak.

As for hepatitis, its typical incubation period actually *exceeds* its exposure period by about 5 days. Even on the leading edge, its average shortest incubation period spans a whopping 90% of its typical exposure period. As a result, unlike with all the other etiologies, a health department may not have the opportunity to learn of a hepatitis outbreak until all the would-be exposures have already occurred—rendering any control of the outbreak impossible.

The last notable distinct feature of chemical outbreaks that enhances the value of having a consolidated bureaucracy to respond to them is the simple fact that the responsible chemicals lack DNA. All the other etiologies—parasites, bacteria, and viruses (including hepatitis)—have a genetic sequence that defines them and which can be used to cluster related cases. As these pathogens replicate, they develop random mutations that they pass on to their progeny. So, for example, after a pathogen with unique mutations A & B starts an outbreak, all cases in that outbreak will be infected with pathogens that also have those same mutations. If these cases each have a lab test done to determine the cause of their symptoms, upon confirmation of general etiology, their lab samples can be processed in greater detail to extract DNA fingerprints[64]; and those fingerprints can be submitted to public health authorities for comparison.

Now, before fingerprint submission, these cases might appear to the authorities to be unrelated—especially if those cases report the illness in different jurisdictions. Authorities have a harder time identifying an outbreak and isolating its cause both when they have fewer cases in their own jurisdiction to compare to one another and when they have relatively little information on those cases.

On the first point, outbreak investigations are largely an exercise in finding common features, and small numbers make that exercise more difficult. If a local public health office has two reports of something that sounds like norovirus, it lacks immediate cause to infer an outbreak; norovirus is a common foodborne illness, so the two cases could be entirely unrelated. Further, assuming the office has the time to interview these cases (and they commonly do not), the responses could omit the causal common factor thanks to memory lapses or include a wealth of common factors unrelated to the outbreak—both of which could mislead investigators.

On the second point, early on in an investigation, authorities may know only that a few vague cases of foodborne illness have presented. Unless and until they secure more information on those cases—such as symptoms, time of onset, and dietary history—they might just as reasonably pronounce those cases isolated events as associate them with one another.

However, after the DNA fingerprints come in, public health authorities can cluster cases both within and across jurisdictions fairly quickly. In my example, they might find that all of the recent cases' clinical samples are positive for the A & B mutations, which strongly suggests that the cases have a common cause. As a result, the affected jurisdictions can pool their resources more efficiently to identify the food that's responsible for the outbreak and more quickly to prevent new cases from arising by removing that food from further distribution and consumption.

Stepping out of the realm of what's possible and into the realm of what's actually done, these DNA clustering tools are not systematically used in the US on a national scale for any etiology except bacteria. For the entire study period, the CDC consolidated and analyzed bacterial DNA fingerprints through PulseNet;[65] so the negative influence of jurisdictional barriers on bacterial outbreak investigations was likely substantially attenuated. For the remaining non-chemical etiologies, programs that were similar to PulseNet but either

[64]I cannot call these fingerprints "sequences" because, in many cases, that's not what they are. Whole genome sequencing (WGS) is certainly the most complete DNA fingerprinting method; but the more common method used in the U.S. public health community over the course of the study period is pulsed-field gel electrophoresis (PFGE), which does not generate complete sequences. See Centers for Disease Control and Prevention (2019).

[65]Centers for Disease Control and Prevention 2016a.

more localized or transient may have played a similar attenuating role; but there is no evidence for the widespread implementation of any such program during the study period.

Appendix C Sample

C.1 Sample Description

In this section, I address the chemical outbreak sample's geographic and temporal distribution. Starting with distribution by metro concentration, as Figure C.1, while most outbreaks (81%) happen in single-state metros, 50 outbreaks occur in multi-state metros. Further, the frequency of outbreaks is fairly consistent across the range of metro concentration ratings (cbsaHerfindahlNorm). These factors afford a decent amount of variation and reasonable assurance that the findings are not driven by any single concentration rating or individual metro.

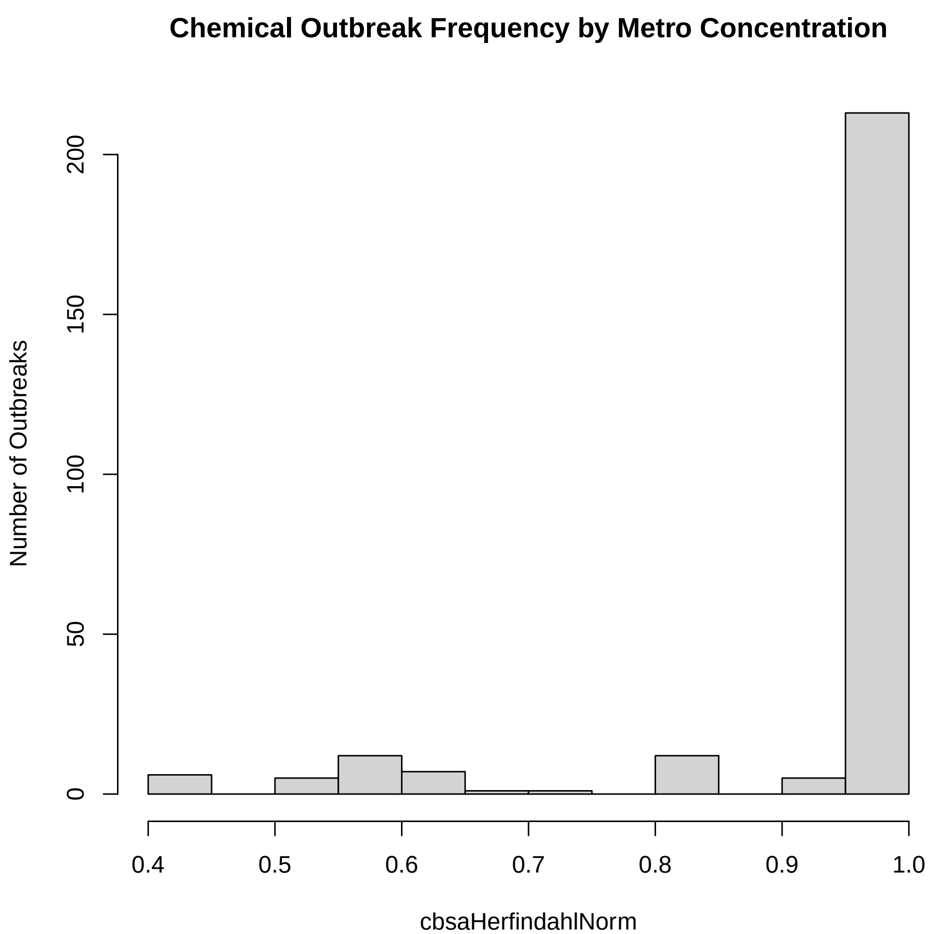

Figure C.1: Chemical Outbreaks by Metro Concentration.

As Figure C.2 demonstrates, average outbreak length is fairly consistent over the sample period, suggesting that time is not an important omitted variable. Admittedly, the ends of the LOESS curve do suggest an upward trend; but the trend is not statistically significant when included in a regression alongside

https://doi.org/10.33774/apsa-2024-82cvk ORCID: https://orcid.org/0000-0001-6810-1260 Content not peer-reviewed by APSA.

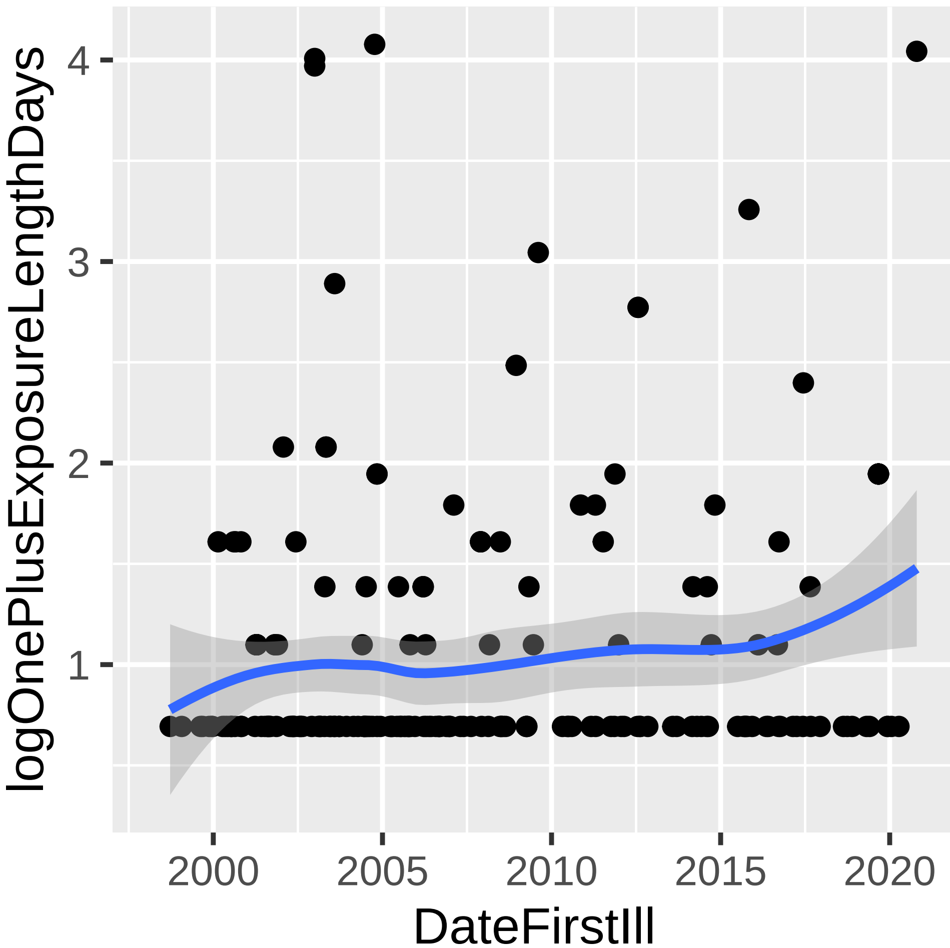

Figure C.2: Outbreak Length over Time. The LOESS curve ($\alpha = 0.5$, polynomial degree is 2) shows a prevailing flat trend. Though the ends suggest an upward trend, this is likely illusory; the lower end's confidence band comfortably includes the sample average, and the upper end is drawn up by the Minnesota COVID outlier outbreak.

metro concentration, and it does not rob the latter of its statistical significance. Further, the early end's confidence bands easily include the middle period's prevailing average; and the late end's upward turn is largely driven by the COVID-era outlier outbreak that caused issues in the governor partisanship analysis (see §3.2.2).

For geographical distribution, chemical outbreaks occur more or less evenly throughout the country. Figure C.3 shows greater frequency in the country's major population centers, though the Miami metro is something of an outlier. This is to be expected given the tropical organisms that often give rise to chemical toxicity; but, reassuringly, my results do not change when I omit Miami outbreaks from the dataset.

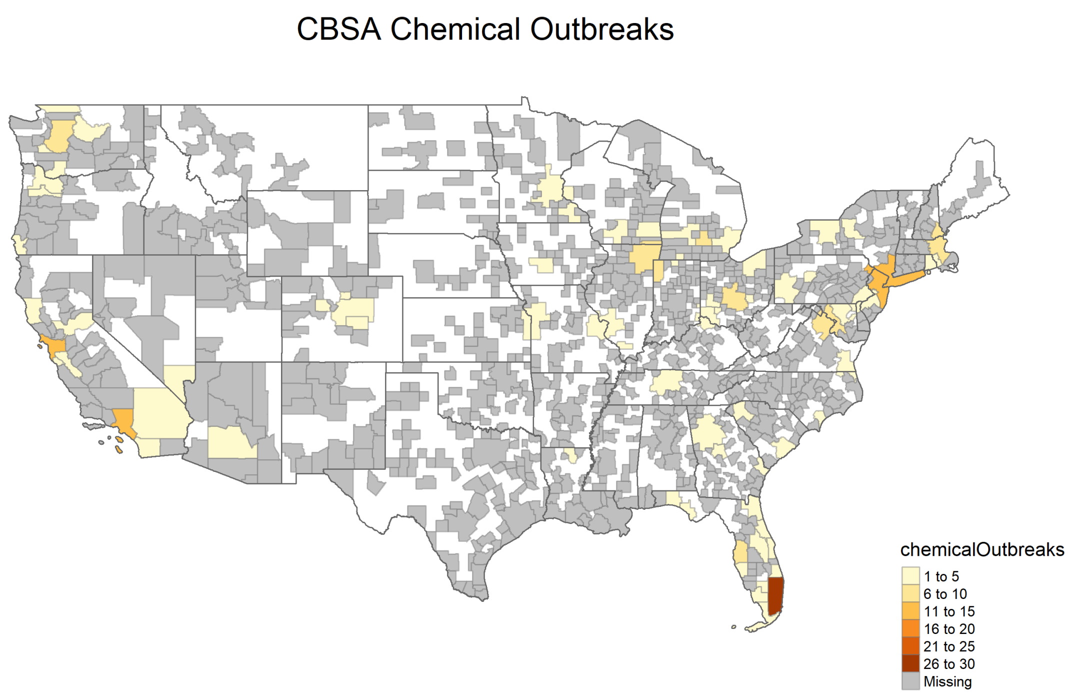

Figure C.3: Total Chemical Outbreaks by CBSA.

https://doi.org/10.33774/apsa-2024-82cvk **ORCID:** https://orcid.org/0000-0001-6810-1260 Content not peer-reviewed by APSA. **License:** All Rights Reserved

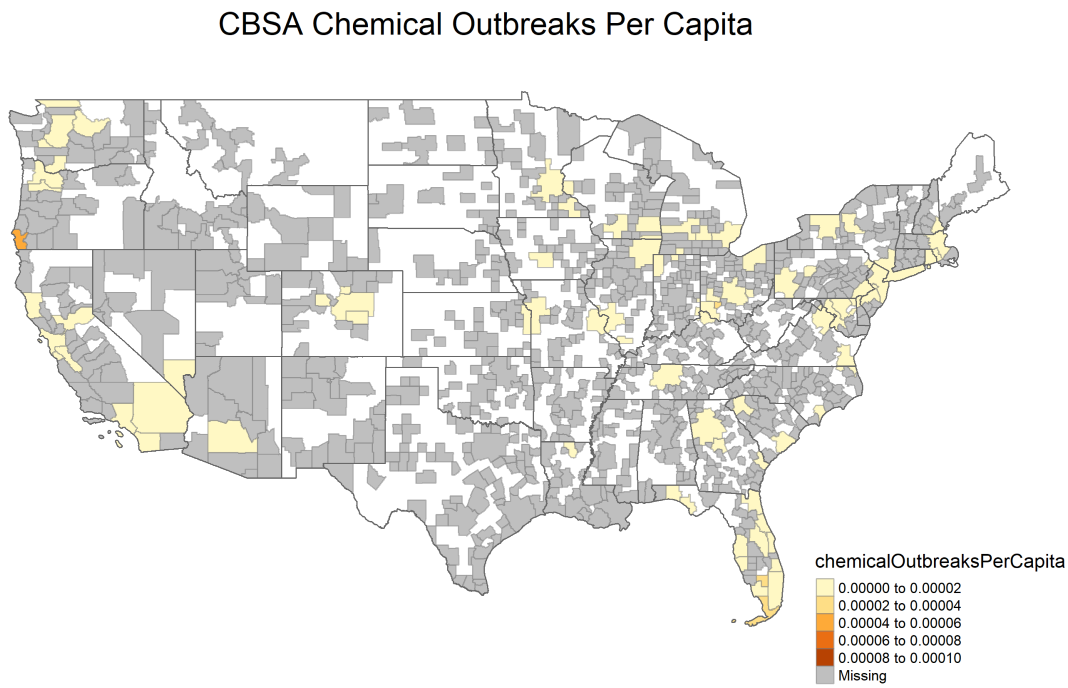

Figure C.4: Total Chemical Outbreaks per Capita by CBSA.

As Figure C.4 shows, when I plot the per capita incidence rate, all the major population centers including Miami just look normal—with a very slight elevation in the rest of South Florida for the same reason I already discussed. The only standout here is Brookings, Oregon; but the weight of its *one* outbreak is amplified by its tiny population of 20,000. This is almost certainly noise; but I rerun the Simple Model without Brookings anyway just to be sure, and the results comfortably survive.

C.2 Sample Balance

https://doi.org/10.33774/apsa-2024-82cvk **ORCID:** https://orcid.org/0000-0001-6810-1260 Content not peer-reviewed by APSA. **License:** All Rights Reserved

	DV: cbsaHerfindahlNorm	
	Balance Model	
(Intercept)	1.54362 ***	(0.00000)
logCBSAPopDensityMilesSq	-0.02709	(0.16382)
logCBSAPopulation	-0.01956	(0.11851)
cbsaHeatingDegreeDaysAnnualAvg	-0.00003 ***	(0.00000)
cbsaCoolingDegreeDaysAnnualAvg	-0.00002 *	(0.02680)
cbsaAnnAvgsStationElevation	0.00007 **	(0.00505)
monthFirstIll1	0.07777	(0.22934)
monthFirstIll2	0.02118	(0.64986)
monthFirstIll3	0.04942	(0.31692)
monthFirstIll4	0.09960 **	(0.00621)
monthFirstIll5	0.05184	(0.21592)
monthFirstIll6	0.05389	(0.19661)
monthFirstIll8	0.03227	(0.35962)
monthFirstIll9	0.04707	(0.19635)
monthFirstIll10	0.10814 **	(0.00757)
monthFirstIll11	0.08381	(0.05944)
monthFirstIll12	0.07614	(0.08971)
postSEDRIC	-0.01233	(0.52633)
logOnePlusIncubationShortDays	-0.03505	(0.72931)
logOnePlusIncubationMedianDays	-0.00594	(0.90441)
cbsaPortTotalVesselCallsAnnualAvg	-0.00005 ***	(0.00000)
cbsaPortTotalVesselCallsIsMissing	-0.08533 **	(0.00600)
N	191	
R2	0.46310	
logLik	150.91593	
BIC	-181.02956	

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table C.1: Predicting metro concentration to check balance.

https://doi.org/10.33774/apsa-2024-82cvk **ORCID:** https://orcid.org/0000-0001-6810-1260 Content not peer-reviewed by APSA. **License:** All Rights Reserved

Appendix D Local Control & Presidential Vote

In the chapter's main text, I demonstrate that state borders prolong outbreaks of chemical foodborne illness by fracturing the public health bureaucracy's communication networks and control structures. I also show that this effect is robust to the inclusion of several omitted variables whose absence might have been driving my results. Among these variables, partisan control of county governments might have been influential since counties often have substantial control over day-to-day public health operations and since Republican county governments could be less interventionistic than their Democratic counterparts. At the same time, however, there was good reason, ex ante, to believe that local partisanship would be minimally influential, if at all.

Local partisanship behaves differently from national partisanship. Unlike national parties, local parties have the ability finely to customize their brand to fit local preferences. For example, a Texas Democrat might carry a gun while a California Republican might support tighter gun control. As a result, the local parties in a given county may converge on policy in a variety of areas—particularly those like food safety that are not nationally politicized. To the extent that this is the case, while food safety policy might still vary substantially across counties, these differences will be attributable far more to regional eccentricities than to which local party holds the county reins.

Nevertheless, to ensure that I do sufficient due diligence, I give the hypothesis the benefit of the doubt and set out to test local partisanship's influence on chemical outbreak length anyway. Unfortunately, this aim is complicated by the fact that local party control in the United States is not uniformly measured. Since the Progressive Era, local US elections have been kept largely non-partisan in an attempt to reduce corruption, to encourage citizen/official engagement unmediated by party, and to emphasize local issues and managerial competence over national policy.[66] Further, even when local elections are partisan, large-scale datasets of those elections' results can be difficult to find—especially when the period of interest does not lie entirely within the age of Big Data.

To circumvent these issues, I gauge local partisan control of county governments using county-level presidential vote as a heuristic and evaluate the heuristic's efficacy in Oregon using, primarily, the party declared on officials' voter registrations. I find that county presidential vote is an excellent predictor of party control of county legislative boards, though the prevalence of appointed and officially non-partisan country executives makes them harder to fit into a partisan framework.

[66] Northup 1987.

D.1 The Heuristic and the Case of Oregon

County-level presidential election results are a useful potential heuristic for local party control because of their uniquely comprehensive coverage and national comparability. On the first point, unlike any other measure of local party, county-level presidential vote is available nationally all the way back to 1788, easily covering my roughly 25-year dataset of the nation's chemical foodborne illness outbreaks.[67] On comparability, county-level presidential vote provides a simultaneous measure on the same set of candidates across all jurisdictions every four years.

Unfortunately, validating my heuristic for local party control even in a limited fashion presents a serious challenge for the same reason that I need the heuristic in the first place: local elections and government are largely non-partisan in the United States. However, one solution to this comes from the fact that local officials are also voters. I will infer that a county board member's or administrator's partisan affiliation matches that of her voter registration; but this in turn requires that I have systematic public data on officials' registered party affiliations as voters.

This requirement turns out to be rather restrictive. To meet it, states must allow voters to declare a party affiliation on their voter registration; they must make voter registration rolls available to the public; and they must include party affiliation on the publicly released rolls along with enough supplemental information to allow me uniquely to link a voter registration record to an official.[68] Among the qualifying states, only 6 have at least 10 counties favoring each major party in 2020 and thereby provide sufficient variation for estimation: Colorado, Florida, Louisiana, North Carolina, Oregon, and Pennsylvania.[69]

Strictly speaking, I could use any of these states to validate my measure of local partisan control. However, because I have limited resources—and since information on Oregon's county government structures is readily available from the Association of Oregon Counties—estimating effective party control in that state is fairly straightforward.[70] Therefore, to check the effectiveness of county-level presidential vote as a predictor of the partisanship of county governments, I unpack the partisanship of each county in Oregon in the aftermath of the 2020 election. Though admittedly limited in scope and far from being nationally representative, this at least provides a cohesive basis for evaluation.

Concerning election year, I limit myself to 2020 primarily because of the prohibitive costs of further expansion. Because of voter registration data vendor access restrictions, I must individually link each official

[67] See Inter-university Consortium for Political and Social Research n.d., Congressional Quarterly 2022, and MIT Election Data Science Lab 2018.

[68] Typically, name, address, and age/birthyear are sufficient to allow for successful disambiguation.

[69] See National Conference of State Legislatures 2022 and Congressional Quarterly 2022.

[70] See their governments table, Association of Oregon Counties 2022.

to their voter registration record; and this process can be time-intensive when multiple registration records are reasonable matches for an official, when officials go by a different name in public life than their legal name, and in a variety of other situations. However, further factors encourage a focus on 2020.

Most importantly, because my voter registration dataset only includes current information, its capacity to support an evaluation of the heuristic's performance in 2012 or even 2016 is more limited than it is for 2020. First, while unlikely, an official's party registration may have been different during those elections; and that difference would not be reflected in my dataset. Second, and more concerning, officials who held office in 2012 and 2016 may have left office since that time and moved to other parts of the state or county. If they moved to another part of the state, their voter registration record could be harder uniquely to identify since I would no longer be able generally to restrict attention to voters at addresses in or near the official's county and, instead, would have to depend on the official having a combination of name and age that was unique throughout the state. Worse, if the official moved to a state where party declarations were not both allowed and publicly reportable, I would not be able to source their party from their voter registration. Finally, if an official from 2012 or 2016 had passed away since that time, their voter registration record would likely not be available in the current version of the rolls, which is the only version to which I have access.[71]

D.2 Data

I draw the data for this analysis from a variety of sources. To generate the roster of Oregon county officials, I rely on official county websites,[72] the State of Oregon's *Blue Book*,[73] the Internet Archive's Wayback Machine,[74] and media reports. For voter registration information, I use L2's DataMapping inferace.[75] And for county-level 2020 presidential election results, I use *Congressional Quarterly*'s *Voting and Elections Collection*.[76]

D.3 County Government Structure

County government structure is critical for determining effective party control. I draw this information primarily from the Association of Oregon Counties' aforementioned report but qualify it with information from each county's official website. On the legislative side, the State of Oregon allows its counties to have either 3-member or 5-member boards. Generally—though not exclusively—these boards are officially non-partisan. On the executive side, not all counties have one; but, in those that do, the executive may be either

[71] See ORS 247.555 (2022) and ORS 247.570 (2022).
[72] Listed at State of Oregon 2023.
[73] State of Oregon Secretary of State 2021.
[74] The Internet Archive 2023a.
[75] L2, Inc. 2022.
[76] Congressional Quarterly 2022.

appointed by the board or directly elected by the citizens. I analyze these paths to office both together and separately. Overwhelmingly, elected county executives also serve as members of the county's legislative body—often under the title "county judge."

D.4 Generating the Roster of Officials

I assemble the dataset using information for officials who held office as of April 1, 2021 (the "record date"). This date allows for any personnel changes resulting from the November 2020 election to take place and for those changes to be reflected on each county's website.

Oregon only holds some of its county elections at the same time as the presidential election. Nevertheless, for the sake of that group, it is valuable for me to delay the record date well into the year following the election so that I can test how well presidential election results reflect the partisan leanings of the local officials who are likely to be in office for the greatest amount of time following the presidential election.

I locate each county's website using the Oregon government's official list, and I use the Internet Archive's Wayback Machine to source archived versions of each county's website.[77] For each office, I record the officeholder's information using the first archived version on or after the record date of a county webpage that reports that information.

In 3 of 36 counties, the first available archived county webpage versions are from 2022.[78] In each of these cases, I record the initial list of officials from those versions, but I also verify that list against the Wayback Machine's mid-2021 archived versions of the State of Oregon's *Blue Book*.[79] For the one official among these who is not listed in the *Blue Book*, I verify his county service using media articles from 2021 which cite both his name and his position.[80]

D.5 Record Linkage & Recorded Variables

To link each official to their voter registration and voting history record in the L2 system, I search for individuals with the official's *public* name in the county for which the official works.[81] This sometimes

[77] See State of Oregon 2023 and The Internet Archive 2023b.

[78] These three counties were Hood River, Wasco, and Washington.

[79] State of Oregon Secretary of State 2021. Similar information is also available in the print volume [CITE], but its vintage is unclear.

[80] The position not documented in the *Blue Book* is the County Administrator for Wasco County; and its occupant was Tyler Stone according to the *Columbia Gorge News* (Gibson 2021), *Gorge County Media* (2021), and *Oregon Live/The Oregonian* (Rogoway 2022).

[81] While probabilistic record linkage methods would be ideal to employ in this context (see Fellegi and Sunter 1969 and Enamorado, Fifield, and Imai 2019), applying these methods would require complete access to the L2 dataset, and this is impractical in L2's DataMapping interface.

returns multiple results while at other times returning no results, and I deploy a different set of strategies for each of these situations.

In the case of multiple results, if all possible matches declare the same party on their voter registration, I simply record that party and move on; otherwise, I add further details until only one declared party remains. I source these details from official biographies on county websites, state compensation records, candidate filings, professional directories, and media reports. Details I check when they are publicly available include middle name or initial, city of residence and its proximity to the official's district and the county seat, city of residence of any publicly noted spouse and adult children, profession, and age range based on photos and educational/professional history.

In cases of no results, I modify the query until I get a plausible match or until I have exhausted the list of plausible variations. Modifications include switching in likely legal names for apparent nicknames, assuming the public first name is actually a middle name, substituting a middle initial for the middle name or dropping it entirely, allowing for multiple last names (which are often entered into the database in a sui generis fashion), and making the search statewide instead of countywide. I address the officials not included in the L2 Oregon data in the next section.

Once I have identified a matching record, I record values for a number of variables to facilitate both analysis and replication: public name, the county for which they work, their role in county government, the name and party on their voter registration, voter registration date, birthyear, city and county of residence, the party of the last major-party primary in which they voted, and their L2 voting frequency.[82]

D.6 Determining an Official's Party Affiliation

To determine each official's partisan affiliation, I primarily assume that each official identifies with the party that they declare on their voter registration. For an average voter, this could be a flawed measure since savvy voters might declare affiliation with the majority party in their state, U.S. House district, or other region so that they can participate in that party's primaries and, in so doing, wield greater direct influence over the political outcomes about which they most care. However, for political candidates and officeholders, since a voter's declared party affiliation is generally a matter of public record in Oregon, declaring for a party other than the one with which they actually affiliate could be read as betrayal—a publicly declared lack of faith in the party's capacity to win; thus, the party on their voter registration is likely to be their genuine party affiliation.

[82] This is an ordinal variable based on the number of elections in which the voter is known to have voted, with higher values indicating greater participation. Unfortunately, L2 does not publicly document the procedures for its generation.

When voter registration party is unavailable for an official—either because they do not specify a party on their registration or because their registration is confidential—I determine their political party using public candidate declarations and/or third-party assertions.[83] For the first, an official often declares their party affiliation publicly when their county allows for partisan county elections or when they run for a non-county partisan office (e.g. state or U.S. legislator). For the second, media organizations and PACs often ascribe a partisan or ideological affiliation to nonpartisan officials; and local political parties sometimes actively marshal support for nonpartisan candidates. Naturally, each of these measures has its limitations.

For public candidate declarations, political opportunism and change over time could invalidate the measure. On the first point, an official might not have strong personal political views and might simply wish to serve in office; instead of acting on personal conviction, this individual could adapt their party affiliation, platform, and behavior in office to each electing constituency's preferences.

While this kind of person is plausible, multiple factors work against their strategy. First are reputational concerns: Voters, PACs, and party organizations all want to know they can rely on an official; and partisan "flip-floppers" do not inspire confidence on that score. Second—and tightly related to the first point—are political networks: Relationships of mutual assistance among political operatives and sources of funding, volunteers, campaign strategy, and campaign infrastructure can take years to cultivate and are generally far more easily and robustly developed within rather than across party lines; thus, a given politician will likely be hard-pressed to work within and draw resources from more than one of these nexuses freely over the course of their political life.

With that said, to the extent that such local flip-floppers do exist, the influence of local party in public health is likely just attenuated. Local partisanship could be less tightly defined and less extreme than its national counterpart; so the local policies that go along with a given party label could vary substantially, making party minimally informative for public health policy and outcomes. In particular, there would likely be a reduced risk that partisan differences among local officials lead to non-cooperation in outbreak investigation and control. In this case, while my heuristic for party would suffer, my broader thesis that borders lengthen outbreaks would enjoy greater plausibility since it would have one less alternative with which to compete.

On the issue of change over time, ideological drift of both parties and politicians could theoretically result in a given politician changing parties; but, at least at the national level and in recent history, this has been empirically unusual. The Southern Realignment that saw Southern Democrats become Southern

[83] Though much voter registration information is a matter of public record in Oregon, registrations may be made confidential for a number of reasons, including the prevention of domestic violence and the protection of those involved in any official capacity with the conduct of elections. See OAR 165-005-0130 (2022), ORS 247.967 (2022), and ORS 247.967 (2022)).

Republicans roughly from 1960 to 1990 is the last instance of any serious volatility in politician partisan affiliation in the US. It is far more typical that such individuals' party affiliations and ideological positions remain stable over their political lifetimes.[84]

Concerning the reliability of media, PAC, and local political party assessments, these organizations all have an incentive accurately to report a politician's partisan and ideological leanings. First, modern media organizations are evaluated on the twin criteria of credibility and the extent to which they cater to the ideological preferences of their readers. If they report a politician's party incorrectly, and if that error results in material harm to their readers' political interests, the organization could easily lose market share to a more reliable alternative.

PACs' and local political parties' incentives are even more direct. Both types of organizations exist for the advancement of political causes; and, if they do not accurately assess and report a politician's political leanings, they risk funneling campaign funds and volunteers to people who could actively undermine the organizations' political objectives.

Whatever doubt these justifications may leave can be further alleviated by the fact that these supplemental measures only affect the analysis through 2 of Oregon's 36 counties. In Clackamas County, Tootie Smith's Republican affiliation is evinced by her public party declaration in a 2014 contest for a US House seat and as a member of the Oregon House from 2001-2005;[85] endorsement by the Oregon Transformation Project, a conservative PAC;[86] and the assertion of two local media organizations, *Williamette Week*[87] and *Oregon Live/The Oregonian*.[88] The same two media organizations also assert a Republican affiliation for Paul Savas, Clackamas County's other officially partyless councilmember.[89] Finally, in Lane County, Heather Buch's Democratic affiliation is evinced by the Lane County Democratic Party's work[90] for her reelection and by her having received an endorsement as a progressive from the Democratic PAC Democracy for America.[91]

[84]See Poole and Rosenthal 2007. As they note, a member of Congress's ideological position (DW-NOMINATE score) does appear to change fairly substantially when they change parties; and this, admittedly, is the reverse of the direction of causality that I hypothesize here. However, since party switching itself remains rare in today's political environment, such ideological changes are unlikely to come into play.

[85]*Tootie Smith* n.d. and Mesh 2012.

[86]Jaquiss 2019.

[87]Jaquiss 2022.

[88]Crombie 2021.

[89]Jaquiss 2022 and Stringer 2022.

[90]Democratic Party of Lane County n.d.

[91]Democracy for America n.d.

D.7 Estimating Effective County Party Control

Ideally, I would estimate party control of the whole county government. Unfortunately, however, it is impossible to determine a superior branch for each county solely on the basis of county government structure. On one hand, each Oregon county executive is typically singular and, as a consequence, theoretically more energetic, reflecting the national government's form and the president's recent ascendance in the face of congressional gridlock; but a number of factors in Oregon work in favor of county legislative power. By comparison with the national government, first, county legislatures are small, ranging in size from 3 to 5 people; and this allows for easier policy coordination among the members. Second, with a unicameral system, the county legislatures have fewer veto points. Third, a smaller bureaucracy and a closer proximity thereto make legislative oversight easier. Finally, in many counties, the county executive is appointed by the county legislature and serves at its pleasure.

Given these ambiguities—which likely carry weight both within Oregon and nationally—instead of attempting to divine a superior branch for each county, I simply estimate effective party control for each county's executive and legislative branches independently. The executive branch is typically led by a single individual, so the branch's party affiliation can be inferred from that one individual. On the other hand, the legislative branch's party affiliation can be estimated using the majority affiliation of its constituent members.

D.8 Descriptive Analyses

For the descriptive analyses, I convert the county's continuous Republican 2-party presidential vote in the 2020 election into an indicator for whether the Republican candidate had a simple majority of that vote. The resulting contingency tables show that county 2-party presidential vote share predicts county legislative control well while evidence on its efficacy in predicting county administrative control is mixed.

On the legislative side, as shown in Table D.1, when I determine party control of the county legislature using only the party on the councilmembers' voter registrations, presidential vote share successfully predicts 82% of cases. Encouragingly, this is comparable to DW-NOMINATE's first dimension's success rate in predicting the voting behavior of members of Congress.[92] As shown in Table D.2, when I supplement voter registrations with candidate, party, PAC, and media indicators, the prediction success rate remains essentially unchanged, with a slight drop to 80%.

[92] See Poole and Rosenthal 2007, p. 34.

	countyPresVote2PartyMajorityRep	
countyLegisMajorityRegRep	*No*	*Yes*
No	5 (0.15)	3 (0.09)
Yes	3 (0.09)	22 (0.67)

Table D.1: County Legislative Control by Registered Party. In each cell, raw counts are on the left while table-wide proportions are on the right and enclosed in parentheses. The county's 2-party presidential vote successfully predicts 82% of cases in this setup.

	countyPresVote2PartyMajorityRep	
countyLegisMajorityRep	*No*	*Yes*
No	6 (0.17)	3 (0.09)
Yes	4 (0.11)	22 (0.63)

Table D.2: County Legislative Control by Any Party Measure. The county's 2-party presidential vote successfully predicts 80% of cases in this setup.

On the administrative side, I determine party using only voter registrations—mostly because of information availability. On one hand, all of the elected administrators registered an affiliation with one of the major parties; so no supplemental sources were necessary. On the other, 5 of the 13 appointed administrators (38%) were registered non-partisans; and the supplemental sources provided absolutely no help in determining these administrators' party affiliations. These people apparently did not ever run for partisan public office; and PACs, parties, and media outlets did not comment on their partisanship. Thus, either voter registration party was all that was necessary, or no publicly available source had the information.

As shown in **??**, when I pool both elected and appointed administrators, Republican 2-party presidential vote share only successfully predicts 64% of county administrators' party affiliations. In contrast, as **??** shows, when I restrict only to *elected* administrators, the rate of successful prediction jumps to an impressive 89%. Naturally, from a statistical significance perspective, this success is heavily qualified by the small sample size of 9.

Nevertheless, taken together, these results suggest that elected administrators are true politicians while appointed administrators are bureaucrats. Elected administrators likely tie their behavior and branding freely and openly to political parties since they keep their jobs through reelection. In contrast, since appointed administrators answer to the county legislature, if they maintain a studied non-partisan public image while competently managing the county and faithfully executing the legislature's orders, they have a reasonable chance of keeping their jobs even when party control of the county legislature changes hands.

	countyPresVote2PartyMajorityRep	
countyAdminRegRep	*No*	*Yes*
No	7 (0.28)	4 (0.16)
Yes	2 (0.08)	12 (0.48)

Table D.3: All County Administrators, Elected and Appointed. Republican Version. Presidential vote share successfully predicts 76% of county administrators' party affiliations.

	countyPresVote2PartyMajorityRep	
countyAdminRegRep	*No*	*Yes*
No	1 (0.11)	1 (0.11)
Yes	0 (0)	7 (0.78)

Table D.4: Elected Administrators Only. Republican Version. Presidential vote share correctly predicts 89% of cases, but the sample size is small. Since all elected administrators are registered with one of the two major parties, the Democratic version reveals no additional information and is omitted.

For the outbreak analyses, this implies that 2-party presidential vote share is an effective heuristic for party control of county government, both on the legislative and the administrative sides. The measure ably performs for all elected county branches, even if the data on the administrative side is sparse; and, while its performance with appointed branches is less impressive, this is likely because personal partisan identity does not play as strong a role in an appointed official's daily work. In short, where party actually matters, 2-party presidential vote share predicts it fairly well.

	countyPresVote2PartyMajorityDem	
countyAdminRegDem	*No*	*Yes*
No	14 (0.56)	7 (0.28)
Yes	2 (0.08)	2 (0.08)

Table D.5: All County Administrators, Elected and Appointed. Democratic Version. Presidential vote share successfully predicts only 64% of county administrators' party affiliations when the positive registration party is Democratic.

D.9 Regression Analyses

In contrast with the contingency tables, the regression analyses have the advantage of being able to evaluate the predictive accuracy of the *continuous* presidential vote share variable I use in the main analyses—not just a dichotomous variant. The outcome variable in each model is an indicator for Republican control of the given branch, and the estimation method is probit regression.

On the legislative side, for models using only voter registration for party, I handle missing data using a worst-case approach. First, I create the set of all possible complete datasets, varying the values for the incomplete observations in each. Next, I estimate the model separately on each of those datasets. Finally, I report only the estimates from the model that is most pessimistic—that is, the one with the highest p value on the presidential vote variable. The results using these methods largely echo the findings from the descriptive section.

Republican 2-party presidential vote share predicts Republican legislative control well. As Table D.6 shows, a county's Republican presidential vote is positively and statistically significantly associated with Republican control of the county legislature, whether I measure a councilmember's party using only their voter registration or all available public sources. The dichotomous predictor has similar results. As Table D.7 shows, the Republican presidential candidate's having taken a majority of the county's 2-party vote share is positively and statistically significantly associated with the Republican having held control of the county legislature, regardless of how an official's party is measured. While the effect is smaller and less precise for the voter-registration-only model, this is a conservative estimate that may be artificially deflated by my approach for missing data.

On the administrative side, mixed motivations and insufficient data again rear their ugly heads. As **??** shows, when I pool elected and appointed administrators, county presidential vote does not predict their partisanship effectively; but, as I discuss above, this is to be expected since appointed administrators likely benefit from keeping their job firmly separate from their personal political preferences. Only marginally more

	DV: countyLegisMajorityRegRep	
	Registered Only	All Party Sources
(Intercept)	-2.167 *	-2.915 **
	(0.026)	(0.009)
countyPresVote2PartyShareRep	0.052 **	0.066 **
	(0.004)	(0.002)
N	36	35
logLik	-14.972	-12.569
BIC	37.111	32.249

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. p values in parentheses.

Table D.6: Republican Legislative Control & Continuous Presidential Vote.

concerning are the results for just elected administrators. While presidential vote does not demonstrate a statistically significant capacity to predict elected administrator party in either of the models in **??**, the limited sample size and limited variance meant that the models were reasonably likely to miss an effect even if there was one.

Overall, to the extent that the data speak, they demonstrate that a county's presidential vote predicts estimated partisan control of county government well. Predictions on the legislative side are strong and statistically significant. Predictions for elected administrators are strong in the limited data that do exist. And, while predictions for appointed administrators are not successful, this is perfectly consistent with these people keeping job and party separate; instead of following their personal partisan inclinations, they likely do the bidding of their elected legislative bosses for the sake of job security. Together, these should account for effective partisan control.

Importantly, I cannot claim that these findings are neatly generalizable to other elections and states. This analysis covers the counties of just one state during only one election, and they do not constitute a random sample from the universe of county-election pairs in American politics; they are simply the pairs for which data was available and most easily accessible. However, while the view is limited, it is nevertheless promising and consistent with my use of presidential vote as a heuristic for party control of county government.

	DV: countyLegisMajorityRegRep	
	Registered Only	All Party Sources
(Intercept)	-0.000	-0.253
	(1.000)	(0.528)
countyPresVote2PartyMajorityRep	1.020 *	1.428 **
	(0.040)	(0.006)
N	36	35
logLik	-18.094	-15.903
BIC	43.355	38.917

*** p < 0.001; ** p < 0.01; * p < 0.05. p values in parentheses.

Table D.7: Republican Legislative Control & Majority Presidential Vote.

	DV: countyAdminRegRep	
	Continuous Pres Vote	Indicator Pres Vote
(Intercept)	-2.640 *	-0.765
	(0.011)	(0.100)
countyPresVote2PartyShareRep	0.049 **	
	(0.005)	
countyPresVote2PartyMajorityRep		1.439 *
		(0.013)
N	25	25
logLik	-12.345	-13.765
BIC	31.127	33.967

*** p < 0.001; ** p < 0.01; * p < 0.05. p values in parentheses.

Table D.8: Republican Executive Control & Presidential Vote, Elected and Appointed Administrators.

	DV: countyAdminRegDem	
	Continuous Pres Vote	Indicator Pres Vote
(Intercept)	-1.211	-1.150 **
	(0.118)	(0.004)
countyPresVote2PartyShareDem	0.005	
	(0.759)	
countyPresVote2PartyMajorityDem		0.386
		(0.530)
N	25	25
logLik	-10.940	-10.796
BIC	28.317	28.029

*** p < 0.001; ** p < 0.01; * p < 0.05. p values in parentheses.

Table D.9: Democratic Executive Control & Presidential Vote, Elected and Appointed Administrators.

	DV: countyAdminRegRep	
	Continuous Pres Vote	Indicator Pres Vote
(Intercept)	-38.939	-5.576
	(0.333)	(0.995)
countyPresVote2PartyShareRep	0.544	
	(0.328)	
countyPresVote2PartyMajorityRep		6.726
		(0.994)
N	9	9
logLik	-1.677	-3.014
BIC	7.748	10.423

*** p < 0.001; ** p < 0.01; * p < 0.05. p values in parentheses.

Table D.10: Republican Executive Control & Presidential Vote, Elected Administrators Only. The Democratic version has identical coefficients in the variables of interest.