



**HARVARD Kennedy School**  
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

# **Sunlight and Protection against Influenza**

## **Faculty Research Working Paper Series**

---

**David J.G. Slusky**

University of Kansas

**Richard J. Zeckhauser**

Harvard Kennedy School

**February 2018**

**Updated July 2020**

**RWP18-007**

Visit the **HKS Faculty Research Working Paper Series** at:

[https://www.hks.harvard.edu/research-insights/publications?f%5B0%5D=publication\\_types%3A121](https://www.hks.harvard.edu/research-insights/publications?f%5B0%5D=publication_types%3A121)

The views expressed in the **HKS Faculty Research Working Paper Series** are those of the author(s) and do not necessarily reflect those of the John F. Kennedy School of Government or of Harvard University. Faculty Research Working Papers have not undergone formal review and approval. Such papers are included in this series to elicit feedback and to encourage debate on important public policy challenges. Copyright belongs to the author(s). Papers may be downloaded for personal use only.

# Sunlight and Protection Against Influenza

David J.G. Slusky  
Richard J. Zeckhauser

## Abstract

Recent medical literature suggests that vitamin D supplementation protects against acute respiratory tract infection. Humans exposed to sunlight produce vitamin D directly. This paper investigates how differences in sunlight, as measured over several years across states and during the same calendar week, affect influenza incidence. We find that sunlight strongly protects against getting influenza. This relationship is driven almost entirely by the severe H1N1 epidemic in fall 2009. A 10% increase in relative sunlight decreases the influenza index in September or October by 1.1 points on a 10-point scale. A second, complementary study employs a separate data set to study flu incidence in counties in New York State. The results are strongly in accord.

*Keywords: Seasonal Influenza, Sunlight, Vitamin D, Natural Experiment, H1N1*

*JEL codes: I10, I12, I18, Q5, N32*

---

## Affiliations:

Slusky: Department of Economics, University of Kansas, 1460 Jayhawk Blvd., 415 Snow Hall, Lawrence, KS, 66045, [david.slusky@ku.edu](mailto:david.slusky@ku.edu)

Zeckhauser: Harvard Kennedy School, Harvard University, 79 JFK St, Cambridge, MA, 02138, [richard\\_zeckhauser@harvard.edu](mailto:richard_zeckhauser@harvard.edu)

Acknowledgements: We thank Nils C. Wernerfelt for exceptional advice and guidance throughout this project. We are also immensely grateful to Janet Currie and Diane Alexander for their help in implementing their geographically granular influenza measure, to Mark Anderson and Dan Rees for providing their early twentieth century mortality data, and to Corey White for sharing his data on the influenza vaccine match rates by season. We also thank Susan Graber, Ryan Brown, and conference participants at the 2018 American Society of Health Economists conference for their comments and suggestions. We are especially grateful to two anonymous referees who provided extensive and insightful comments and suggestions.

Declarations of interest: None. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

© 2020 by David J. G. Slusky and Richard J. Zeckhauser. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission, provided that full credit, including © notice, is given to the source.

## 1. Introduction

Seasonal influenza has been with humans throughout history (Viboud and Epstein 2016). It imposes extreme costs on contemporary societies, with the Spanish flu epidemic of 1918 – which actually stretched from 1917 to 1920 -- being a notable high outlier (CDC 2018). Beyond the significant discomfort to those it strikes, influenza saps productivity when individuals cannot work (Duarte et al. 2017) and absorbs health care resources (Schanzer and Schwartz 2013). Influenza also has less known long-range consequences. Notably, individuals exposed to influenza in utero have lower earnings as adults, are more likely to depend on government assistance (Almond 2006; Schwandt 2017), and are more likely to suffer from serious health problems later in life (Lin and Liu 2014). They are also more likely to have a heart attack (Kwong et al. 2018). Influenza severity can create capacity constraints on hospitals, magnifying existing disparities in whom a hospital chooses to admit (Alexander and Currie 2017). Finally, influenza infections can cause an untimely death. The parallels between the consequences of and measures to combat flu and COVID-19 have stimulated far greater attention to flu in light of the latter pandemic.

Influenza is a type of viral respiratory infection. Traditional public health measures to combat it include vaccination (Maurer 2009; and White 2019) and paid sick leave to keep contagious workers at home (Barmby and Larguem 2009; Pichler and Ziebarth 2017; Asfaw, Rosa, and Pana-Cryan 2017; Pichler, Wen, and Ziebarth 2020). Coincidental reductions in interpersonal contact (such as from holiday school closings and public transportation strikes) can also reduce prevalence (Adda 2016). Finally, a recent meta-analysis shows that ingested vitamin D pills help to protect against these types of infections (Martineau et al., 2017).<sup>1</sup>

---

<sup>1</sup> The Martineau meta-analysis imposed stringent criteria for including a trial. This minimizes concerns about a variety of confounding factors, such as selection effects.

This paper analyzes the potential of another mechanism for securing vitamin D: direct bodily production of vitamin D when exposed to sunlight (Holick 2007). This paper tests this mechanism's performance directly by studying population-level vitamin D production by humans experiencing sunlight exposure, using a reduced form analysis of the relationship between flu and sunlight.<sup>2</sup> While we can ingest vitamin D from many sources, such as fish and fortified milk, passive exposure to sunlight is a much more effective source.<sup>3</sup> Sunlight as a source has two added benefits. First, unlike ingested vitamin D, which can become toxic at a certain concentration, the self-production mechanism does not generate toxic quantities (Holick 2007). Second, a far greater percent of the population in an area is exposed to sunlight than secures significant vitamin D for ingestion. Thus, the level of exposure is more likely to be in the range where herd immunity is significant.

Normally, flu season is in the winter, when both the average and variance of the sunlight level are low. The experience with the H1N1 flu epidemic in 2009 was dramatically different. The epidemic peaked much earlier, in the late summer and early fall. Those few months also had lower than average sunlight. This is sufficient concurrent variation in both sunlight and flu to study the relationship.

---

<sup>2</sup> This reduced form approach unfortunately prevents us from differentiating among competing mechanisms, such as ultraviolet light decreasing folate status and hampering immune function (Branda and Eaton 1978; Cohn 2002; Borradaile et al. 2014; Jones et al. 2018) or alternatively deactivating the virus directly (Sagripanti and Lytle 2007).

<sup>3</sup> The minimum amount of sunlight exposure (on head, neck, arm, and hands, without sunscreen) necessary to reach saturation (and therefore produce an effective allotment) varies greatly by latitude, weather, time of year, and skin tone. In the summer it can be as short as a few minutes; in the winter it can be over an hour. See [http://nadir.nilu.no/~olaeng/fastrt/VitD-ez\\_quartMEDandMED\\_v2.html](http://nadir.nilu.no/~olaeng/fastrt/VitD-ez_quartMEDandMED_v2.html) to calculate the minimum effective exposure time given a certain set of conditions.

Also, an interesting feature of these calculations is that while excessive unprotected sun exposure can lead to sunburn and increased risk of skin cancer, the minimum exposure time for adequate vitamin D is always less than the minimum exposure for a sunburn.

This relationship between sunlight and flu has been studied in the broader medical literature (as by Charland et al. 2009; Grant and Giovannucci 2009; and Soebitantyo et al. 2015). Our study time period also overlaps the H1N1 outbreak in 2009. The relationship between the 2009 H1N1 outbreak and Vitamin D has been studied specifically (e.g., Bruce et al. 2010; Momplaisir et al. 2012; Khare et al. 2013; Urashima et al. 2014), albeit with inconclusive results. Grant and Giovannucci (2009) is probably the closest to our study in that it uses sunlight data (specifically, UVB irradiance) and case-fatality rates for the 1918-1919 H1N1 influenza pandemic. It discusses many of the biological pathways by which vitamin D can help against influenza, including reducing production of cytokines and the associated inflammation. While multiple flu strains (e.g., H1N1 and H5N1, responsible for avian flu) induce a cytokine response, vitamin D did not reduce the risk of death for mice infected with H5N1 (Grant and Giovannucci 2009).

The studies on the 2009 H1N1 outbreak do not substantively investigate the relationship between it and sunlight. This paper is therefore the first to investigate the relationship between the 2009 H1N1 outbreak and sunlight, and the first to perform its analysis at two levels of aggregation (across states in the U.S. and across counties in New York). The results at the two levels are consistent.<sup>4</sup>

Our paper also fits more broadly in the economics literature of the health effects of sunlight. In previous work, we found a strong protective impact of a pregnant woman's exposure to sunlight on later-in-life asthma in her child (Wernerfelt, Slusky, and Zeckhauser 2017). Others have found

---

<sup>4</sup> In Appendix E, we also extend our analysis to mortality, both for the modern period covered in the data in this paper, but also for the 1918 influenza pandemic given that there are a number of similarities. They are the only two recorded H1N1 epidemics. They are the two largest epidemics since 1900, though the 1918 epidemic was far worse. The majority of mortality for the earlier epidemic occurred between October 1918 and January 1919, a similar pattern to the 2009 epidemic, suggesting that sunlight might have played a role. The 1918 case has been studied extensively in the economics literature as a driver of adverse outcomes apart from health (Almond 2006; Brown and Thomas 2018; Beach, Ferrie, and Saaverda 2018; Fletcher 2018). It has received much less attention in the economics literature as an adverse outcome itself (Clay, Lewis, and Severini 2018; Clay, Lewis, and Severini 2019). None of these latter studies considered sunlight.

that sunlight exposure in utero favorably affects birth weight and gestational age (Trudeau, Conway, and Menclova 2016; Zhang, Chen, Wang, and Zhang 2020), and also favorably affects fertility (Conway and Trudeau 2019).

## 2. Data

For influenza data, we used data from the CDC. For cases of flu, we used the CDC index that aggregates data reports from the individual state health department influenza surveillance points, and then harmonizes the aggregate to a consistent 10-point scale. The scale measures the ratio of visits to outpatient healthcare providers by those with symptoms of influenza, relative to all outpatient visits (regardless of symptoms). Each point on the scale represents an additional standard deviation above the mean. Weekly state-level data are available, from October 2008 to the present.<sup>5</sup> Some states, however, are missing individual weeks of data. Dropping the jurisdictions with missing flu data or sunlight data leaves us with 28 states for our primary analysis sample (CDC 2017a).<sup>6</sup>

We combined this flu data with the National Solar Radiation Database (NSRDB)’s daily sunlight data for 2003-2016. It covers all states but Alaska, plus the District of Columbia. This data reports the solar radiation for a particular set of coordinates (in watts per square meter). We calculate our primary independent variable by downloading the hourly sunlight data for the population-weighted county centroid, averaging across each week, and then constructing a county

---

<sup>5</sup> See Appendix A for more details about the how the index is calculated.

<sup>6</sup> Those 28 states are: Alabama, Arizona, California, Georgia, Hawaii, Illinois, Indiana, Kansas, Maine, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Hampshire, New Jersey, Ohio, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, West Virginia, Wisconsin, and Wyoming. As shown in Appendix Table 1, when we include all 49 states with sunlight and flu data (plus the District of Columbia), using whatever data is available for each month, we find consistent results.

population-weighted average across counties for each state-week (Census 2010).<sup>7</sup> The dataset also includes data on temperature and humidity. While our influenza data only begins in 2008, earlier data was used solely for placebo tests (NREL 2018). For one of our robustness checks, we also include precipitation data (which is not in NSRDB) from NOAA’s Global Surface Summary of the Day (NOAA 2017), which utilizes data from 1,218 weather stations spread throughout the United States. We assigned this data to states by matching the station closest to the population-weighted centroid for each county and then averaging with in each state-week across counties, as described above (Census 2010).<sup>8</sup>

We supplement our state-level analysis with a county-level analysis, following Alexander and Curie (2017), who construct a ZIP-code level weekly flu measure for New Jersey. Here, we use SPARCS hospital discharge data (New York State Department of Health 2015) for all of New York for October 2008 (the earliest month we have CDC flu data) to June 2014 (the last year for which we have discharge data and the last quarter for which we have bed data).

### 3. Methodology

As described above, Martineau et al. (2017)’s meta-analysis of randomized controls demonstrated significant benefits of vitamin D supplements for reducing the likelihood that an individual will contract an acute upper respiratory infection. Randomized controlled trials have served as the gold standard for epidemiological investigation. This analysis follows an alternate path to methodological soundness. It employs quasi-experimental variation to effectively create equivalent randomization, a common econometric technique. Implicitly, this approach controls for

---

<sup>7</sup> Appendix D shows comparable results using census tract level sunlight for the 100 largest counties in the U.S.

<sup>8</sup> The correlation between the week-month average temperature variables from the two data sets is 0.9947, suggestions that there is no issue with combining weather variables from both.

a wide number of variables. It thereby avoids the inevitable selection problems that arise when individuals must volunteer for randomized controlled trials. The current study thus employs an independent variable over which individuals had effectively no control: the deviation of a state's sunlight from its normal level.

Ideally, an econometric study would run a two-stage instrumental variable analysis, where the first stage used sunlight to predict vitamin D levels and the second stage used predicted vitamin D levels to predict influenza. Unfortunately, we lack any large scale, geo-tagged data on vitamin D levels. In its stead, our analysis employs a “reduced form” estimate of sunlight’s impact on influenza. Given that sunlight levels in a geographic area for a particular week vary randomly over the years, this provides us with a robust estimate.

Vitamin D is fat-soluble (unlike vitamin C, for example, which is water-soluble) and, therefore, has a half-life of between two weeks and two months (Mawer, Schaefer, Lumb, and Stanbury 1971; and Jones 2008). Thus, we are most interested in, and therefore calculate, the sunlight received over the week of the influenza report and seven weeks prior. This choice of windows is conservative, for two reasons: if the half-life is long, it includes the right window; if the half-life is short, it includes extra information, which would bias our results downwards.<sup>9</sup>

Our variable is a weighted average (by county population). Such weighting is important, because the more populous areas have a greater impact on the flu index, which is a function of the count of outpatient visits.

---

<sup>9</sup> Appendix Table C10 re-estimates our main results for past windows of two to seven weeks and finds all are statistically significant and directionally consistent, with the magnitude increasing with the length of the window.



We estimate the impact of the percent of deviation of sunlight (the change in log points, following Wernerfelt, Slusky, and Zeckhauser 2017) from its mean on deviations of the flu index from its mean as follows:

$$Flu_{swy} = \alpha + \gamma \ln(sunlight_{swy}) + \theta \mathbf{Weather}_{swy} + \mathbf{stateweek}_{sw} + \mathbf{fluyear}_y + \varepsilon_{swy}$$

$Flu_{swy}$  is the flu index for state  $s$  in week  $w$  in flu year  $y$ . By flu year, we mean the twelve months from July 1 of one year to June 30 of the next year, denoted by the first of the two years (e.g., flu year 2009 is July 2009 to June 2010). Given that influenza troughs in the summer, this prevents an artificial break between December and January in the middle of a flu season.  $Sunlight_{swy}$  refers to the average sunlight for week  $w$  and the prior seven weeks (as described above) for state  $s$  in year  $y$ .<sup>10</sup> We use the epidemiological definition of a week (always Sunday-Saturday), which is the CDC standard.<sup>11</sup>  $\gamma$  is our coefficient of interest.

There is a chance that our results could be picking up some other kind of weather variation. Hence, we include a vector of weather controls in **Weather**. These controls include temperature, given that it is known to have health effects (Deschenes 2013; Barreca, Deschenes, and Guldi 2018; Barreca et al. 2016; and Huetal, Miller, and Molitor 2020). Also, following Barreca, Deschenes, and Guldi (2018) and Wernerfelt, Slusky, and Zeckhauser (2017), we control for the number of days per week that a state experiences extreme cold (daily low temperature below 15°F). Such a measure is merited, because the influenza virus can survive better at lower temperatures (Polozov et al. 2008). Absolute humidity can also play a role in influenza mortality. Prior work identifies a negative nonlinear relationship between humidity and influenza, where levels below 6

---

<sup>10</sup> Wernerfelt, Slusky, and Zeckhauser (2017) use data from the American Time Use Survey to show that relatively increases in sunlight increase relative time spent outdoors. We rely on their validation of this measure of sunlight and do not repeat their analysis.

<sup>11</sup> [https://wwwn.cdc.gov/nndss/document/MMWR\\_Week\\_overview.pdf](https://wwwn.cdc.gov/nndss/document/MMWR_Week_overview.pdf), using the Stata command “epiweek”

g of water vapor per kg of air had a substantial impact (per Barreca 2016; Barreca and Shimshack 2012).<sup>12</sup> Finally, we include precipitation, as it is possible that a lack of sunlight is acting through this channel.

Our preferred specification also includes interaction terms (**stateweek**) for state-week fixed effects (for example, week 40 in Kansas), and flu year fixed effects (for example, July 2009-June 2010).<sup>13</sup> Robust standard errors are clustered at the state level.<sup>14</sup> Flu year fixed effects are also particularly appropriate given that the specific strains of influenza differ from year to year and vary significantly in their intensities (hence visits to the hospital if infected) and degrees of contagion. This specification follows our prior work examining the link between sunlight and vitamin D in relation to asthma (Wernerfelt, Slusky, and Zeckhauser 2017).

To evaluate and bolster the robustness of our results, we conducted a second analysis at the county level in New York State. That analysis followed the methodology of Alexander and Currie (2017) for constructing a local measure of influenza intensity. The method of the county analysis can be briefly described as follows:

1. Identify all emergency department discharges and all inpatient discharges with an emergency department indicator (since those admitted from the ED drop out of the ED file).

---

<sup>12</sup> Specific humidity is not directly provided in the NSRDB data, so we calculated it using the available information on dew point and atmospheric pressure and the Tetens equation. See [http://snowball.millersville.edu/~adecaria/ESCI241/esci241\\_lesson06\\_humidity.pdf](http://snowball.millersville.edu/~adecaria/ESCI241/esci241_lesson06_humidity.pdf) for the necessary formulas.

<sup>13</sup> Previous literature on the relationship between sunlight and flu (including Charland et al. 2009; Grant and Giovannucci 2009; and Soebitantyo et al. 2015) does not make use of fixed effects models. Given the substantial variation in latitude, weather sunlight and flu severity across states, fixed effects are crucial to ensure that estimates measure the impact of relative sunlight variation on relative flu variation, as opposed to merely identifying simple correlations.

<sup>14</sup> We also estimate our main state-level and county-level results adjusting for spatial correlation, using the "acreg" function available at <https://acregstata.weebly.com/> (Colella et al. 2019). Using a conservative distance cutoff of 10,000 km (so that any state in the contiguous United States and certainly any county in New York State could be spatially correlated with another), we find results that are still statistically significant nationally (p-value < 0.001) and for New York (p-value = 0.018).

2. Keep those emergency discharges with an influenza diagnosis (CCS<sup>15</sup> code of 123) and inpatient discharges with an influenza diagnosis that was present on arrival.
3. Use the admitted date to assign an epidemiological week.

We use these counts of influenza discharges to construct a county-level measure, again following Alexander and Currie (2017). Briefly, the steps are as follows:

1. For each hospital, merge in bed data (New York State Department of Health 2016)<sup>16</sup> and divide the number of admissions in that week by the number of beds to get the per bed admissions rate.
2. For each county centroid (Census 2010), calculate the great circle distance to the geocoded coordinates of each hospital's address.
3. For each hospital within 100 miles, divide the per bed influenza rate by the distance between the county and the hospital and then sum to get the county level influenza flu index

We then merged this county-level influenza measure with the county-level sunlight and weather data from NSRDB. New York has 62 counties, hence slightly more than the 28 units in the main analysis.

#### 4. Results

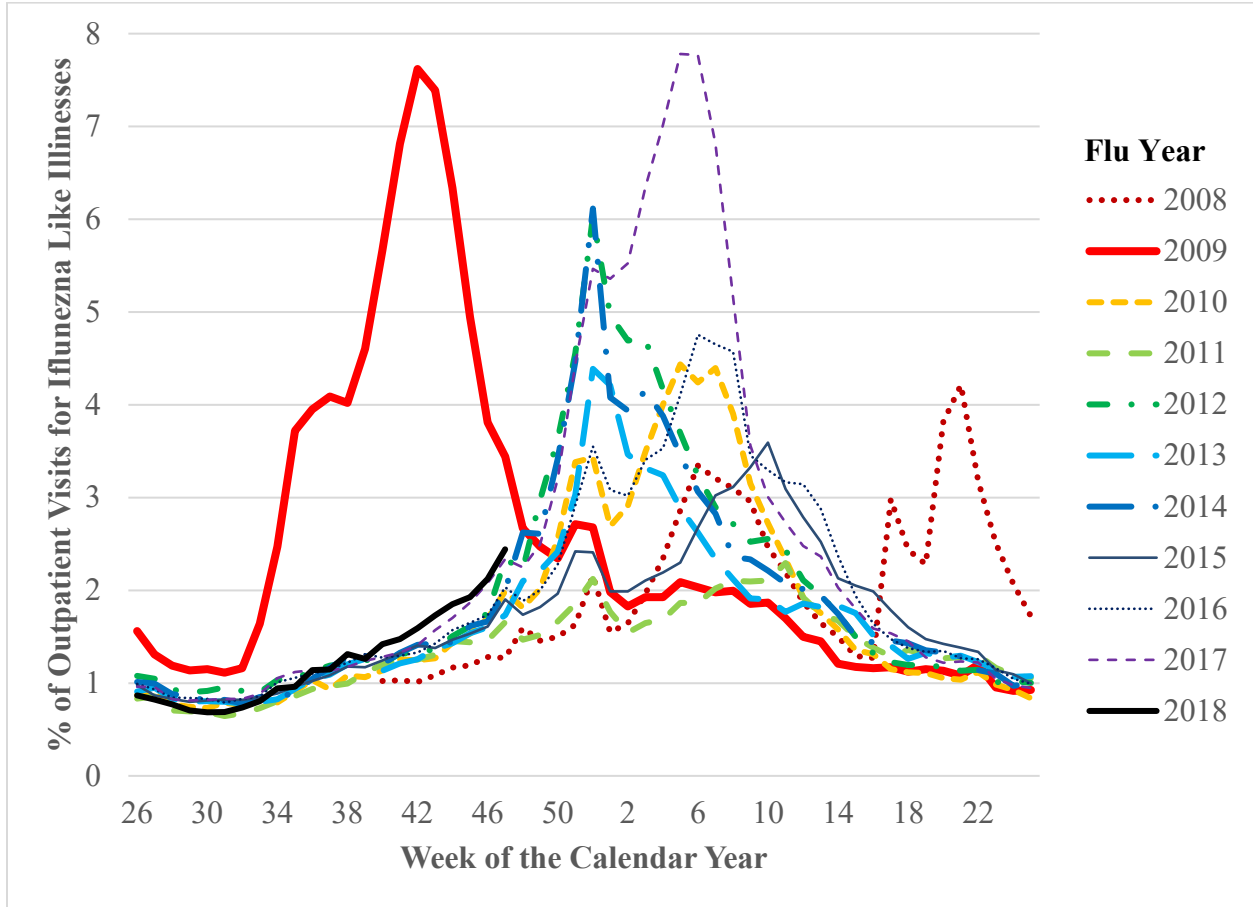
As mentioned above, our time period overlaps the H1N1 epidemic of 2009. This is crucial, because of both the timing and severity of that season (CDC 2017b), as shown in Figure 1. Warmer colors represent earlier years, and cooler colors later years. Here a year refers to a July-June flu year denoted, as described above, by the first calendar year involved, i.e., the one starting in July.

---

<sup>15</sup> <https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp>

<sup>16</sup> The current number of beds is available on the New York State Department of Health's website. Historical information through the second quarter of 2014 was obtained in response to an email request.

**Figure 1: Weekly National Influenza Intensity**



Here, we see that the 2009 flu year was extremely atypical. It was by far both the most severe and occurred much earlier in the year (weeks 33 to 48, corresponding to August to November) than any other. Its level of greatest severity occurred during a time of the year with more sunlight overall

as well as much more sunlight variation around long-term average levels). No other flu year approached its extreme outlier status on any of these measures.<sup>17,18</sup>

Table 1 shows summary statistics for the flu index and population-weighted average sunlight levels, as well as for other weather variables (used as additional controls.)

**Table 1: Summary Statistics**

	(1) N	(2) Mean	(3) StDev	(4) Min	(5) Max
Flu index	12,068	2.049	2.289	1	10
Sunlight (W/m <sup>2</sup> )	12,068	183.1	71.61	46.49	358.3
Temperature (°F)	12,068	53.92	17.27	4.033	89.22
Days/week temp <15°F	12,068	12.26	6.460	1.662	28.35
Specific humidity (g water vapor / kg air)	12,068	0.403	0.993	0	6.693
Days/week specific humidity < 6 g/kg	12,068	1.731	2.133	0	7
Precipitation (inches / day)	12,068	0.0819	0.0505	0.000006.43	0.365

Note: Unit of observation is a year-week for each of the 28 states that have complete flu and sunlight data.

We see that the flu index varies between 1 and 10, with an average level of 2. Sunlight also varies widely, specifically by latitude, weather, and season. Temperature and humidity also vary extensively.

Table 2 shows our initial regression results for the impact of sunlight on the influenza index, using the state-month and year fixed effect strategy described above.

<sup>17</sup> We also examine variation in population-weighted sunlight averages (in kilojoules per square meter per day). Appendix Figure B1 shows the three-year (2009-2016) average. We see the expected pattern, which is that the United States is sunnier in the south and west. Appendix Figure B2 then shows the variation by state in the average influenza index. Here we see a somewhat different pattern than in Appendix Figure B1. Some sunny states have high flu levels (such as Texas and California), and some low flu levels (for instance, Arizona). Moreover, some less sunny states also have high flu levels (such as Illinois), and some have low flu levels (such as Maine and New Hampshire). Still, the correlation between the two is 0.53, which is statistically significant at the 1% level (p value = 0.0037). This simple analysis, though, pools seasons and years. Other state-specific factors may influence influenza levels, which makes controlling for state-specific fixed effects important.

<sup>18</sup> Appendix Figure B3 addresses this issue. It plots the ranges of influenza and sunlight by month. The top half of that figure shows that there is flu activity in the late summer, fall, and winter, but that activity is minimal in the spring and summer (except in outlier situations). The lower half shows the expected seasonal variation in sunlight levels, with large amounts of sunlight in the spring and summer and substantially less in the fall and winter.

**Table 2: Main Results of Sunlight on Flu, All Months**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-1.403** (0.597)	-2.324*** (0.738)	-1.310** (0.577)	-1.824** (0.779)
Observations	12,068	12,068	12,068	12,068
R-squared	0.066	0.069	0.069	0.074
Weather levels		X		X
Weather thresholds			X	X

Notes: 2008-2016. All regressions include state-week and flu year fixed effects. The panel consists of the 28 states that have complete flu and sunlight data. Weather level controls include the log of temperature, specific humidity, and precipitation. Weather thresholds include extreme cold (daily low temperature below 15°F) and low humidity (below 6 g of water vapor per kg of air). Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Column (1) of Table 2 shows that a 10% increase in the average sunlight over the past two months reduces the 10-point flu index by 0.1403.

However, Table 2 includes months that have minimal influenza activity, and also months that have low levels of sunlight. Including these months dramatically reduces the magnitude of the coefficients, and diminishes measured seasonality in the results.

Table 3 re-estimates our model for each month of flu data after including the impact of that month and the prior month's sunlight. State-week and flu-year fixed effects are still included.<sup>19</sup>

<sup>19</sup> We assign weeks to a month based on the last day (Saturday) of the week. Given variations in what day of the week a month starts on and the length of each month, each month does have exactly the same sample size.

**Table 3: Month by Month<sup>20</sup>**

	(1) Jan	(2) Feb	(3) Mar	(4) Apr	(5) May	(6) Jun	(7) Jul	(8) Aug	(9) Sep	(10) Oct	(11) Nov	(12) Dec	(13) Sep & Oct
Log sunlight for that month and the prior month	1.390 (1.110)	0.514 (1.978)	2.548 (2.054)	1.461 (1.594)	1.656* (0.902)	-1.592 (2.021)	0.459 (1.053)	1.697 (1.012)	-7.315** (3.136)	-9.625*** (2.012)	-0.376 (2.297)	1.909 (1.679)	-10.56*** (1.438)
N	1,036	896	980	952	1,008	952	980	1,008	924	1,148	1,092	1,092	2,072
R-squared	0.648	0.344	0.239	0.203	0.092	0.065	0.075	0.084	0.472	0.781	0.546	0.447	0.663

Notes: 2008-2016. All regressions include state-week and flu year fixed effects and all weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

<sup>20</sup> Because the flu data begins in October 2008, the regressions for October, November, and December have an additional year of observations for each of the 28 states included in the primary analytic sample.

It shows that September influenza (that is, August and September sunlight), and to a lesser extent October influenza (that is, September and October sunlight) are driving our results. These months meet the dual requirements (as shown in Appendix Figure B3) of non-trivial level of influenza activity and still-substantial levels of sunlight. For these two months, shown in Column (13), a 10% increase in relative sunlight levels leads to a 1.1-point decline in the influenza index.

Table 3 shows that the statistically significant results are found primarily in the late summer and early fall, and Figure 1 shows that the majority of flu cases in this time of year were in the H1N1 epidemic of 2009. Those results might lead one to wonder whether our results are present in years other than 2009. Table 4 shows several analyses for that year alone, and for all of the other years.

**Table 4: The Role of 2009**

*Panel A: Only 2009*

Outcome Variable	(1)	(2)	(3)	(4)	(5)	(6)
		Flu Index		Difference in Flu Index from the State-Week Mean		
Log sunlight for that month and the prior month	-5.256*** (0.694)	-21.43*** (2.035)	-20.41*** (2.560)			
Difference in log sunlight for that month and the prior month from the state-month mean				-11.78*** (3.459)	-22.66*** (5.461)	-39.64*** (5.495)
N	1,456	392	252	1,456	392	252
R-squared	0.240	0.508	0.405	0.097	0.273	0.445
Months	All	Aug-Oct	Sep-Oct	All	Aug-Oct	Sep-Oct

Notes: All regressions include weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. Only flu year 2009. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



Panel B: Years Other Than 2009

Outcome variable	(1)	(2) Flu Index	(3)
Log sunlight for that month and the prior month	2.113** (0.811)	0.0814 (0.517)	0.0836 (0.777)
Observations	10,612	2,688	1,820
R-squared	0.093	0.017	0.022
Months	All	Aug-Oct	Sep-Oct

Notes: All regressions include state-week and flu year fixed effects and all weather controls. Flu years 2008 & 2010-2016. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

Columns (1)-(3) in Panel A show the results of regressing the flu index for a state and month on the log sunlight for that month and the prior month for only the 12 months in 2009. For any category of months (all, August-October, or September-October) there is a *negative* relationship between sunlight levels and flu that is strong both statistically and in magnitude.

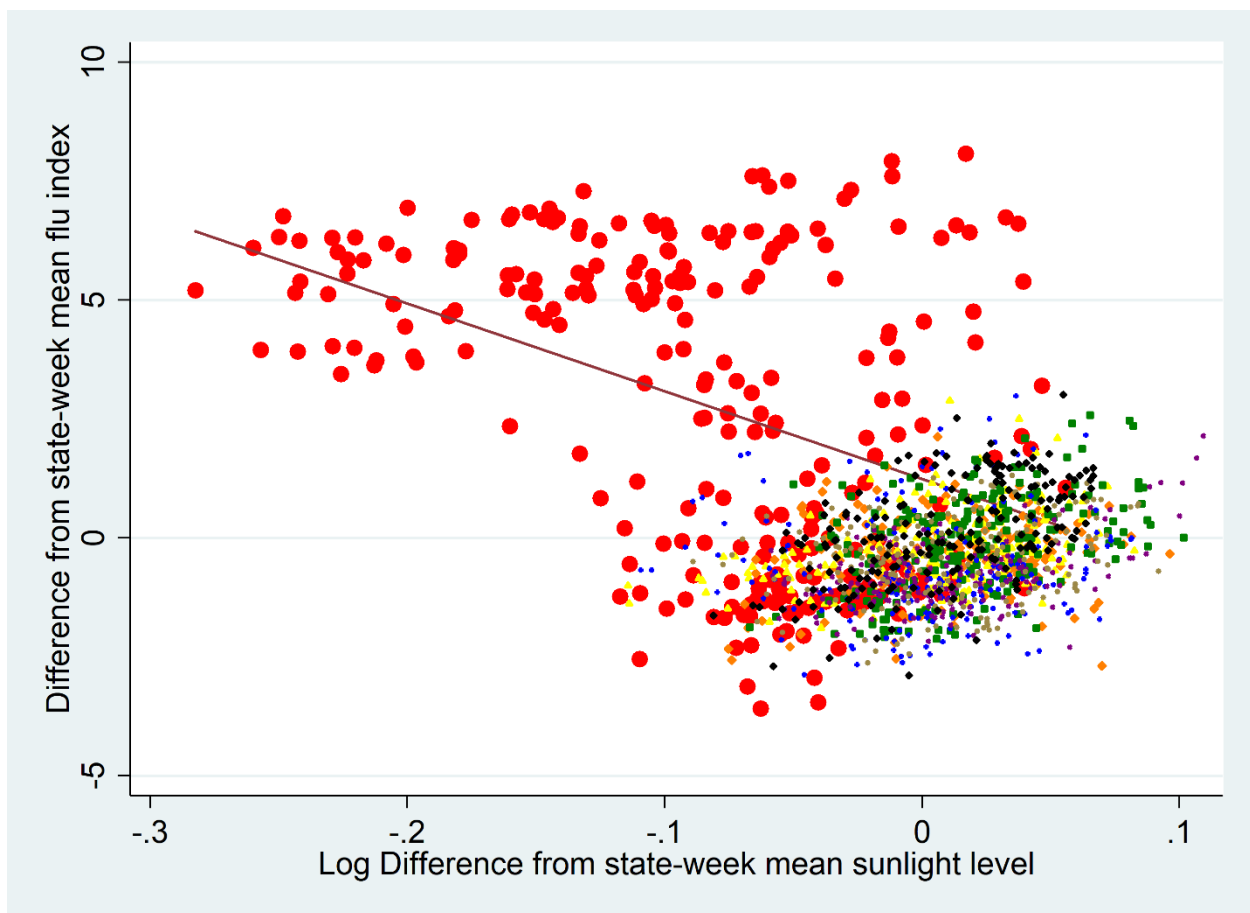
In theory, this association could be endogenous: individuals who chose to live in each of these states could have chosen it understanding of how sunny it is on average each month. Therefore, following the spirit of the main analysis above, we instead calculate differences from the state-week mean value, as log points for sunlight, and as points on the flu index for its incidence. This then measures not the level of sunlight but the difference in sunlight when compared to an average year. That independent variable should be exogenous. Analysis with these endogenous variables yields much larger results. The results differ little with and without August flu. Going forward, we therefore stick to September and October flu.

Finally, in Panel B, we re-estimate the main model in its normal form (i.e., year and state-week fixed effects), but excluding the data from 2009. We find, if anything, marginally statistically

significant positive results. This further supports our hypothesis that 2009 was a special flu year, with an early and intense season. That year is responsible for any cross-year results.

We can also see this result in graphical form. Figure 2 graphs the deviations in the September and October influenza index and the log level of August/September and September/October sunlight from the mean for each state and month. The vertical axis represents the residual after regressing the difference in the state-week mean flu index on the difference in the state-week mean weather controls.

**Figure 2: State-Week Deviations for Flu and Sunlight, September and October**



Notes: Red Circles = 2009; Orange Diamonds = 2010, Yellow Triangles = 2011, Green Squares = 2012, Blue Pluses = 2013, Purple X's = 2014, Brown Small Circles = 2015, Black Small Diamonds = 2016. Line is linear best fit for 2009. The vertical axis indicates the residual after regressing the difference in the state-week mean flu index on the difference in the state-week mean weather controls.

The horizontal axis displays our independent variable, the log of sunlight by date and month. The vertical axis plots our dependent variable, flu index by state and month, in the difference variables in log-points calculated for Table 4 above. Thus, if sunlight is protective, then the greater its level for a state and a month, the lesser will be the flu index for that state and month. As can be seen by the vertical placement of observations in Figure 2, consistent with Figure 1 above, the 2009 flu season was substantially more severe than any other season in our sample. In addition, there was substantially less sunlight that year than in the average year, and much more variability in those deviations.<sup>21</sup> These deviations facilitate estimation. Hence, in 2009 a clear and substantial negative relationship emerges between relative differences in sunlight and relative differences in flu level.

We now turn to examine, and hopefully replicate, our results at the sub-state level. To replicate, we first need to show that our results are consistent for at least some subsets of states.<sup>22</sup> Table 5 shows our results stratifying by quartiles of overall average sunlight levels.<sup>23</sup> Here we include all states (even those with missing weeks of influenza data) and all months, both to maximize statistical power and because the state we will eventually look within (New York) has only nearly complete CDC data. (It is missing one week of data out of 451 weeks.)

---

<sup>21</sup> Appendix Figure B4 stratifies each of the two panels in Appendix Figure B3 into 2009 and the other years, and also shows that the fall of 2009 had comparatively more flu and less sunlight.

<sup>22</sup> As we'll show below, our results are robust to omitting each state one at a time, and so are not driven by any individual state.

<sup>23</sup> 1<sup>st</sup> Quartile: Connecticut, Delaware, District of Columbia, Maine, Michigan, Minnesota, Montana, New Hampshire, North Dakota, Oregon, Vermont, Washington, and Wisconsin

2<sup>nd</sup> Quartile: Idaho, Illinois, Indiana, Iowa, Massachusetts, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, South Dakota, and West Virginia,

3<sup>rd</sup> Quartile: Arkansas, Kansas, Kentucky, Maryland, Missouri, Nebraska, North Carolina, South Carolina, Tennessee, Utah, Virginia, and Wyoming

4<sup>th</sup> Quartile: Alabama, Arizona, California, Colorado, Florida, Georgia, Hawaii, Louisiana, Mississippi, Nevada, New Mexico, Oklahoma, and Texas

**Table 5: Results Stratified by Average Sunniness of State**

Quartile of Sunniness	(1) 1 <sup>st</sup> (least sunny)	(2) 2 <sup>nd</sup>	(3) 3 <sup>rd</sup>	(4) 4 <sup>th</sup> (sunniest)
Log sunlight for that month and the prior month	-2.514*** (0.652)	-2.463** (0.971)	-2.522** (1.123)	-0.833 (1.357)
Observations	5,083	5,095	5,100	5,592
R-squared	0.075	0.074	0.093	0.091
Number of state-weeks	663	636	636	689
States	13	12	12	13

Notes: 2008-2016. All months. All regressions include state-week and flu year fixed effects and weather controls. Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

Apart from the sunniest quartile of states, our results are broadly consistent across quartiles. We can therefore proceed to analyze within New York (which falls in the second quartile).<sup>24,25</sup>

In Table 6, we that see across a variety of flu measures and sample months the results are highly consistent with those at the national level, as shown above.<sup>26</sup>

<sup>24</sup> To validate the county-level approach, following Alexander and Currie, (2017) Appendix Figure B5 compares the weekly New York State flu index from the CDC (2017a) with the total number of influenza admissions that week in New York. While our discharge-based measure is less spikey than the CDC's measure, the two track each other remarkably well. This is all the more significant given that the CDC measure is based on outpatient office visits and not emergency department visits.

<sup>25</sup> We expand upon Table 5 in Appendix B. Appendix Table B1 investigates further whether our fall (September and October) results are driven by overall sunniest, the H1N1 epidemic in the fall of 2009, or sunshine variation specifically in the fall of 2009. We find consistent statistically significant results for the above median sunniest states and the below median states. We find consistently insignificant results for both groups of states when we exclude 2009. If we instead group states by average deviations from the mean log sunlight level (the horizontal axis in Figure 2), the results are driven entirely by the states that had deviations with magnitudes larger than the median. This suggests our results are driven both by H1N1 in 2009 and by larger sunlight deviations in those months. They are not driven by overall sunniness or lack thereof.

Appendix Table B2 considers the typical peak flu months of October to March for the less sunny and more sunny states. While the point estimates are comparable for the two groups, the effect is highly statistically significant ( $p = 0.003$ ) for the less sunny states and not statistically significant ( $p = 0.136$ ) for the more sunny states.

<sup>26</sup> Appendix Table B3 stratifies the results in Table 6 by race and by age, as the New York State county-based measure data comes from individual-level discharge data, which contains both of these variables. The hypothesis here is that sunlight would be more protective to the non-elderly population (under age 65) who are more likely to go outside, and white individuals given that they convert sunlight to Vitamin D faster. While all of the results are statistically significant at the 1% level, the results are much larger in magnitude for white and non-elderly individuals.

**Table 6: New York State County-Level Analysis for Hospital-Based Influenza Measure**

	(1) Per Bed	(2) Per Bed	(3) Per Bed	(4) Count	(5) Count
Log sunlight for that month and the prior month	-0.268*** (0.0333)	-0.997*** (0.245)	-0.774*** (0.108)	-59.72*** (12.31)	-85.74*** (13.76)
Dropping Outlier Counties			X		
Observations	20,651	3,139	3,087	20,651	3,139
R-squared	0.066	0.241	0.368	0.038	0.370
Months	All	Sep-Oct	Sep-Oct	Sep-Oct	Sep-Oct

Notes: 2008-2014. All regressions include county-week and year fixed effects and weather controls.<sup>27</sup> Robust standard errors clustered at the county level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

We now stratify the New York results by years, as above. Table 7 is analogous to Table 4, first looking at only 2009, and then at the rest of the years excluding 2009. This separation tells if our results appear in only 2009, given that it is the source of our late summer/early fall flu variation, or in other years as well.

<sup>27</sup> While there are no state-weeks with zero precipitation, there are a handful of county-weeks that have zero precipitation, making it impossible to take the log. We therefore include the level of precipitation instead. In Appendix Table C7, we show that for the state level results (where we can include either log or level of precipitation) it only makes a minimal difference which we include (less than 2% in the magnitude of the coefficient). Therefore we do not think this accommodation here is consequential.

**Table 7: The Role of 2009, New York Counties***Panel A: Only 2009*

Outcome Variable	(1) Hospital-Based Flu Measure	(2)	(3)	(4) Difference in Flu Measure from the County-Week Mean	(5)	(6)
Log sunlight for that month and the prior month	-0.202*** (0.0290)	-0.976*** (0.220)	-1.514*** (0.288)			
Difference in log sunlight for that month and the prior month from the state-month mean				-0.560*** (0.0609)	-2.306*** (0.567)	-2.795*** (0.607)
N	3,220	867	558	3,220	867	558
R-squared	0.160	0.293	0.303	0.102	0.193	0.224
Months	All	Aug-Oct	Sep-Oct	All	Aug-Oct	Sep-Oct

Notes: Flu year 2009 only. Robust standard errors clustered at the county level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

*Panel B: Years Other Than 2009*

Outcome variable	(1) Hospital-Based Flu Measure	(2)	(3)
Log sunlight for that month and the prior month	0.0127 (0.0175)	-0.0158*** (0.00400)	-0.0211*** (0.00783)
Observations	17,431	3,901	2,581
R-squared	0.134	0.079	0.087
Months	All	Aug-Oct	Sep-Oct

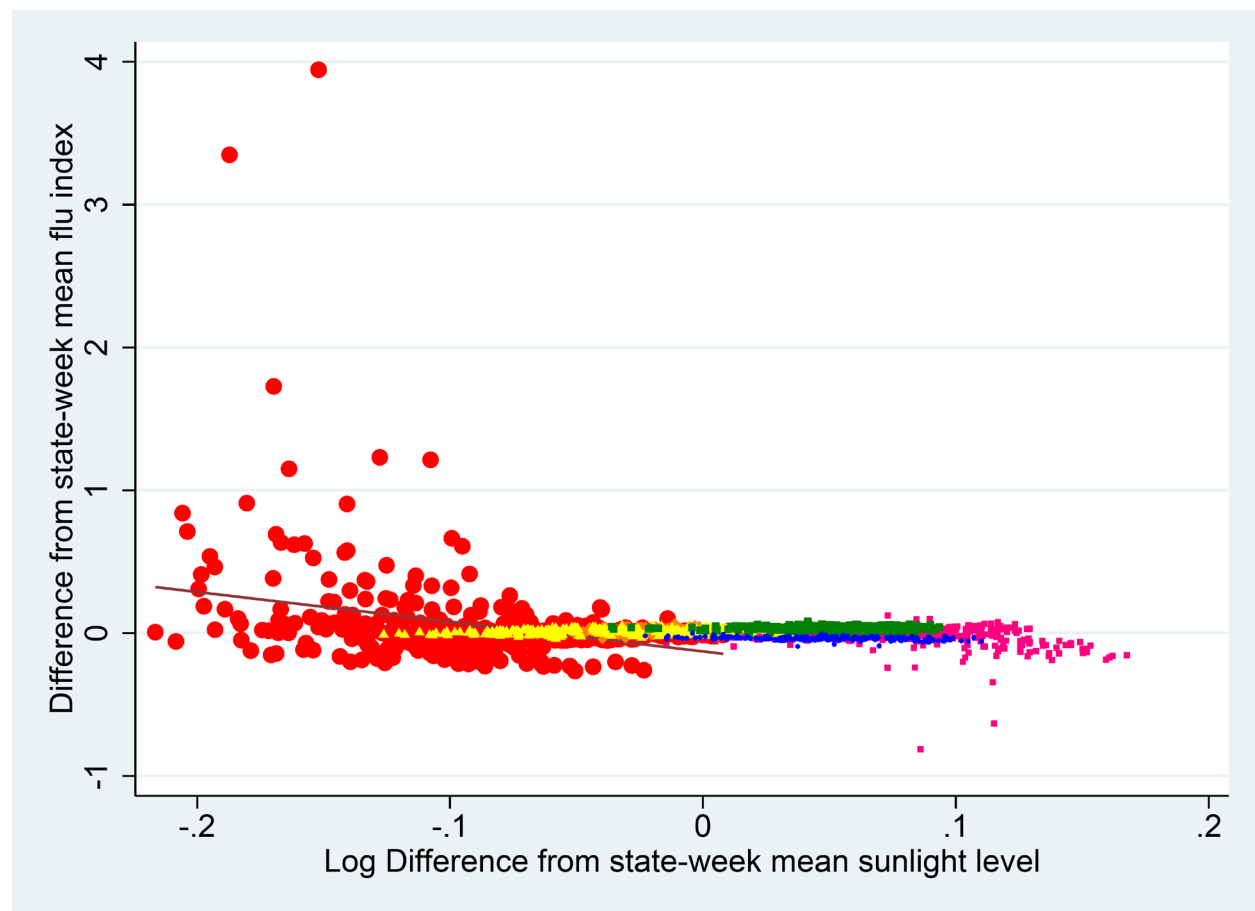
Notes: Flu years 2008, 2010-2014. All regressions include county-week and flu-year fixed effects and weather controls. Robust standard errors clustered at the county level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

Panel A shows consistent results, both in the cross section and using differences from county-month averages across all years. The panel B, in Column (1), finds no statistically significant

result, analogous to Table 4 above. In Columns (2) and (3), despite excluding 2009, identifies the statistically significant result that relatively greater levels of fall sunlight led to relatively lower influenza intensity. The coefficients, however, are a tiny fraction of what they were for 2009.

We now consider Figure 3, analogous to Figure 2. It compares the differences in sunlight and flu within counties across years. Again, the vertical axis represents the residual after regressing the difference in the state-week mean flu index on the difference in the state-week mean weather controls.

**Figure 3: County-Month Deviations for Flu and Sunlight, September and October**



Notes: Pink Small Squares = 2008, Red Circles = 2009; Orange Diamonds = 2010, Yellow Triangles = 2011, Green Squares = 2012, Blue Pluses = 2013, Purple X's = 2014, Brown Small Circles = 2015, Black Small Diamonds = 2016. Line is linear best fit for 2009. Y axis is the residual after regressing difference in state-week mean flu index on difference in state-week mean weather controls.

Here we see a similar relationship to the one above, where the variation is driven by 2009 (consistent with Figures 1 and 2 above). The relationship, as expected, slopes downward (more sunlight implies less flu). The results are robust to dropping the two outliers at the top left of the graph (with a difference greater than 3) —see Table 6, Column (3).

## 5. Robustness Checks

The tables in Appendix C conduct additional robustness checks. Appendix Table C1 repeats the Table 2 analyses, but includes an unbalanced panel of all contiguous states, Hawaii, and D.C. (i.e., even those with missing influenza data in some weeks). That table shows a comparable result. It also employs both linear and quadratic specifications. All three specifications find strongly statistically significant results, though obviously at different coefficient magnitudes.

Appendix Table C2 drops each of the 28 states in the primary specification, one at a time, to test whether the main result persists if any one state is excluded. The answer is yes. Appendix Table C3 performs the analysis for only sunlight from each day of the week (e.g., the average sunlight on Sundays in a given month). It finds that the sunlight for every day of the week has an impact.

Appendix Table C4 performs the analysis at a monthly level with state-month fixed effects. The primary variable calculated as the average over the previous two months. The results are consistent with those found at the week level.

The next three tables all contain variants on our weather controls. Appendix Table C5 adds weather controls one at a time and finds minimal effect on the primary finding. Appendix Table C6 uses more flexible dummy variables for ventiles (e.g., 20 bins that at the 5, 10, 15, etc. percentiles of the variables data) of each weather control, and finds comparable results. Finally,



Appendix Table C7 includes the level of precipitation instead of its logarithm; it too finds comparable results.

Appendix Table C8 controls for the annual flu vaccine match rate, using data from White (2019). Because this control is collinear with the flu year fixed effects, the results employ one or the other but not both. Including the flu vaccine match rate produces much larger results in magnitude, but a much lower  $r$  squared. That outcome is consistent with this effectively being a more rigid form for including flu year fixed effects.

Appendix Table C9 utilizes data on the flu vaccination rates from the Behavioral Risk Factor Surveillance System (BRFSS) from (CDC 2019). BRFSS asks individuals not only whether they've had a flu vaccination in the past 12 months, but also what month they had it in. Given that, as described above, we define a flu year as from July-June, we can use these two variables to create a variable for whether each respondent had a flu vaccination in that season. This information can be aggregated up to a rate at the state-week level. That rate appears in Column (1).

This measure is very noisy, so some adjustments are necessary. First, we linearly interpolate when there is no respondent in a particular state-week. These interpolated values appear in Column (2). Then we can also compute a 7-week moving average to smooth out some of the noise. Those averages are shown in Column (3). Finally, sometimes the vaccination rate appears to go down. We thus employ a ratchet for each season, so that come November of each flu year it only goes up and we ignore any data suggesting it would go back down (through May). The ratcheted values appear in Column (4). Our results are robust to utilizing any of these measures.

Appendix Table C10 reconsiders our choice to use average sunlight data over the previous eight weeks. It looks instead at all windows from two weeks to seven weeks. The results are all

statistically significant at the 1% level, and the magnitude of the coefficient grows with each week beyond two that is added to the window, which suggests a persisting effect of recent sunlight.

## 6. Discussion

6.1 Impact on welfare. How important are the impacts we identify on welfare? As described above, each point on the influenza index represents one standard deviation above the mean share of the non-flu week's ratio of outpatients presenting with symptoms of influenza to all outpatients (CDC 2017a). (That data is also available on the actual outpatient counts, though those counts are not broken down at the state level (CDC 2017b.) In the 2005-2008 "pre-period," this mean share is 1.03%, and the standard deviation is 0.394 percentage points.<sup>28</sup>

Figure 2 shows that the range of relative sunlight levels for September and October within state-weeks across years. That range is roughly plus or minus 0.05 log points, that is, 10 percentage points. Thus, our coefficient for log sunlight shown in Table 3 corresponds to a 1.056-point reduction in the influenza index, which can be interpreted as approximately 1.056 standard deviations. Given that one standard deviation is 0.394 percentage points in flu incidence, -1.056 standard deviations represents 0.416 percentage points, which represents a 40% reduction on the mean of 1.03%.

The average annual number of total outpatients in September and October (weeks 35 to 43) in our study years (2009-2016) (from CDC 2017b) is 6,342,726. A 0.416 percentage point reduction in flu incidence would lead to 26,390 fewer outpatient cases.

6.2 Herd Immunity. Giving 100 people in a town of perhaps 10,000 people a vitamin D supplement will offer extremely few externalities of protection. But give that same town extra

---

<sup>28</sup> See Appendix A for additional calculation details.

sunlight, and most members of the community will produce their own vitamin D. This mass effect will convey an externality of protection, namely herd protection against influenza, a highly communicable disease.<sup>29</sup> Posit that supplements and sunlight-produced vitamin D are equally powerful. That externality should make the protective effect of sunlight far greater for an individual than would a vitamin D supplement.

To test this conjecture, we compared our results to those for vitamin D supplementation. The Martineau et al. (2017) meta-analysis of 25 randomized controlled trials of vitamin D supplementation found an adjusted odds ratio of 0.88 for acute respiratory tract infections. 42.2% of the control group experienced at least one acute respiratory tract infection, but only 40.3% of the treatment group did, thus an unadjusted difference of 1.9 percentage points (4.5%). A possible contributor to this disparity are the externalities promoting herd immunity when sunlight is the protective factor. An alternate explanation, of course, is that part or all of the disparity arises because sunlight produces more vitamin D or more effective vitamin D than supplements.

6.3 Possible alternative mechanisms. Multiple other mechanisms might be at work. Increased time spent outdoors enjoying the sunlight (and therefore less time in confined indoor spaces exposed to sick individuals), for example could reduce interpersonal transmission.

Sunlight can also protect against influenza via a path quite apart from the production of vitamin D. For example, ultraviolet light deactivates the virus directly (Sagripanti and Lytle 2007). Data beyond that in this paper would be required to assess the relative contributions of time spent in close proximity to others, virus deactivation, and vitamin D production. However, we can be

---

<sup>29</sup> This herd immunity obviously would also benefit those who do not go outdoors, as the more outdoorsy people with whom they come in contact would be less likely to be infected and thereby contagious.

confident that the vitamin D path is consequential, as the Martineau et al. (2017) meta-analysis demonstrates.

## 7. Conclusion

Sunlight, likely operating through the well-established channel of producing vitamin D, has the potential to play a significant role in reducing flu incidence. A recent meta-analysis of 25 randomized controlled trials of vitamin D supplementation (Martineau et al. 2017) demonstrated significant benefits of such supplements for reducing the likelihood that an individual will contract an acute upper respiratory infection. The current study considers sunlight as an alternate, natural path through which humans can and do secure vitamin D. This study's findings complement and reinforce the Martineau et al. findings.

Our major result is that incremental sunlight in the late summer and early fall has the potential to reduce the incidence of influenza. Sunlight had a dramatic effect in 2009, when sunlight was well below average at the national level, and the flu came early. Our result is potentially relevant not just to the current COVID-19 pandemic, but also to a future outlier H1N1 pandemic. The threat is there; some H1N1 viruses already exist in animals (Sun et al. 2020). One must be cautious, though, with generalizations, given the unique economic circumstances (e.g., a 25-year high unemployment rate) in the fall of 2009.

Apart from its methodological contributions, this study reinforces the long-held assertion that vitamin D protects against acute upper respiratory infections. One can secure vitamin D through supplements, or through a walk outdoors, particularly on a day when the sun shines brightly. When most walk, through herd protection, all benefit.

## References

- Adda, J. 2016. Economic Activity and The Spread of Viral Diseases: Evidence From High Frequency Data. *Quarterly Journal of Economics*, 131(2): 891–941.
- Alexander, D. and J. Currie. 2017. Are publicly insured children less likely to be admitted to hospital than the privately insured (and does it matter)? *Economics & Human Biology* 25: 33–51.
- Almond, D. 2006. Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy*, 114 (4): 672–712.
- Asfaw, A., R. Rosa, and R. Pana-Cyan. 2017. Potential Economic Benefits of Paid Sick Leave in Reducing Absenteeism Related to the Spread of Influenza-Like Illness. *Journal of Occupational Environmental Medicine* 59(9): 822–829.
- Barmby, T. and M. Larguem. 2009. Coughs and sneezes spread diseases: An empirical study of absenteeism and infectious illness. *Journal of Health Economics* 28: 1012–1017.
- Barreca, A.I. 2012. Climate change, humidity, and mortality in the United States. *Journal of Environmental Economics and Management* 63(1):19–34.
- Barreca, A.I., K. Clay, O. Deschenes, M. Greenstone, and J.S. Shapiro. 2016. Adapting to Climate Change: The Remarkable Decline in the US Temperature-Mortality Relationship over the 20th Century. *Journal of Political Economy* 124(1): 105–159.
- Barreca, A.I., O. Deschenes, and M. Guldi. 2018. Maybe Next Month? The Dynamic Effects of Ambient Temperature on Fertility. *Demography*, 55(4):1269–1293.
- Barreca, A.I. and J.P. Shimshack. 2012. Absolute Humidity, Temperature, and Influenza Mortality: 30 Years of County-Level Evidence from the United States. *American Journal of Epidemiology* 176(7): S114–S122.
- Bartlett, M.S. 1957. Measles Periodicity and Community Size. *Journal of the Royal Statistical Society* 120(1): 48–70.
- Beach, J. P., Ferrie, M.H. Saavedra. 2018. Fetal Shock or Selection? The 1918 Influenza Pandemic and Human Capital Development Brian. NBER Working Paper No. 24725.
- Black, F.L. 1996. Measles endemicity in insular populations: Critical community size and its evolutionary implication. *Journal of Theoretical Biology* 11(2): 207–211.
- Borradale, D., E. Isenring, E. Hacker., and M.G. Kimlin. 2014. Exposure to solar ultraviolet radiation is associated with a decreased folate status in women of childbearing age. *Journal of Photochemistry and Photobiology B: Biology* 131(5): 90–95.

- Branda, R. F., and J.W. Eaton. 1978. Skin color and nutrient photolysis: an evolutionary hypothesis. *Science* 201(4356): 625–626.
- Brown, R, and D. Thomas. 2018. On the Long Term Effects of the 1918 U.S. Influenza Pandemic. Working paper available at <https://clas.ucdenver.edu/ryan-brown/working-papers> (last accessed July 6, 2020).
- Bruce, D, J..H Ooi, S. Yu, and M.T. Cantorna. 2010. Vitamin D and host resistance to infection? Putting the cart in front of the horse. *Experimental Biology and Medicine* 235(8): 921–927.
- CDC. 2017a. Influenza-Like Illness (ILI) Activity Level Indicator Determined by Data Reported to ILINet <https://gis.cdc.gov/grasp/fluview/main.html> (last accessed July 6, 2020).
- CDC 2017b. National, Regional, and State Level Outpatient Illness and Viral Surveillance. <https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html> (last accessed July 6, 2020).
- CDC 2017c. Overview of Influenza Surveillance in the United States. <https://www.cdc.gov/flu/weekly/overview.htm> (last accessed July 6, 2020).
- CDC 2018. Transcript for CDC Update on Flu Activity. <https://www.cdc.gov/media/releases/2018/t0126-flu-update-activity.html> (last accessed July 6, 2020).
- CDC 2019. Behavioral Risk Factor Surveillance System [https://www.cdc.gov/brfss/annual\\_data/annual\\_data.htm](https://www.cdc.gov/brfss/annual_data/annual_data.htm) (last accessed July 6, 2020).
- Census 2010. US Population Weighted Center. [https://www2.census.gov/geo/docs/reference/cenpop2010/county/CenPop2010\\_Mean\\_CO.txt](https://www2.census.gov/geo/docs/reference/cenpop2010/county/CenPop2010_Mean_CO.txt) (last accessed July 6, 2020).
- Charland, K.M.L., D.L. Buckeridge, J.L. Sturtevant, F. Melton, By. Reis, K.D. Mandl, and J.S. Brownstein. 2009. Effect of environmental factors on the spatio-temporal patterns of influenza spread. *Epidemiol. Infect.* 137:1377–1387.
- Clay, K., J. Lewis, E Severnini. 2018. Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic. *Journal of Economic History* 78(4): 1179–1209.
- Clay, K., J. Lewis, E Severnini. 2019. What explains cross-city variation in mortality during the 1918 influenza pandemic? Evidence from 438 U.S. cities. *Economics & Human Biology* 35: 42–50.
- Cohn, B.A. 2002. Sunlight, skin color, and folic acid. *Journal of the American Academy of Dermatology* 46(2):317–318.
- Conway, K.S. and J Trudeau. 2019. Sunshine, fertility and racial disparities. *Economics & Human Biology* 32: 18–39.

- Deschenes, O. 2013. Temperature, human health, and adaptation: A review of the empirical literature. *Energy Economics*, 46(C): 609–619.
- Duarte, F., S. Kadiyala, S.H. Masters, and D.Powell. 2017. The Effect of the 2009 Influenza Pandemic on Absence from Work. *Health Economics*, 26(12): 1682–1695.
- Fletcher, J. 2018. New Evidence on the Impacts of Early Exposure to the 1918 Influenza Pandemic on Old-Age Mortality: A Research Note. *Biodemography and Social Biology* 64(2): 123–126.
- Grant, W.B., and E. Giovannucci. 2009. The possible roles of solar ultraviolet-B radiation and vitamin D in reducing case-fatality rates from the 1918–1919 influenza pandemic in the United States. *Dermato-Endocrinology*, 1(4):215–219.
- Holick, M.F. 2007. Vitamin D deficiency. *New England Journal of Medicine*, 357(3): 266–281.
- Huetal, G., N.H. Miller, and D. Molitor. 2020. Adaptation and the Mortality Effects of Temperature Across U.S. Climate Regions. Forthcoming, *Review of Economics and Statistics*.
- Jones, G. 2008. Pharmacokinetics of vitamin D toxicity. *Am J Clin Nutr*, 88(2):582S–586S.
- Jones, P., Lucock, M., Veysey, M., and Beckett, E. 2018. The Vitamin D-Folate Hypothesis as an Evolutionary Model for Skin Pigmentation: An Update and Integration of Current Ideas. *Nutrients*, 10(5), 554.
- Khare, D, N.M. Godbole, S.D. Pawar, V. Mohan, G. Pandey, S. Gupta, D. Kumar, T. Dhole, and M.M. Godbole. 2013. Calcitriol [1, 25[OH]<sub>2</sub> D<sub>3</sub>] pre- and post-treatment suppresses inflammatory response to influenza A (H1N1) infection in human lung A549 epithelial cells. *European Journal of Nutrition*. 52(4): 1405–1415.
- Kwong, J.C., K.L. Schwartz, M.A. Campitelli, H. Chung, N.S. Crowcroft, T. Karnauchow, K. Katz, D.T. Ko, A.J. McGeer, D. McNally, D.C. Richardson, L.C. Rosella, A. Simor, M. Smieja, G. Zahariadis, and J.B. Gubbay. 2018. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. *New England Journal of Medicine*, 378(4): 345–353.
- Lin, M.J. and E.M. Liu. 2014. Does in utero exposure to Illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment. *Journal of Health Economics*, 37: 152–163.
- Martineau, A.R., D.A. Jolliffe, R.L. Hooper, L. Greenberg, J.F. Aloia, P. Bergman, G. Dubnov-Raz, S. Esposito, D. Ganmaa, A.A. Ginde, E.C. Goodall, C.C. Grant, C.J. Griffiths, W. Janssens, I. Laaksi, S. Manaseki-Holland, D. Mauger, D.R. Murdoch, R. Neale, J.R. Rees, S. Simpson, I. Stelmach, G.T. Kumar, M. Urashima, C.A. Camargo. 2017. Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. *BMJ* 356 :i6583.

- Momplaisir, F., Frank, I., Meyer, W., Kim, D., Kappes, R., & Tebas, P. 2012. Vitamin D Levels, Natural H1N1 Infection and Response to H1N1 Vaccine among HIV-Infected Individuals. *Journal of AIDS & Clinical Research*, 3(4): 152.
- Maurer, J. 2009. Who has a clue to preventing the flu? Unravelling supply and demand effects on the take-up of influenza vaccinations. *Journal of Health Economics*, 28: 704–717.
- Mawer, E.B., K. Schaefer, G.A. Lumb, and S.W. Stanbury. 1971. The metabolism of isotopically labelled vitamin D3 in man: the influence of the state of vitamin D nutrition. *Clin Sci*, 40: 39–53.
- NREL. 2018. National Solar Radiation Database. <https://nsrdb.nrel.gov/> (last accessed July 6, 2020).
- NOAA. 2017. Federal Climate Complex Global Surface Summary of Day Data. <ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt> (last accessed July 6, 2020).
- New York State Department of Health. 2015. Statewide Planning and Research Cooperative System (SPARCS) <https://www.health.ny.gov/statistics/sparcs/> (last accessed July 6, 2020).
- New York State Department of Health. 2016. NY State Health Profiles. <https://profiles.health.ny.gov/hospital/alpha> (last accessed July 6, 2020).
- Pichler, S., and N.R. Ziebarth. 2016. The Pros and Cons of Sick Pay Schemes: Testing for Contagious Presenteeism and Noncontagious Absenteeism Behavior. *Journal of Public Economics* 156: 14–33.
- Pichler, S., K. Wen, and N.R. Ziebarth. 2020. Positive Health Externalities of Mandating Paid Sick Leave. Available at [https://www.researchgate.net/publication/336832189\\_Positive\\_Health\\_Externalities\\_of\\_Mandating\\_Paid\\_Sick\\_Leave](https://www.researchgate.net/publication/336832189_Positive_Health_Externalities_of_Mandating_Paid_Sick_Leave) (last accessed July 6, 2020).
- Polozov, I. V., L. Bezrukov, K. Gawrisch and J. Zimmerberg. 2008. Progressive ordering with decreasing temperature of the phospholipids of influenza virus. *Nature Chemical Biology* 4(4): 248-255.
- Sagripani, J.L. and C.D. Lytle. 2007. Inactivation of Influenza Virus by Solar Radiation. *Photochemistry and Photobiology* 83: 1278–1282.
- Schanzer, D.L., and Schwartz, B., 2013. Impact of Seasonal and Pandemic Influenza on Emergency Department Visits, 2003–2010, Ontario, Canada. *Acad Emerg Med* 20: 388–397.
- Schwandt, H. 2017. The Lasting Legacy of Seasonal Influenza: In-utero Exposure and Labor Market Outcomes. IZA Working Paper 10589.
- Soebitantyo, R.P., D. Gross, P. Jorgensen, S. Buda, M. Bromberg, Z. Kaufman, K. Prosenc7, M.



- Socan, T.V. Alonso, M.A. Widdowson, R.K. Kiang. 2015. Associations between Meteorological Parameters and Influenza Activity in Berlin (Germany), Ljubljana (Slovenia), Castile and León (Spain) and Israeli Districts. *PLOSOne* 10(8):e0134701.
- Sun, H. et al. 2020. Prevalent Eurasian avian-like H1N1 swine influenza virus with 2009 pandemic viral genes facilitating human infection. *Forthcoming in PNAS*.
- Trudeau, J., K.S. Conway, A. K. Menclova. 2016. Soaking Up the Sun: The Role of Sunshine in the Production of Infant Health. *American Journal of Health Economics* 2016 2(1): 1–40.
- Urashima M, Mezawa H, Noya M, Camargo CA Jr. 2014. Effects of vitamin D supplements on influenza A illness during the 2009 H1N1 pandemic: a randomized controlled trial. *Food Funct.* 5(9):2365–70.
- Viboud, C., and S.L. Epstein. 2016. First flu is forever. *Science* 354 (6313): 706–707.
- Wernerfelt, N., D.J.G. Slusky, and R.J. Zeckhauser. 2017. Second Trimester Sunlight and Asthma: Evidence from Two Independent Studies. *American Journal of Health Economics*, 3(2): 227–253.
- White, C. 2019. Measuring the Social and Externality Benefits of Influenza Vaccination. *Forthcoming in the Journal of Human Resources*.
- Zhang, X., Y. Wang, X. Chen, X. Zhang. 2020. Prenatal Sunshine Exposure and Birth Outcomes in China. *Science of The Total Environment*, 713: 136472.
- ZIP Code Database. 2013 <https://www.unitedstateszipcodes.org/zip-code-database/> (last accessed July 6, 2020).

## Appendix A: The CDC Flu Index Calculation

We use the weekly count of outpatient visits (both total and only those due to influenza) from the CDC (2017b) along with the documentation in the ILI data (CDC 2017c) to conduct calculations regarding the influenza index. That index corresponds to the number of standard deviations the share of outpatient visits that report influenza symptoms that week exceeds all non-influenza weeks. A “non-influenza week” is defined as a week in which it and its preceding week had fewer than 2% of all outpatient visits to healthcare providers indicating influenza.

Our study period is 2008-2016. Hence, we use the October 2005-September 2008 period as a “pre-period” to calibrate our index. We begin with the formal start of the season, which the CDC defines as week 40 (the first week of October). Unfortunately, whereas the ILI data (CDC 2017a) is available at the state level, the outpatient visit count data is only available nationally. Therefore, we conduct our calculations at that level.

Nationally, of the 156 weeks in October 2005-September 2008, 108 fit the above definition of “non-influenza.” The mean share for those 108 weeks is 1.03%, and their standard deviation is 0.39 percentage points.

Given this, the method for calculating the influenza index is now to take all weeks, calculate the z-score[s] (that is, number of standard deviations above or below the mean), and then apply the following index definition:

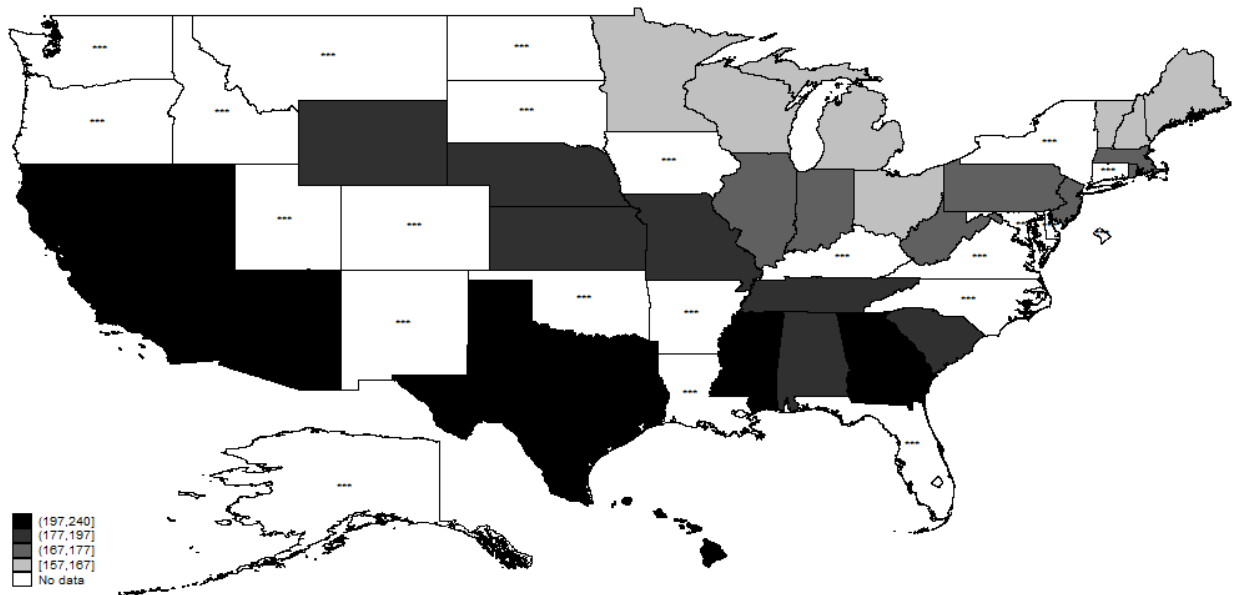
Flu index =

$$\begin{array}{lll} 1 & \text{if} & Z < 0 \\ \text{int}(Z) + 2 & \text{if} & 0 < Z < 8 \\ 10 & \text{if} & Z > 8 \end{array}$$

So, in the interior range of the index, we can consider an additional index point as representing an additional standard deviation.

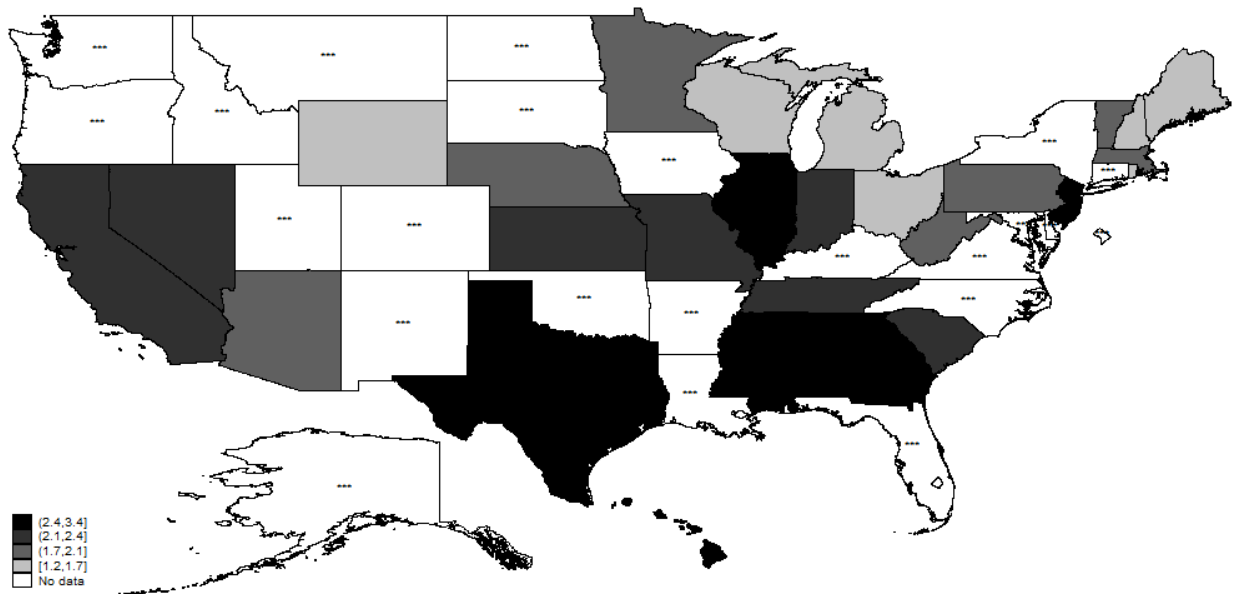
## Appendix B: Additional Figures

**Appendix Figure B1: Population-Weighted Geographic Sunlight Variation**



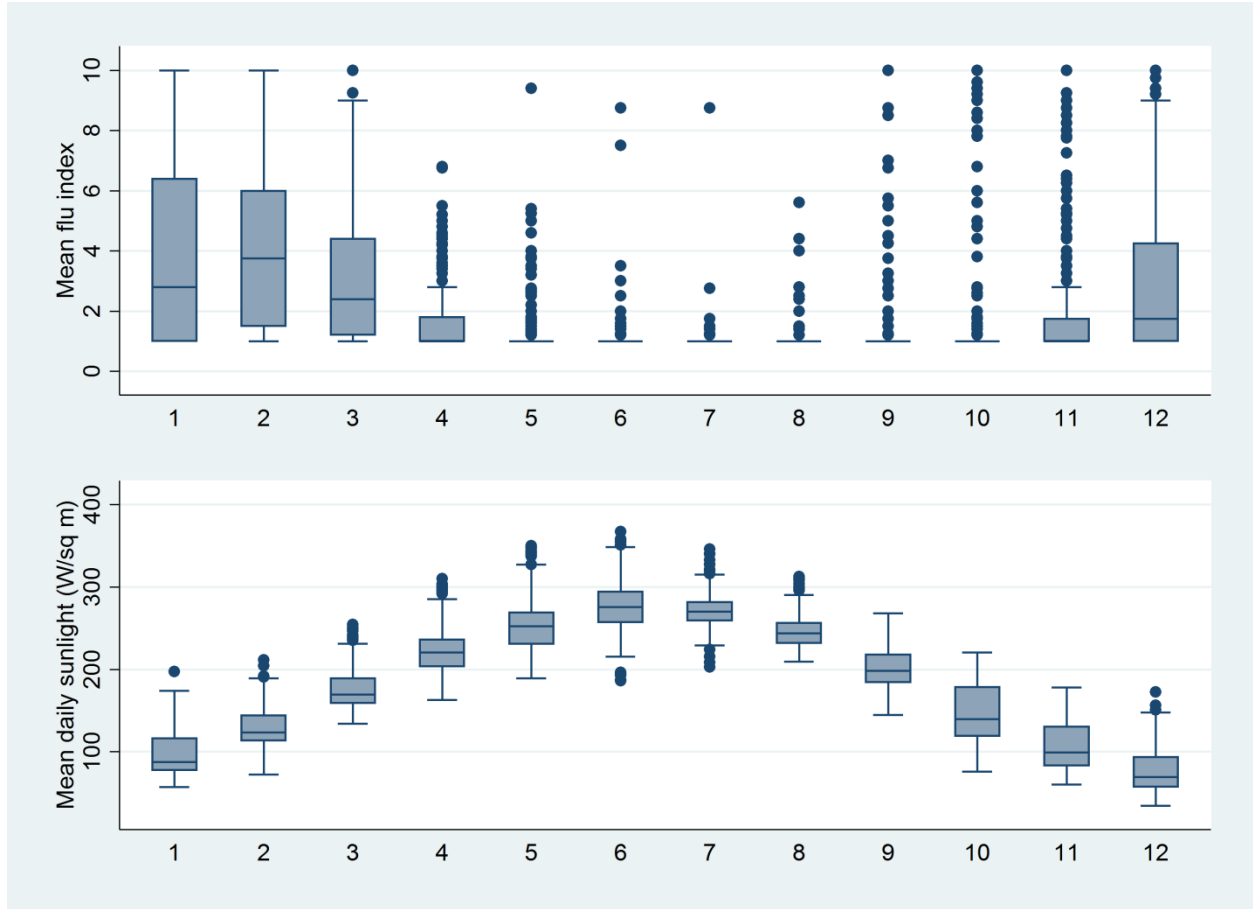
Notes: 3-year average (2009-2016) of daily county sunlight, weighted by county population. “No data” and “\*\*\*” refer to incomplete influenza data for that state.

**Appendix Figure B2: Geographic Flu Variation**



Notes: 3-year average (2009-2016) of weekly state-level flu index. “No data” and “\*\*\*” refer to incomplete influenza data for that state.

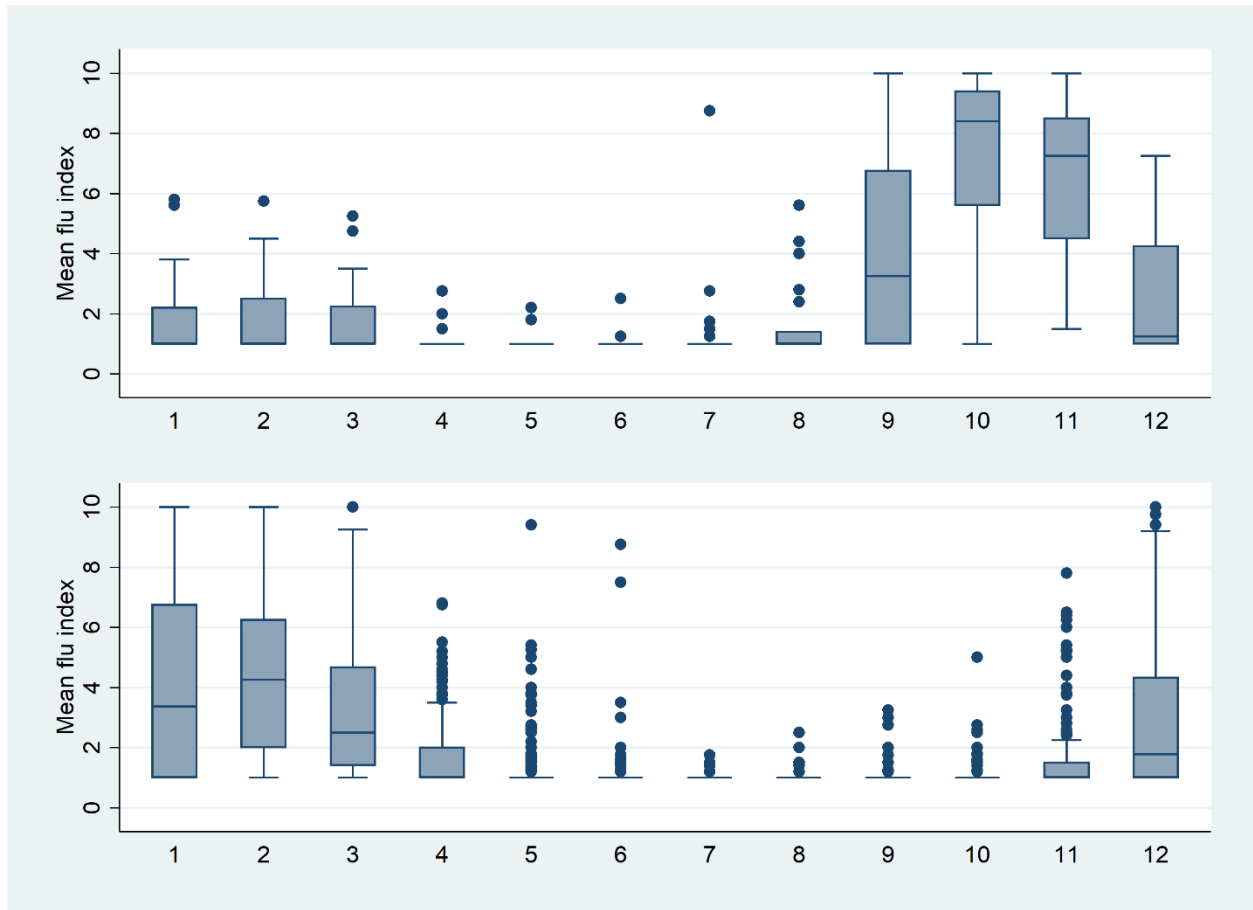
**Appendix Figure B3: Box Plots of Average Flu and Sunlight by Month**



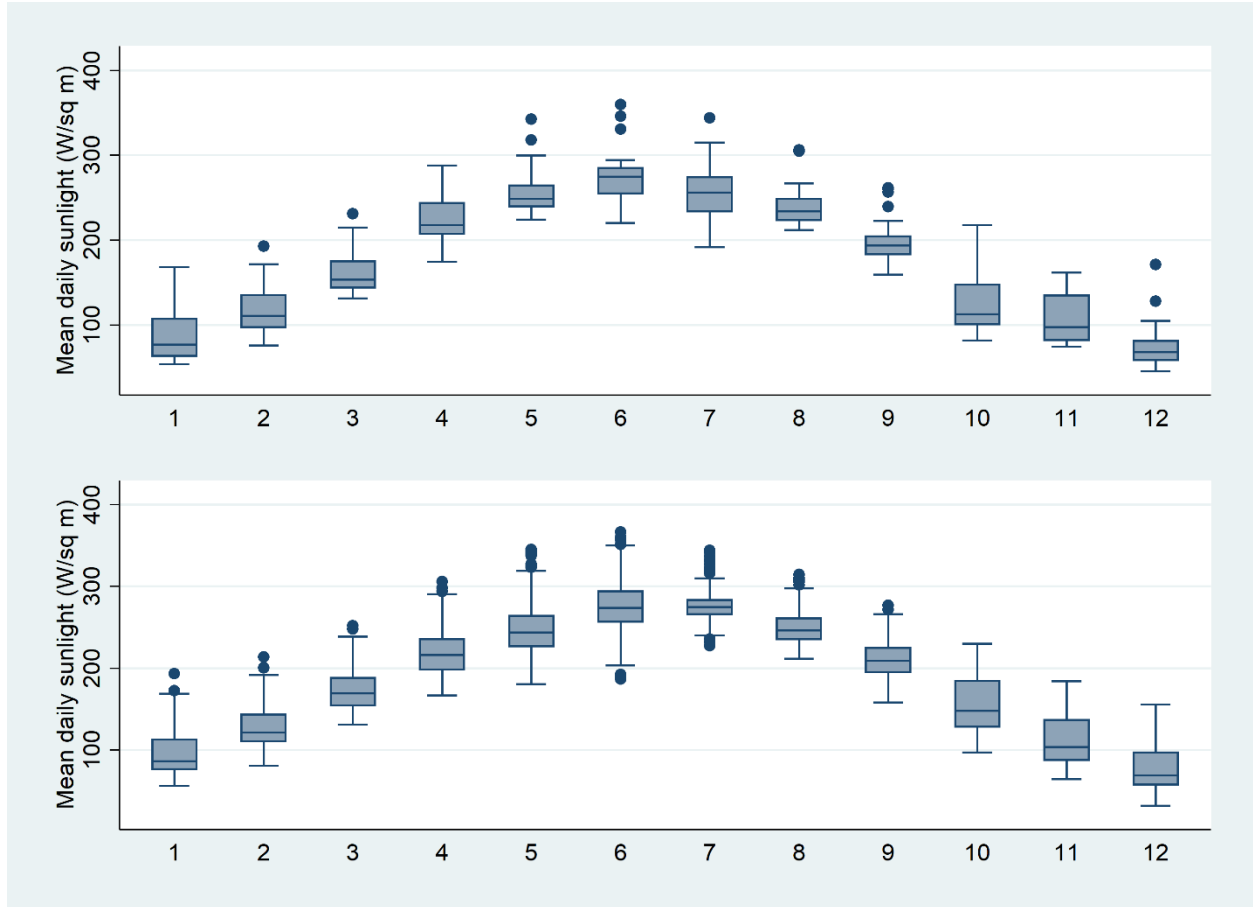
Notes: Covers the 28 contiguous states that have full flu and sunlight data. Outliers are shown in blue dots.

## Appendix Figure B4: Box Plots of Average Flu and Sunlight by Month

*Panel A: Flu, flu-year 2009 (top) and 2010-2016 (bottom)*



*Panel B: Sunlight, flu-year 2009 (top) and 2010-2016 (bottom)*

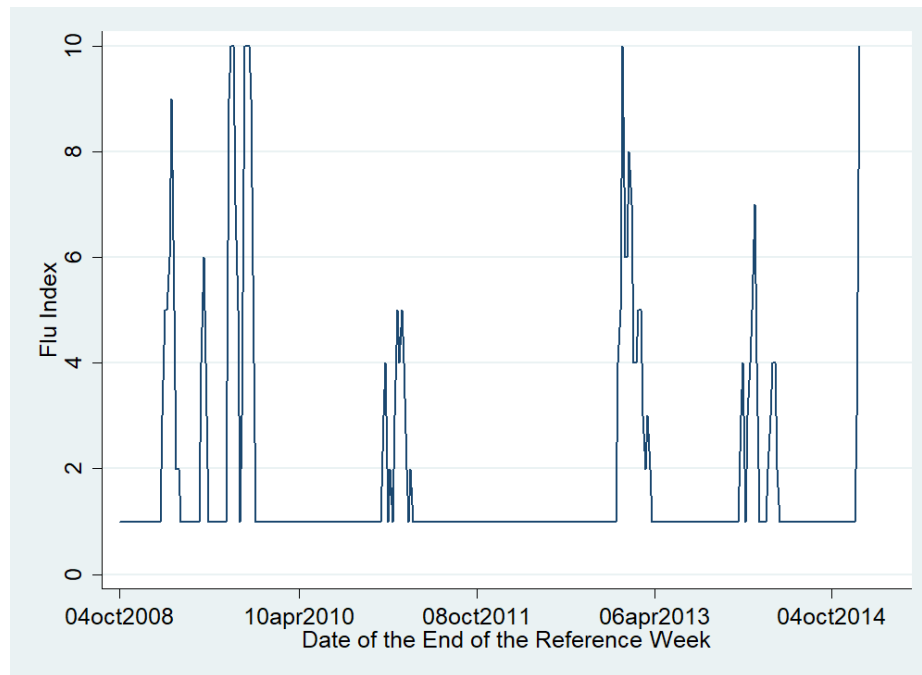


Notes: Covers the 28 contiguous states that have full flu and sunlight data. Outliers are shown in blue dots.

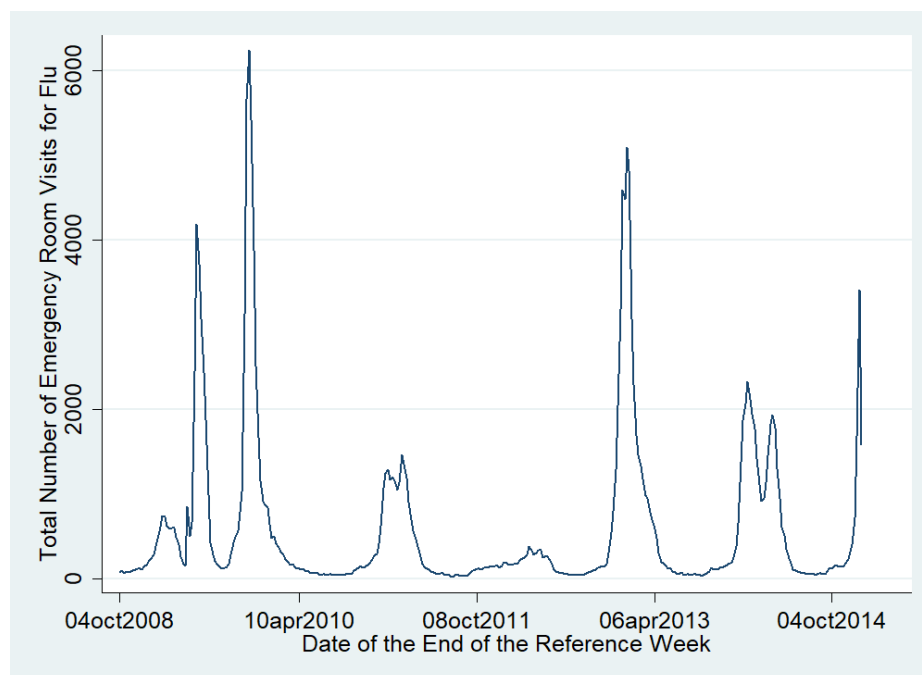
## Appendix Figure B5: Comparison of CDC and Discharge Flu Measures for New York

State

*Panel A: CDC Flu Index*



*Panel B: Hospital Emergency Room Visits for Flu*





**Appendix Table B1: Stratifying by Sun Level and by Deviation of Fall 2009 Sunlight**

	(1)	(2)	(3)	(4)	(5)	(6)
	Sunniness		Sunniness		Fall 2009 Log Sunlight Variation	
Years	Low 2008-2016	High 2008-2016	Low 2008, 2010-2016	High 2008, 2010-2016	Large 2008-2016	Small 2008-2016
Log sunlight for that month and the prior month	-6.031*** (2.145)	-6.658*** (1.277)	0.897 (0.912)	-0.131 (1.284)	-7.680*** (1.864)	-2.359 (1.863)
Observations	1,716	1,821	1,500	1,596	1,709	1,820
R-squared	0.566	0.841	0.042	0.032	0.813	0.602
Average Sunlight (W/m <sup>2</sup> )	161.9	197.5	161.9	197.5		
Average Sept/Oct 2009 Deviation of Log Sunlight from Mean					-0.121	-0.0389

Notes: September and October only. All regressions include state-week and flu year fixed effects and weather controls. All states are included, even those with missing weeks of flu data (hence the differences in sample sizes). Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Low sunniness states: Connecticut, Delaware, District of Columbia, Idaho, Illinois, Indiana, Iowa, Maine, Massachusetts, Michigan, Minnesota, Montana, New Hampshire, New Jersey, New York, North Dakota, Ohio, Oregon, Pennsylvania, Rhode Island, South Dakota, Vermont, Washington, West Virginia and Wisconsin.

High sunniness states: Alabama, Arizona, Arkansas, California, Colorado, Florida, Georgia, Hawaii, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nebraska, Nevada, New Mexico, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Utah, Virginia, and Wyoming.

Large Fall 2009 Log Sunlight Deviation from Mean: Alabama, Arkansas, Delaware, Georgia, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nebraska, New Jersey, New York, North Carolina, North Dakota, Ohio, Oklahoma, Pennsylvania, Tennessee, Texas, West Virginia.

Small Fall 2009 Log Sunlight Deviation from Mean: Arizona, California, Colorado, Connecticut, Florida, Hawaii, Idaho, Maine, Massachusetts, Michigan, Minnesota, Montana, Nevada, New Hampshire, New Mexico, Oregon, Rhode Island, South Carolina, South Dakota, Utah, Vermont, Virginia, Washington, Wisconsin, Wyoming.

**Appendix Table B2: Stratifying by Sun Level for Normal October-March Flu Season**

	(1)	(2)	(3)
	All	Low Sunniness	High Sunniness
Log sunlight for that month and the prior month	-2.241*** (0.682)	-2.325*** (0.704)	-2.024 (1.311)
Observations	10,977	5,409	5,568
R-squared	0.133	0.124	0.156
Average Sunlight (W/m <sup>2</sup> )	180.1	161.9	197.5

Notes: 2008-2016. October-March only. All regressions include state-week and flu year fixed effects and weather controls. All states are included, even those with missing weeks of flu data (hence the differences in sample sizes). Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Low sunniness states: Connecticut, Delaware, District of Columbia, Idaho, Illinois, Indiana, Iowa, Maine, Massachusetts, Michigan, Minnesota, Montana, New Hampshire, New Jersey, New York, North Dakota, Ohio, Oregon, Pennsylvania, Rhode Island, South Dakota, Vermont, Washington, West Virginia and Wisconsin.

High sunniness states: Alabama, Arizona, Arkansas, California, Colorado, Florida, Georgia, Hawaii, Kansas, Kentucky, Louisiana, Maryland, Mississippi, Missouri, Nebraska, Nevada, New Mexico, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Utah, Virginia, and Wyoming.

**Appendix Table B3: Stratifying New York Results by Race and Age**

	(1)	(2)	(3)	(4)	(5)
	All	White	Non-White	Age<65	Age>=65
Log sunlight for that month and the prior month	-0.997*** (0.245)	-0.847*** (0.211)	-0.151*** (0.0361)	-0.975*** (0.238)	-0.0227*** (0.00653)
Observations	3,139	3,139	3,139	3,139	3,139
R-squared	0.241	0.232	0.258	0.240	0.140

Notes: 2008-2014. September and October only. Per bed rate of influenza discharges. All regressions include county-week and year fixed effects and weather controls. Robust standard errors clustered at the county level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

## Appendix C: Additional Robustness Checks

**Appendix Table C1: Results Including States Missing Flu Data for Some Months**

	(1)	(2)	(3)	(4)	(5)	(6)non
States	All	All	Non-Missing	Non-Missing	Non-Missing	Non-Missing
Months	All	Sept & Oct	All	Sept & Oct		
Log sunlight for that month and the prior month	-2.203*** (0.520)	-6.663*** (1.309)				
Sunlight for that month and the prior month			-0.0182*** (0.00528)	-0.0419*** (0.00868)	-0.0345** (0.0142)	-0.249*** (0.0384)
(Sunlight for that month and the prior month) <sup>2</sup>					3.36e-05 (2.92e-05)	0.000494*** (8.87e-05)
F stat					8.247	31.83
P value					0.00160	7.90e-08
Observations	20,870	3,537	12,068	2,072	12,068	2,072
R-squared	0.071	0.701	0.080	0.651	0.090	0.675

Notes: 2008-2016. All regressions include state-week and flu year fixed effects and weather controls. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05,

\* p<0.1

**Appendix Table C2: Results Dropping One State at a Time**

State dropped	Log sunlight for that month and the prior month		R-squared
Alabama	-10.16***	(1.447)	0.65
Arizona	-10.87***	(1.48)	0.66
California	-11.28***	(1.379)	0.66
Georgia	-10.40***	(1.529)	0.653
Hawaii	-10.99***	(1.446)	0.689
Illinois	-10.43***	(1.466)	0.656
Indiana	-10.45***	(1.454)	0.654
Kansas	-10.65***	(1.48)	0.654
Maine	-9.903***	(1.344)	0.68
Massachusetts	-10.33***	(1.466)	0.674
Michigan	-10.50***	(1.495)	0.672
Minnesota	-10.48***	(1.485)	0.668
Mississippi	-10.82***	(1.501)	0.655
Missouri	-10.84***	(1.467)	0.652
Nebraska	-10.32***	(1.453)	0.66
Nevada	-10.72***	(1.456)	0.658
New Hampshire	-10.21***	(1.458)	0.678
New Jersey	-10.78***	(1.451)	0.672
Ohio	-10.48***	(1.469)	0.67
Pennsylvania	-10.63***	(1.49)	0.657
Rhode Island	-10.55***	(1.495)	0.671
South Carolina	-10.62***	(1.49)	0.654
Tennessee	-10.61***	(1.498)	0.651
Texas	-10.04***	(1.375)	0.654
Vermont	-10.99***	(1.423)	0.678
West Virginia	-10.63***	(1.483)	0.664
Wisconsin	-10.69***	(1.471)	0.656
Wyoming	-10.56***	(1.54)	0.66

Notes: N=1998. 2008-2016. September and October only. All regressions include state-week and flu year fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C3: Results By Days of the Week**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log sunlight for that month and the prior month, only for:							
Sunday	-1.831* (0.978)						
Monday		-3.401*** (0.624)					
Tuesday			-5.964*** (0.829)				
Wednesday				-1.327** (0.562)			
Thursday					-3.233*** (0.564)		
Friday						-2.242*** (0.765)	
Saturday							-1.284** (0.558)
Observations	2,064	2,072	2,069	2,070	2,067	2,070	2,072
R-squared	0.636	0.647	0.665	0.627	0.635	0.636	0.629

Notes: 2008-2016. September and October only. All regressions include state-week and flu year fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C4: Monthly Level Results**

Months States	(1) All All	(2) Sept & Oct All	(3) All Non- Missing	(4) Sept & Oct Non- Missing
Log sunlight for that month and the prior month, one year earlier	-2.573*** (0.520)	-9.278*** (1.577)	-2.145*** (0.774)	-13.76*** (1.781)
Observations	4,760	760	2,744	448
R-squared	0.080	0.755	0.081	0.737

Notes: 2008-2016. All regressions include state-month and flu year fixed effects and weather controls. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C5: Results Controlling for Other Weather Measures**

	(1)	(2)	(3)	(4)	(5)
Log sunlight for that month and the prior month	-7.017*** (1.295)	-7.527*** (1.145)	-9.400*** (1.074)	-8.321*** (1.053)	-8.878*** (1.447)
Controls (past two months):					
Log temperature	X				
Days per month below 15°F		X			
Log specific humidity			X		
Days per month specific humidity is below 6 g/kg				X	
Log precipitation					X
Observations	2,072	2,072	2,072	2,072	2,072
R-squared	0.645	0.651	0.654	0.653	0.648

Notes: 2008-2016. All regressions include state-week and flu year fixed effects. September and October only. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C6: Dummy Variables for Ventiles of Weather Controls**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-1.824** (0.779)	-1.631** (0.775)	-10.56*** (1.438)	-9.156*** (1.455)
Observations	12,068	12,068	2,072	2,072
R-squared	0.074	0.103	0.663	0.679
Number of group	1,484	1,484	252	252
Weather Controls	Standard	Ventiles	Standard	Ventiles
Months	All	All	Sep & Oct	Sep & Oct

Notes: 2008-2016. All regressions include state-week and flu year fixed effects. September and October only. The panel consists of the 28 states that have complete flu and sunlight data. Ventiles are specific subsample in column. Robust standard errors clustered at the state level in parentheses.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C7: Levels of Precipitation Instead of Logarithms**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-1.824** (0.779)	-1.829** (0.770)	-10.56*** (1.438)	-10.44*** (1.470)
Observations	12,068	12,068	2,072	2,072
R-squared	0.074	0.073	0.663	0.663
Number of group	1,484	1,484	252	252
Precipitation Control	Log	Linear	Log	Linear
Months	All	All	Sep & Oct	Sep & Oct

Notes: 2008-2016. All regressions include state-week and flu year fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C8: Controlling for Flu Match Rate**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-1.824** (0.779)	-3.194*** (0.719)	-10.56*** (1.438)	-20.20*** (2.723)
Observations	12,068	12,068	2,072	2,072
R-squared	0.074	0.025	0.663	0.437
Number of group	1,484	1,484	252	252
Flu Year Controls	X		X	
Match Rate Control		X		X
Months	All	All	Sep & Oct	Sep & Oct

Notes: 2008-2016. All regressions include state-week fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Appendix Table C9: Controlling for Cumulative Flu Vaccination Rate**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-10.18*** (1.439)	-10.53*** (1.436)	-10.33*** (1.414)	-10.33*** (1.414)
Observations	2,053	2,072	2,072	2,072
R-squared	0.652	0.663	0.664	0.664
Number of group	252	252	252	252
Vaccination Control	Raw Rate	Raw + Interpolated	Moving Average	Season Ratchet Rate

Notes: 2008-2016. September and October only. All regressions include state-week and flu year fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



**Appendix Table C10: Results for Different Past Sunlight Windows**

	(1)	(2)	(3)	(4)	(5)	(6)
Log sunlight for the past:						
Two weeks	-3.330*** (0.595)					
Three weeks		-5.977*** (0.787)				
Four weeks			-7.802*** (0.849)			
Five weeks				-9.518*** (1.009)		
Six weeks					-10.56*** (1.122)	
Seven weeks						-10.91*** (1.281)
Observations	2,068	2,072	2,072	2,072	2,072	2,072
R-squared	0.649	0.673	0.681	0.683	0.680	0.671

Notes: 2008-2016. September and October only. All regressions include state-week and flu year fixed effects and weather controls for the same period as the sunlight variable (e.g., the past two weeks). The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

## Appendix D: Tract Level Analysis

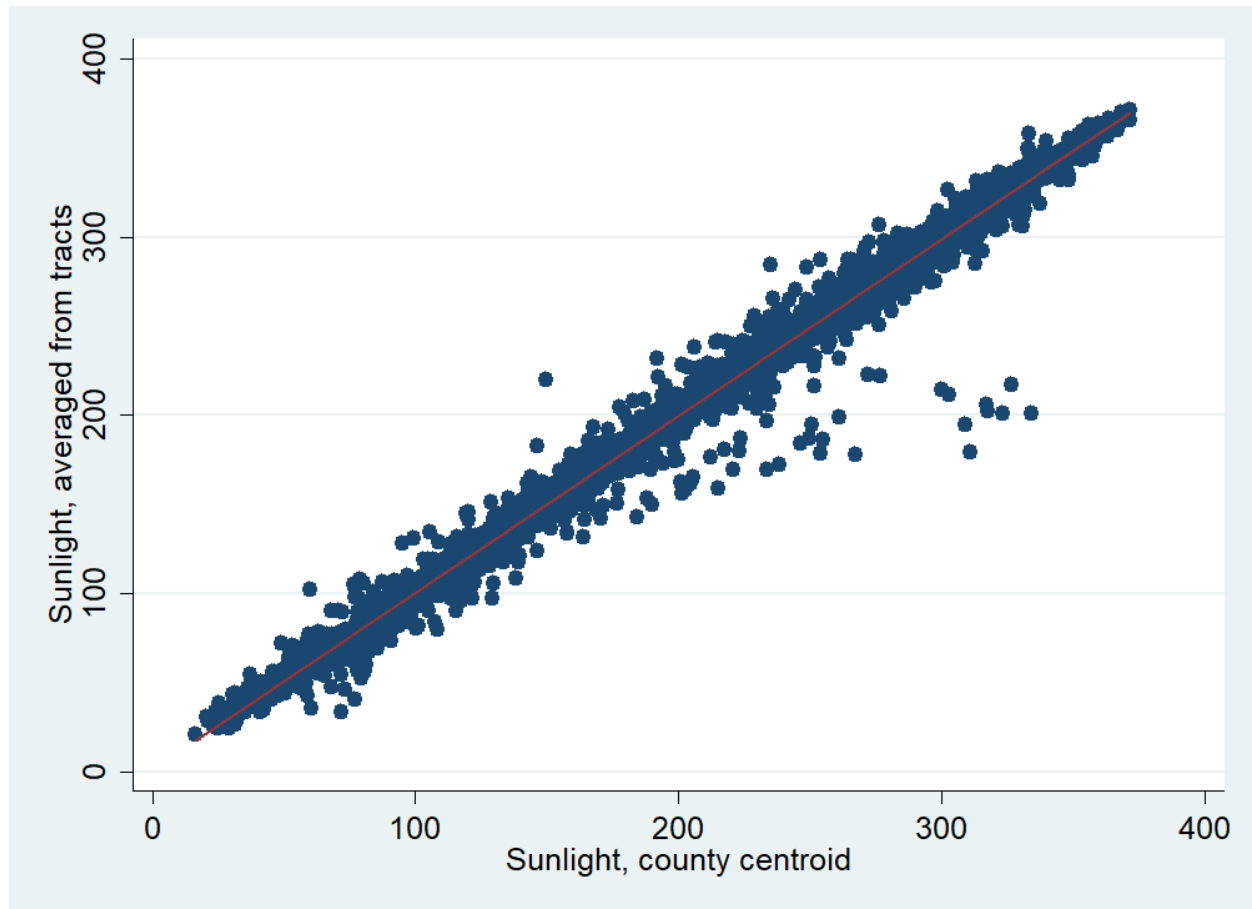
The state-week level sunlight data used in this paper is calculated as a population weighted average of the sunlight at the coordinates of the centroid for each county in each state. This is purely for tractability, as sunlight and weather data (excluding precipitation, which is from another source) must be downloaded for each year for individual set of coordinates, and one can only download 2000 files per IP address per day. With over 3000 counties in the U.S. and several years of data (including the placebo years used in the analysis below) it took over a month to just download the main data for the paper.

One may be worried, though, that for large counties using the county centroid misses intra-county variation in sunlight. There are 74,000 census tracts in the U.S (Census 2000) and so it is infeasible to download sunlight data for all of their centroids for all years.

In lieu of that, this appendix instead downloads sunlight and weather data for each census tract in the 100 largest counties by land area in the U.S. (not including Alaska), only for calendar year 2009 (which is the year that is driving our results). These counties range from San Bernardino in California (20,000 square miles) to Kane County in Utah (4000 square miles), across 16 states (Arizona, California, Colorado, Hawaii, Idaho, Maine, Minnesota, Montana, Nebraska, Nevada, New Mexico, Oregon, Texas, Utah, Washington, and Wyoming). Together they have 7200 census tracts.

This data is then aggregated up to the county level using the same method as the county to state aggregation in the rest of the paper: a population weighted average. Appendix Figure D1 shows a scatter plot of this average sunlight based on tract data vs. sunlight the county centroid for this 100 counties.

**Appendix Figure D1: County Sunlight Averaged from Tracts vs. County Centroid**

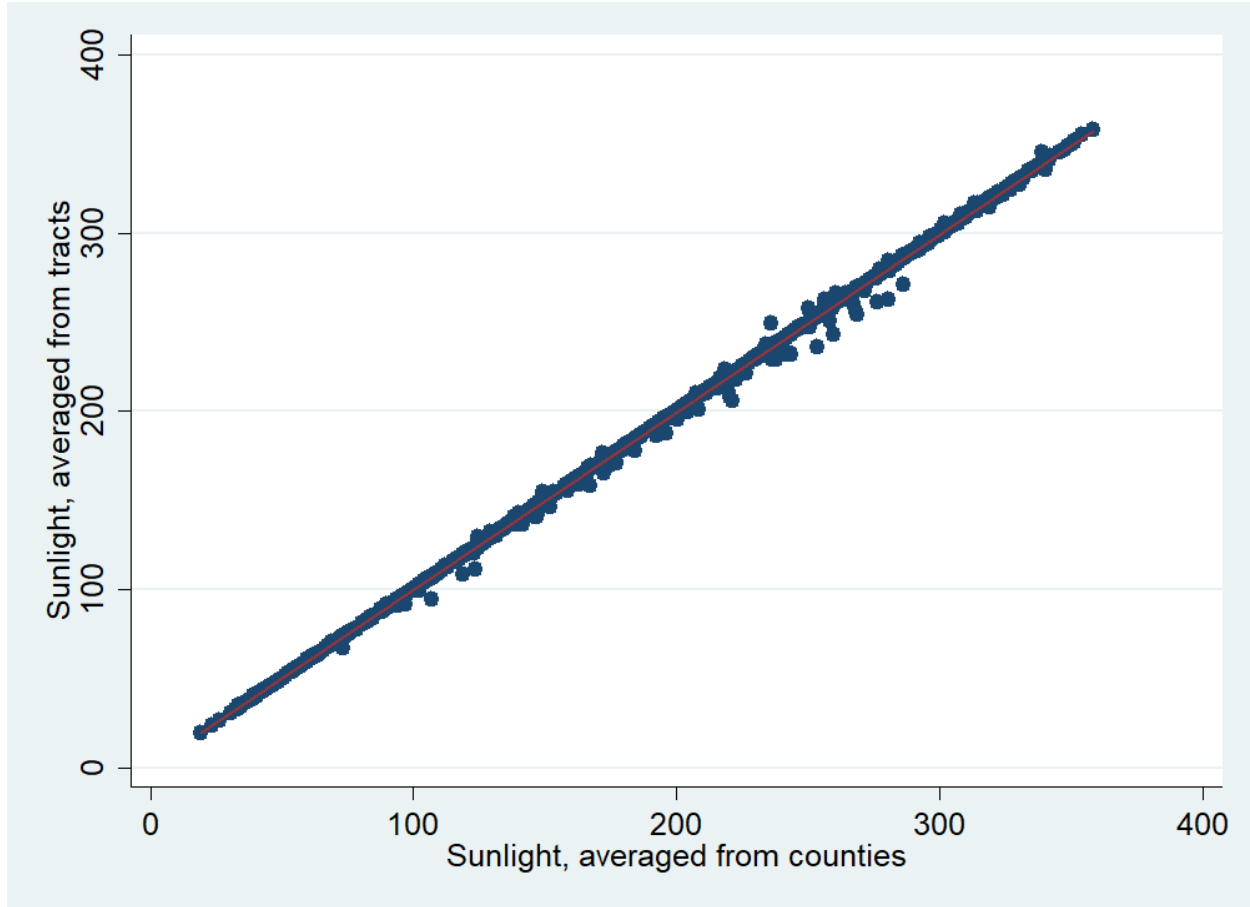


Notes: 100 largest counties in U.S., which are in Arizona, California, Colorado, Hawaii, Idaho, Maine, Minnesota, Montana, Nebraska, Nevada, New Mexico, Oregon, Texas, Utah, Washington, and Wyoming. Red line is best linear fit. 2009 only.

While there are a small number of outliers, the vast majority of points in this graph lie near the best fit line.

Then this county level data is combined with the main data for the rest of the counties in each state and aggregated up the state level. Appendix Figure D2 shows the state level averages, with our normal data on the x-axis and the new data incorporating tract level sunlight on the y-axis.

**Appendix Figure D2: State Sunlight Averaged from Tracts vs. Averaged from Counties**



Notes: Arizona, California, Colorado, Hawaii, Idaho, Maine, Minnesota, Montana, Nebraska, Nevada, New Mexico, Oregon, Texas, Utah, Washington, and Wyoming. 2009. Red line is line of best fit.

This data is very close to the 45 degree line; the correlation is 0.9996

We then repeat the main results of our paper, using the sunlight and weather data from the tract centroids for the 100 largest counties. (Given that the precipitation data is from weather stations, we use the same precipitation as the main analysis.)

**Appendix Table D1: Main Results Using Census Tract Level Sunlight for the Largest Counties**

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-2.132*** (0.509)	-1.739** (0.758)	-6.297*** (1.399)	-10.14*** (1.689)
Observations	20,870	12,068	3,537	2,072
R-squared	0.070	0.073	0.699	0.660
States	All	Non-Missing	All	Non-Missing
Months	All	All	Sep & Oct	Sep & Oct

Notes: 2008-2016. September and October only. All regressions include state-week and flu year fixed effects and weather controls. The panel consists of the 28 states that have complete flu and sunlight data. Robust standard errors clustered at the state level in parentheses.

Here we see results that are almost identical to our main results. Column (4) shows the coefficient (standard error) of our main September/October result are -10.14\*\*\* (1.69), compared to -10.56\*\*\* (1.44) from Table 3 above where the county centroid data was used. So it makes our point estimate 4% smaller and our t-stat remains overwhelmingly significant at 6.0 rather than 7.3.

## References

- Census 2000. United States -- County by State, and for Puerto Rico. GCT-PH1. Population, Housing Units, Area, and Density.  
[https://web.archive.org/web/20090826043629/http://factfinder.census.gov/servlet/GCTTable?\\_bm=y&-ds\\_name=DEC\\_2000\\_SF1\\_U&-CONTEXT=gct&-mt\\_name=DEC\\_2000\\_SF1\\_U\\_GCTPH1\\_US9&-redoLog=false&-caller=geoselect&-geo\\_id=&-format=US-25|US-25S&-lang=en](https://web.archive.org/web/20090826043629/http://factfinder.census.gov/servlet/GCTTable?_bm=y&-ds_name=DEC_2000_SF1_U&-CONTEXT=gct&-mt_name=DEC_2000_SF1_U_GCTPH1_US9&-redoLog=false&-caller=geoselect&-geo_id=&-format=US-25|US-25S&-lang=en) (last accessed July 6, 2020).
- Census 2010a. US Population Weighted Center (Counties).  
[https://www2.census.gov/geo/docs/reference/cenpop2010/county/CenPop2010\\_Mean\\_CO.txt](https://www2.census.gov/geo/docs/reference/cenpop2010/county/CenPop2010_Mean_CO.txt) (last accessed July 6, 2020).
- Census 2010b. US Population Weighted Center. (Tracts)  
[https://www2.census.gov/geo/docs/reference/cenpop2010/tract/CenPop2010\\_Mean\\_TR.txt](https://www2.census.gov/geo/docs/reference/cenpop2010/tract/CenPop2010_Mean_TR.txt) (last accessed July 6, 2020).

## Appendix E: Mortality in 2009 and in the 1918 pandemic

Given the strong relationship between relative levels of influenza cases and sunlight in 2009, one would think that sunlight might also affect mortality rates per case. To study this question, we conduct an analysis for mortality data for recent years. We also repeat our analysis using historical mortality data for the 1918 influenza pandemic.<sup>30</sup>

The methodology for the mortality analysis follows the format above with a few necessary modifications. First, whereas we have CDC influenza index data back to October of 2008, mortality data specific to influenza and pneumonia are only available back to October of 2009. We therefore re-estimate our main results using the same period to show that they are consistent.

Second, working with the early 20<sup>th</sup> century data imposed a number of limitations. The mortality data is at the annual level, and not the month level. Given that 85% of the pandemic mortality occurred between October 1918 and January 1919 (Almond 2006), and based on our other results, we estimate the impact of either August-September or August-December sunlight from a given year on the mortality for that year.

Finally, we have both sunlight and mortality data for very few states before 1918. Thus, we average across all of the stations in a given state and year, and only keeping states that have data for 1918. This provides a balanced panel of 29 states for the years 1918 to 1936.<sup>31</sup>

For mortality data, we use CDC data on both the absolute number of individuals who die from influenza as well as pneumonia in each state and week and the share they comprise of total

---

<sup>30</sup> We know of no data on cases of influenza for the early 20<sup>th</sup> century period, and so unfortunately cannot also repeat our cases-level analysis.

<sup>31</sup> California, Colorado, Connecticut, Illinois, Indiana, Kansas, Kentucky, Louisiana, Massachusetts, Maryland, Maine, Michigan, Minnesota, Missouri, Montana, North Carolina, New Jersey, New York, Ohio, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Utah, Virginia, Vermont, Washington, and Wisconsin.

deaths. Both indicators are commonly reported by the CDC,<sup>32</sup> as some influenza deaths, are associated with secondary complications of influenza, including pneumonia, and so are recorded as such on the death certificate.<sup>33</sup> Unlike the influenza index described above, the CDC does not convert these numbers into a 10-point index (CDC 2019).

Our mortality data for 1900-1936, by state, year, and cause of death, comes from Cutler and Miller (2005) and Anderson, Charles, and Rees (2018). Both sets of authors provided identical data. That data only covers 10 states in 1900, but it gradually increases to 30 states in 1918 and 48 in 1936. Our sunlight data for 1891-1987 for 240 weather stations at the year and month level, is from Steurer and Karl (2012).

#### *Results: Mortality In the Modern Period*

First, consider a figure analogous to Figure 1, but now for the percentage of deaths from pneumonia or influenza.

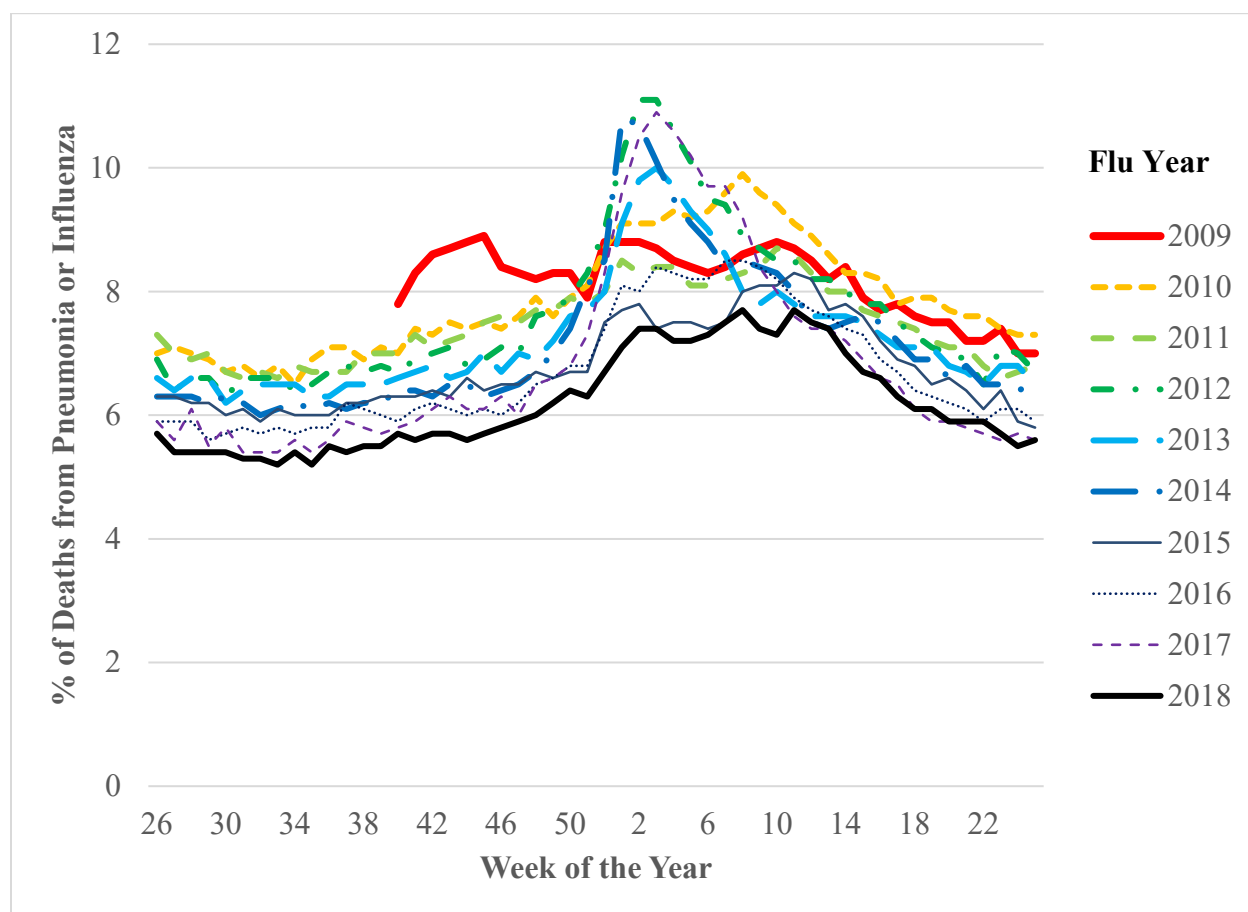
---

<sup>32</sup> E.g., see <https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm>

<sup>33</sup> <https://www.cdc.gov/flu/about/burden/faq.htm>

## Appendix Figure E1: Weekly % of National Deaths from Influenza & Pneumonia

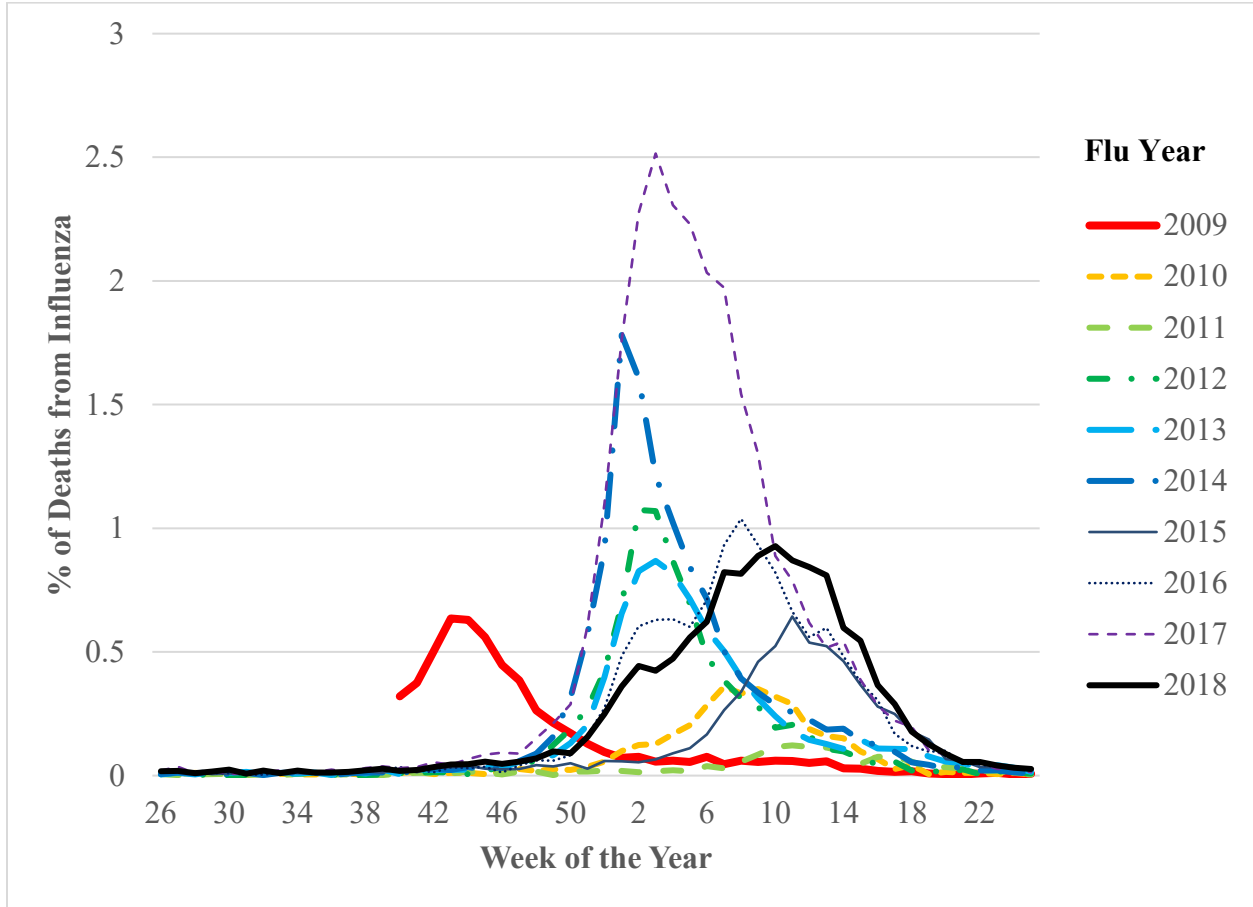
*Panel A: From Pneumonia or Influenza*



Notes: October 2009-2018. Whereas we have CDC influenza index data back to October of 2008, mortality data specific to influenza and pneumonia is only available back to October of 2009



*Panel B: From Influenza*



Notes: October 2009-2018. Whereas we have CDC influenza index data back to October of 2008, we only have mortality data back to October of 2009

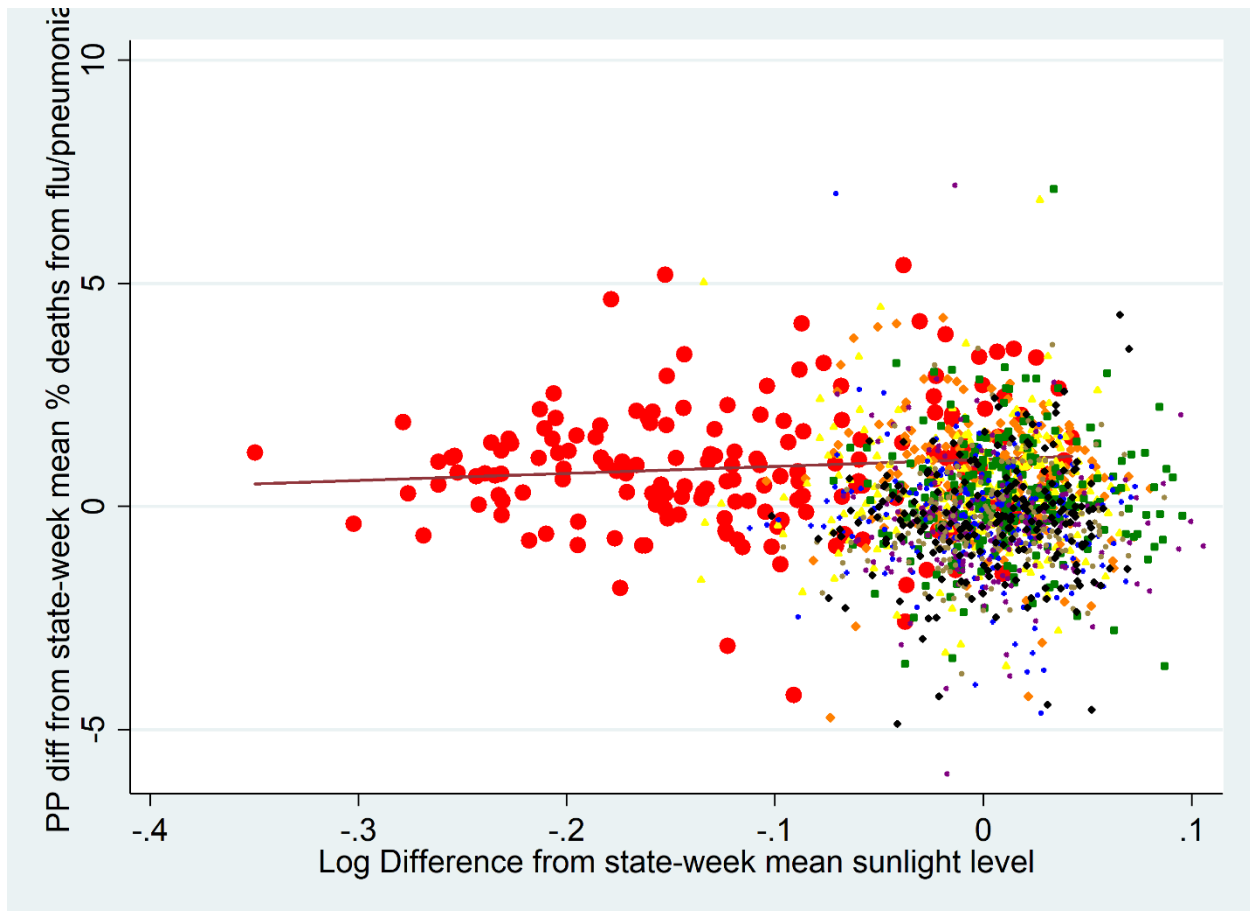
Similarly, in Panel B, we see that 2009 was also only an outlier in timing and not in magnitude for deaths only from influenza.

We can also make a figure analogous to Figure 2 where we plot the log difference in sunlight from the state-month mean for sunlight and the percentage point difference in the share of deaths from the state-month mean. Note that due to the mortality data starting only in October

of 2009, these graphs are only for mortality data for the month of October for the years 2009-2016, and the corresponding sunlight data in September-October of those years.<sup>34</sup>

## Appendix Figure E2: State-Week Deviations for Mortality and Sunlight, October

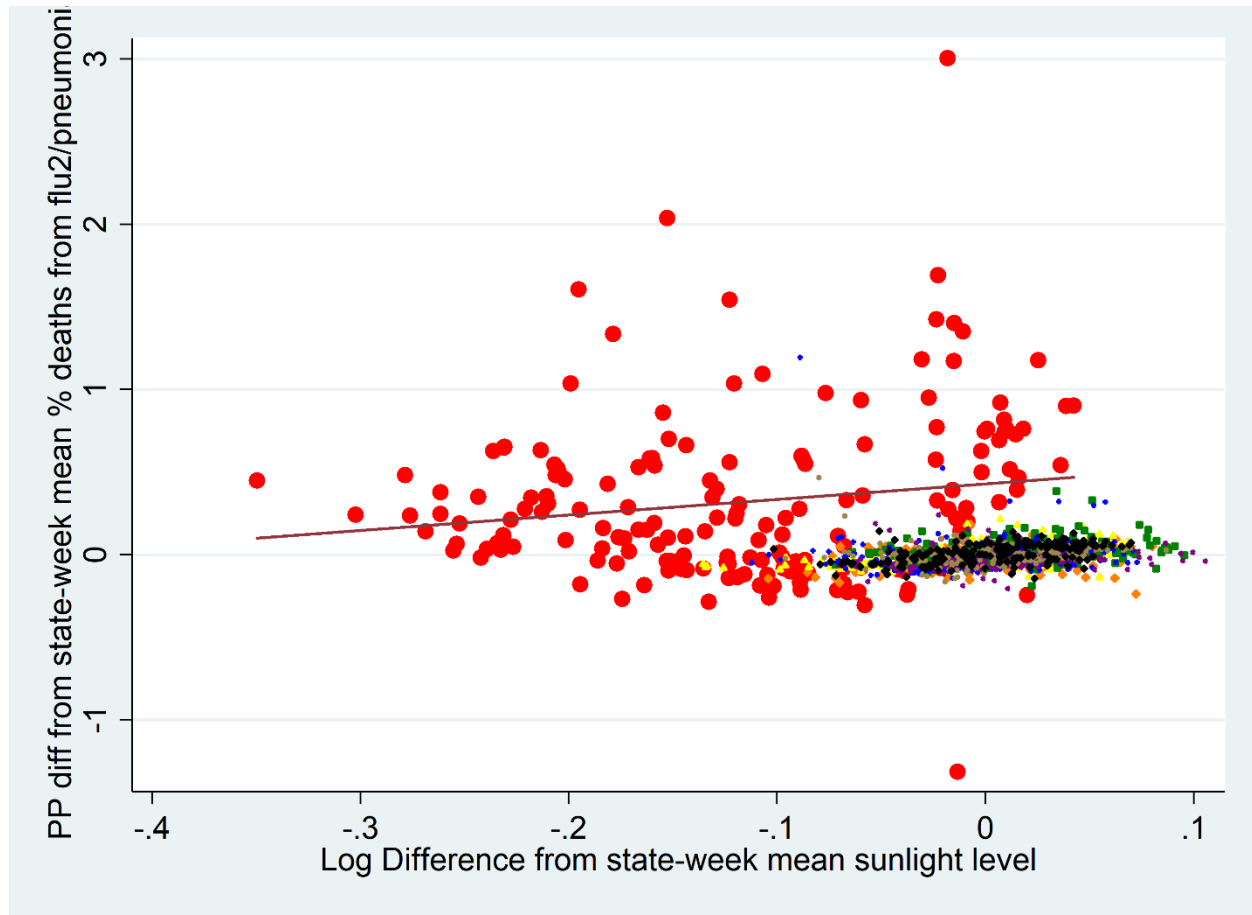
*Panel A: From Pneumonia or Influenza*



Notes: Red Circles = 2009; Orange Diamonds = 2010, Yellow Triangles = 2011, Green Squares = 2012, Blue Pluses = 2013, Purple X's = 2014, Brown Small Circles = 2015, Black Small Diamonds = 2016. Line is linear best fit for 2009. Vertical axis is the residual after regressing difference in state-week mean flu index on difference in state-week mean weather controls.

<sup>34</sup> Appendix Figure B2 shows Figure 2 for influenza cases for only the month of October (and therefore September-October sunlight), and demonstrates that the main relationship found in this paper is still present.

*Panel B: From Influenza*



Notes: Red Circles = 2009; Orange Diamonds = 2010, Yellow Triangles = 2011, Green Squares = 2012, Blue Pluses = 2013, Purple X's = 2014, Brown Small Circles = 2015, Black Small Diamonds = 2016. Line is linear best fit for 2009. Vertical axis is the residual after regressing difference in state-week mean flu index on difference in state-week mean weather controls.

Unlike in Figure 2, above, here we do not see a negative relationship, and conceivably even a weak positive relationship between relative levels of sunlight and mortality.

Appendix Table E1 then repeats our above analysis for the mortality data. Here we use October mortality data for 2009-2016, and September mortality data for 2010-2016.

**Appendix Table E1: Results of Sunlight on Flu Mortality, September and October**

	(1) Cases (Flu Index)	(2) % of Deaths from Flu & Pneumonia	(3) % of Deaths from Flu & Pneumonia	(4) % of Deaths from Flu	(5) % of Deaths from Flu
Log sunlight for that month and the prior month	-6.589*** (1.454)	0.316 (1.002)	-0.430 (1.153)	0.271 (0.197)	-0.180 (0.158)
Observations	1,848	2,860	1,820	2,860	1,820
R-squared	0.762	0.154	0.163	0.493	0.519
States	Non-Missing cases data (28)	All with mortality data	Non-Missing cases data (28)	All with mortality data	Non-Missing cases data (28)

Notes: All regressions include state-week and flu year fixed effects and weather controls. September (2010-2016) and October only (2009-2016). Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

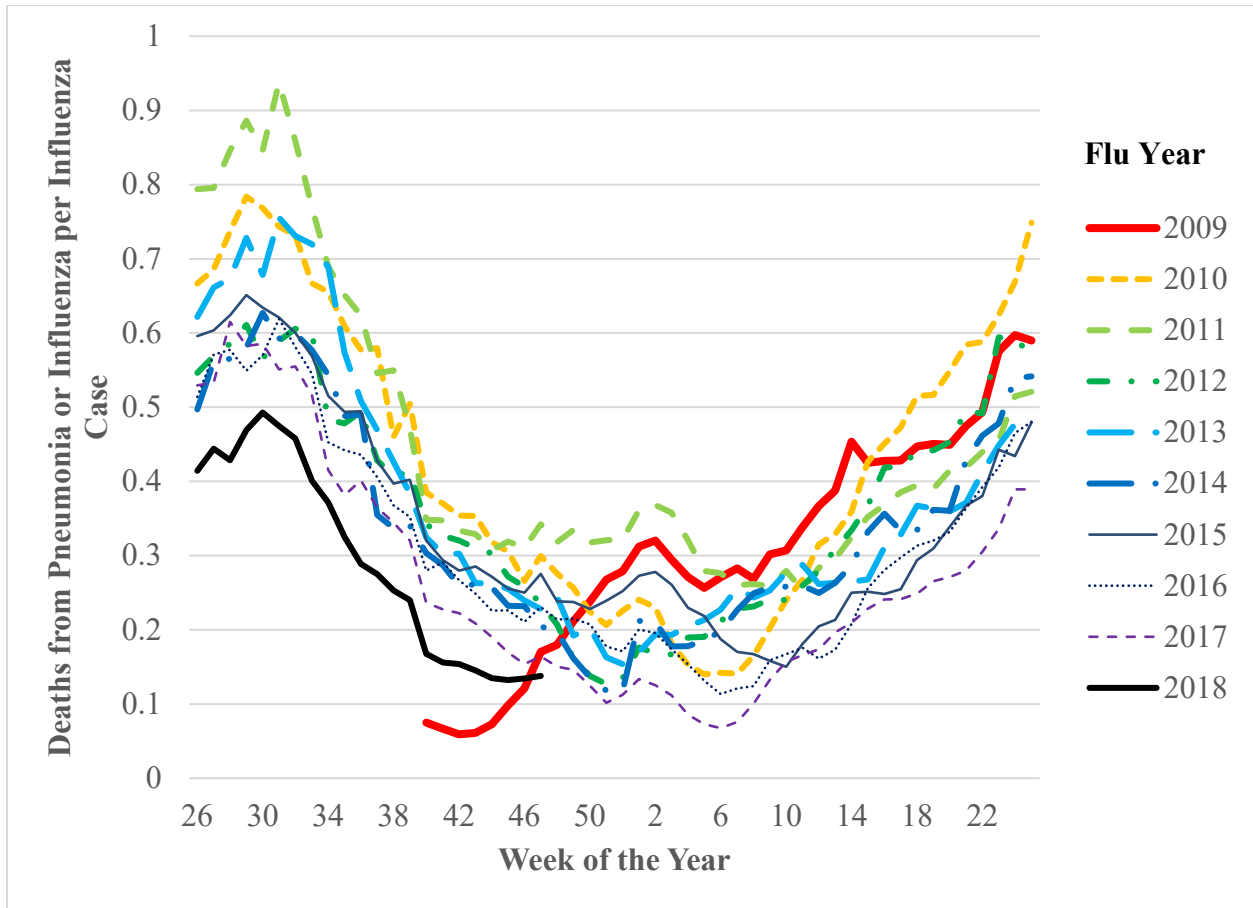
Column 1 shows our main result from Column 6 of Appendix Table C5, estimated for the same months as the mortality regressions. While we lose an order of magnitude when dropping September 2009, our result is still large and statistically significant.

Columns 2-5 show small and statistically insignificant coefficients for the effects on the share of deaths from influenza. This is true regardless of whether one only uses states with non-missing data or all states, or either outcome measure.

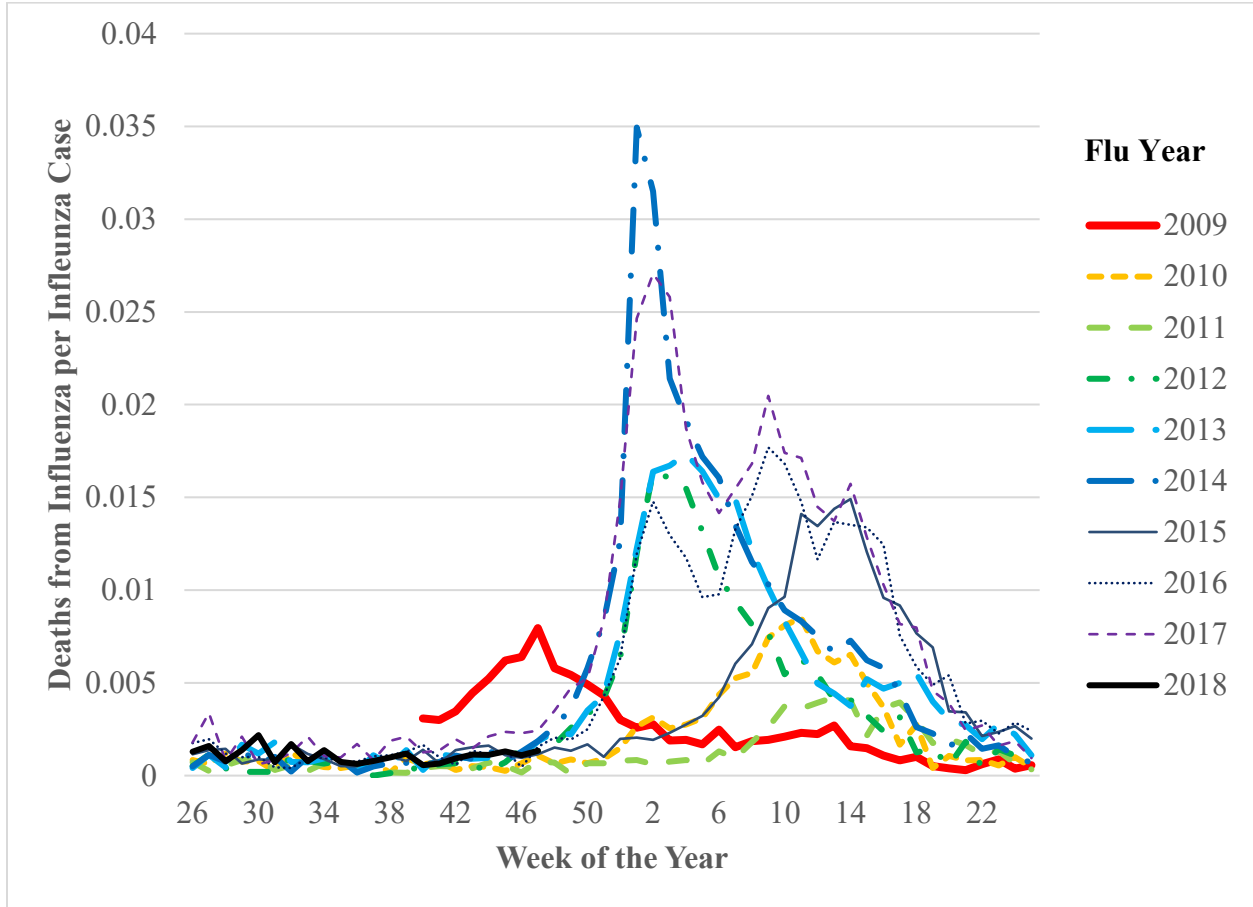
To better understand this null result, we combine the two panels of Appendix Figure E1 with Figure 1 to calculate deaths per case by week and year. The panel A shows deaths for pneumonia or influenza, where Panel B shows only deaths from influenza. The denominator is the same in both graphs, namely cases of influenza.

### Appendix Figure E3: Deaths per Influenza Case

*Panel A: Deaths from Pneumonia or Influenza*



*Panel B: Deaths from Influenza*



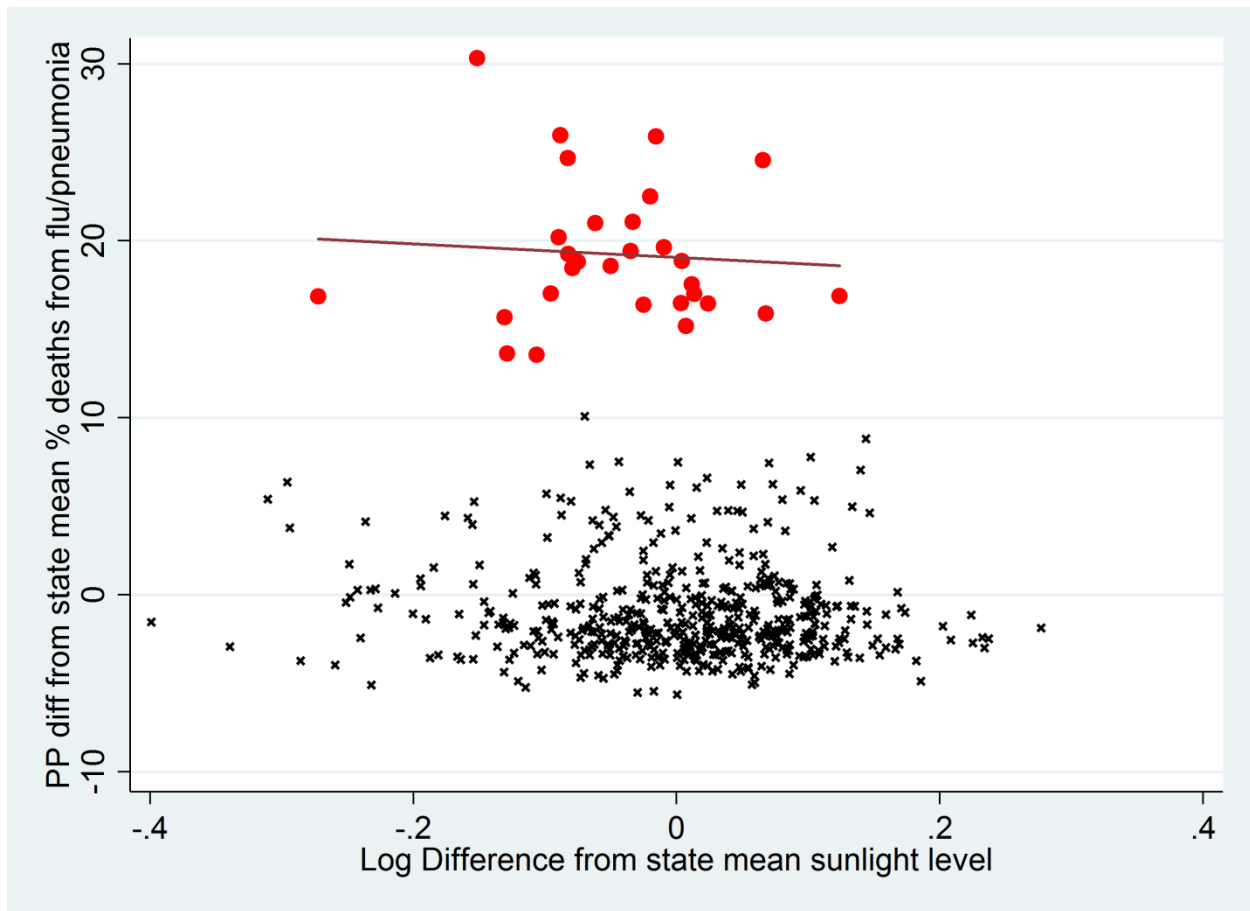
These figures help us solve the puzzle of why sunlight and cases are so closely related in the fall of 2009 but sunlight and mortality are not. The answer is that mortality per case is in general much lower in the fall than it is in the winter. The cases for 2009 were overwhelmingly in the fall.

#### *Historical Mortality, Starting with 1918*

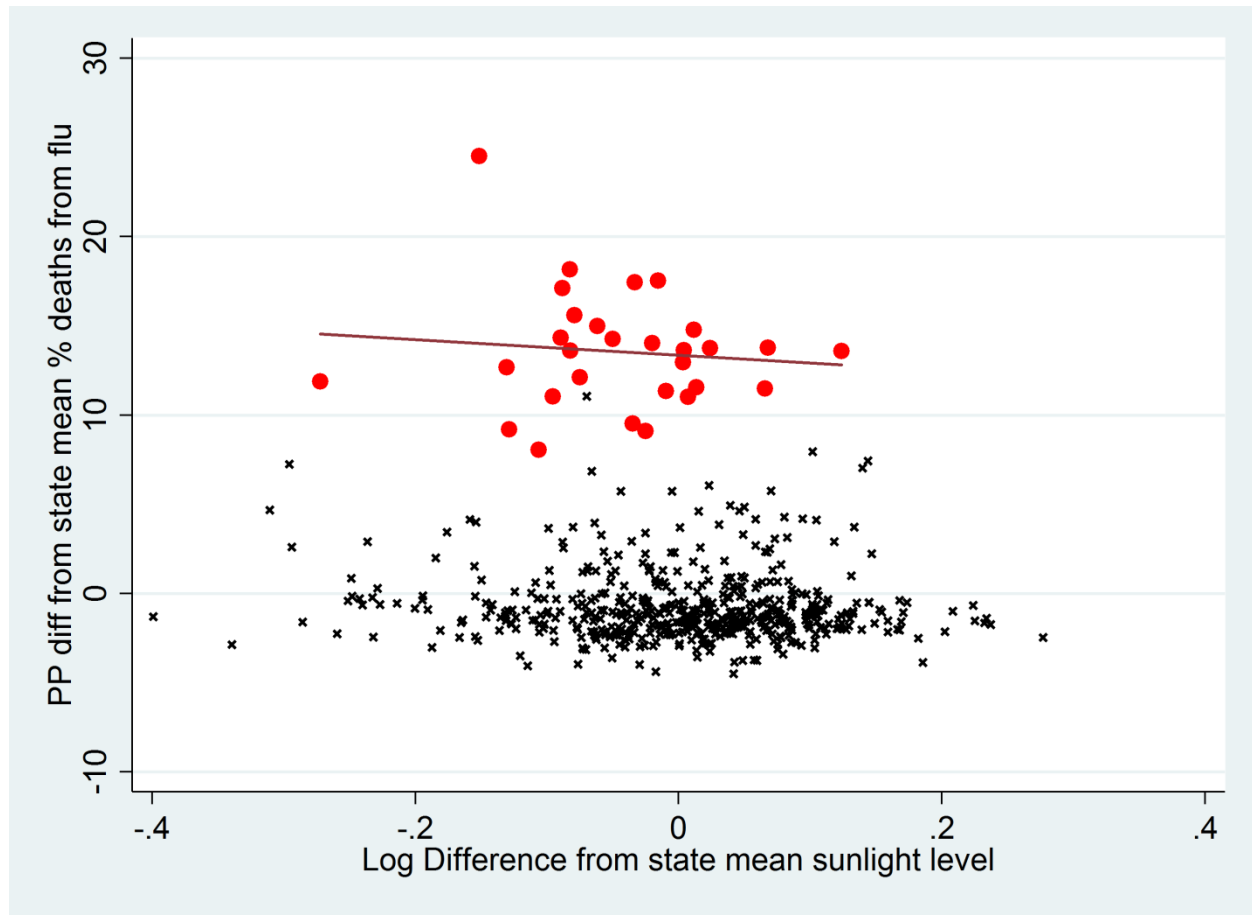
We now turn to the mortality analysis for the 1918 influenza pandemic. We first create figures analogous to Figure 2 and Appendix Figure E2, albeit at the month level. Given that there are 19 years of data, we color all years other than 1918 the same for clarity.

**Appendix Figure E4: State-Year Deviations for Mortality and Sunlight, August-September  
Sunlight**

*Panel A: From Pneumonia or Influenza*



*Panel B: From Influenza*



Notes: Red Circles = 1918; Black x's: 1919-1936. Line is linear best fit for 1918.

First, is it clear from both pictures that the 1918 flu represents a large deviation from the average. In that year, the share of deaths from influenza was 15-20 percentage points higher than subsequent years.

Appendix Table E2 shows analogous null regression results. This is the case regardless of whether sunlight for only August-September or for August-December is used.



**Appendix Table E2: Results of Sunlight on Flu Mortality, 1918-1936**

	(1) % of deaths from flu	(2)	(3) % of deaths from flu & pneumonia	(4)
Log sunlight for August and September	0.462 (0.654)		1.229 (0.762)	
Log sunlight for August-December		-0.577 (0.757)		0.376 (0.864)
Observations	551	551	551	551
R-squared	0.911	0.911	0.932	0.932

Notes: All regressions include state and year fixed effects. Robust standard errors clustered at the state level in parentheses. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ .

As shown above, we find no impact of relative sunlight on mortality rates. This would suggest that the fewer cases we find came from the most healthy individuals (who were unlikely to die from influenza) and not from the least healthy (who were most likely to die) or equally from the entire distribution.

## References

- Almond, D. 2006. Is the 1918 Influenza Pandemic Over? Long-Term Effects of In Utero Influenza Exposure in the Post-1940 U.S. Population. *Journal of Political Economy*, 114 (4): 672–712.
- Anderson, D.M., K.K. Charles, and D.I. Rees. 2018. Public Health Efforts and the Decline in Urban Mortality. NBER Working Paper No. 25027.
- CDC 2019. Pneumonia and Influenza Mortality Surveillance from the National Center for Health Statistics Mortality Surveillance System. <https://gis.cdc.gov/grasp/fluview/mortality.html> (last accessed July 6, 2020).
- Cutler, D. and G. Miller. 2005. The role of public health improvements in health advances: the twentieth-century United States. *Demography*, 42(1):1-22. Data available at <https://www.nber.org/data/vital-statistics-deaths-historical/> (last accessed July 6, 2020).
- Steurer, P.M., and T.R. Karl. 2012. Historical Sunshine and Cloud Data in the United States (revised 1991). Available at <https://cdiac.ess-dive.lbl.gov/ndps/ndp021.html> and <https://www.osti.gov/biblio/1394922> (both last accessed July 6, 2020).

## Appendix F: Placebo Tests Utilizing Data from Other Years

To provide an additional check on the robustness of our results, we conduct placebo tests using sunlight data from other years to predict influenza rates for a particular year. Appendix Table F1 parallels Table 3. It adds lagged sunlight and weather controls for years before the treatment period and leads of sunlight and weather controls for years after treatment period to provide those tests. Our primary coefficient retains its statistical significance, even when multiple other independent variables are included. Moreover, adding these variables only provides a modest upward nudge to the R-square values.

**Appendix Table F1: Retrospective Placebo Results for September and October Flu**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Log sunlight for that month and the prior month											
Treatment year	-16.48*** (2.063)	-17.89*** (1.820)	-17.38*** (1.895)	-14.71*** (2.578)	-13.91*** (2.633)	-14.84*** (2.532)	-16.49*** (2.311)	-14.76*** (2.249)	-11.38*** (2.192)	-12.86*** (2.924)	-15.35*** (2.379)
Year -1		-4.938*** (1.414)	-4.356** (1.776)	-1.856 (1.912)	-2.524 (2.490)	-1.120 (2.621)				-0.308 (3.506)	-4.082 (2.782)
Year -2			-1.902 (2.234)	0.0675 (2.245)	-1.186 (2.839)	0.598 (2.799)				1.243 (3.245)	-1.416 (3.034)
Year -3				2.818 (2.315)	2.237 (2.538)	3.445 (2.780)				3.565 (2.750)	2.555 (2.666)
Year -4					1.532 (2.946)	2.999 (3.155)				-0.452 (2.734)	3.144 (2.836)
Year -5						3.381 (3.034)				5.703** (2.529)	3.317 (2.568)
Year +1							10.69*** (2.518)	8.034*** (2.671)	9.335*** (2.573)	10.44*** (2.597)	
Year +2								-3.046 (3.980)	-1.396 (3.030)	-4.411 (3.428)	-8.064** (3.242)
Year +3									-0.455 (3.342)	-2.308 (3.651)	-5.814 (3.633)
Observations	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120	1,120
R-squared	0.716	0.746	0.754	0.763	0.768	0.774	0.739	0.772	0.786	0.822	0.808

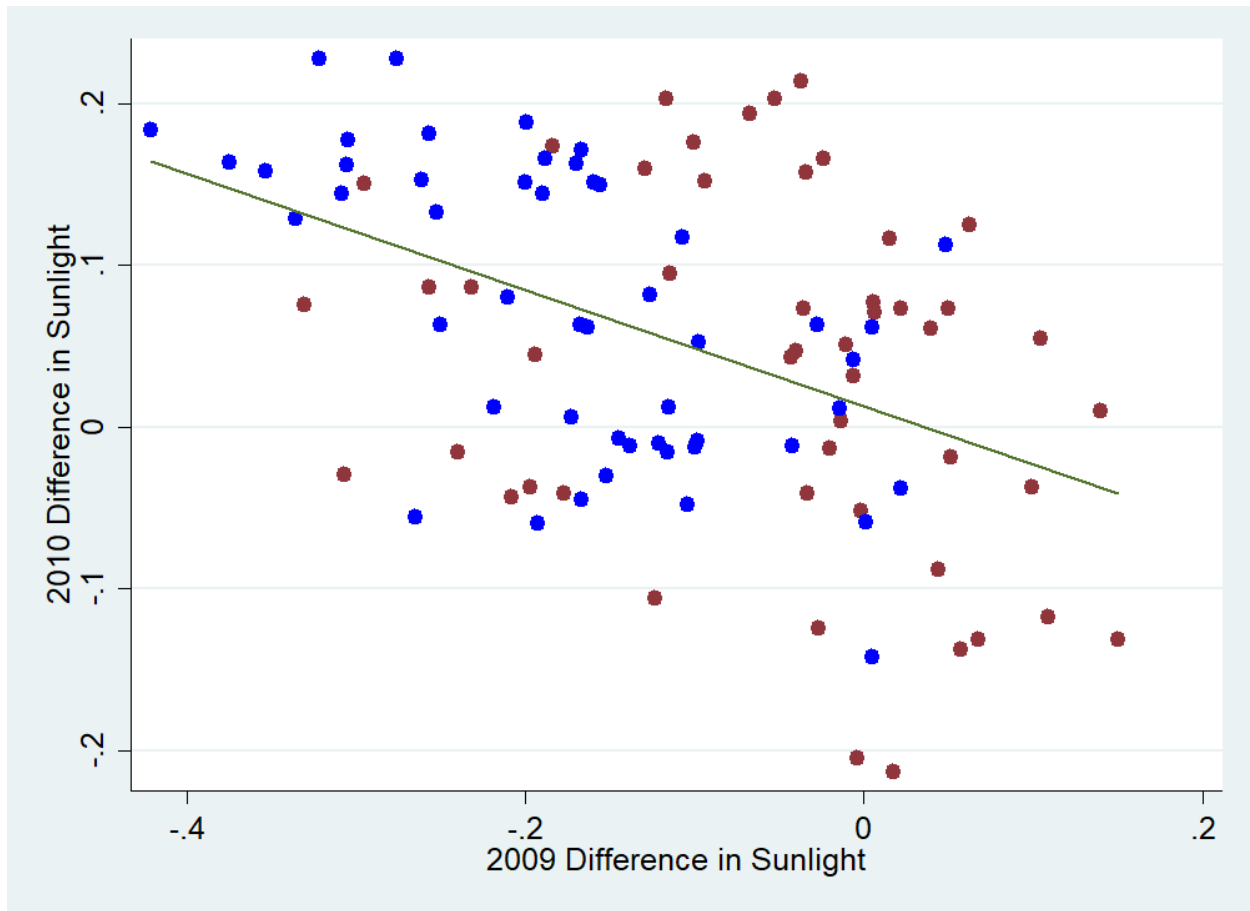
Notes: Flu years 2008-2012. All regressions include state-week and flu year fixed effects and weather controls. September and October only. Robust standard errors clustered at the state level in parentheses. The panel consists of the 28 states that have complete flu and sunlight data. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

This placebo test does identify statistically significant coefficients on other years, most notably on one year forward. For this test to be meaningful requires two previously unstated assumptions. First, the sunlight in a placebo year could not possibly influence the flu in the studied year. Second, sunlight in a placebo year, when looking across states, could not be strongly correlated with sunlight in the studied year.

Unfortunately, the sunlight data fails the second assumption, as there are multiple correlated years, the most extreme being the fall of 2009 and 2010. The following figures show those correlations. The figures show differences from the mean, as that is analogous to flu year fixed effect structure of the regressions. To avoid a mechanism correlation, the differences are relative to the 2003-2007 state-month mean.

First, Appendix Figure F1 shows a scatter plot of state level difference 2009 vs 2010, where September is in red and October is in blue. There are 100 pairs of points, as we including Washington, D.C. but excluding Alaska, consistent with our data.

**Appendix Figure F1: State Differences in Sunlight in 2009 vs 2010**



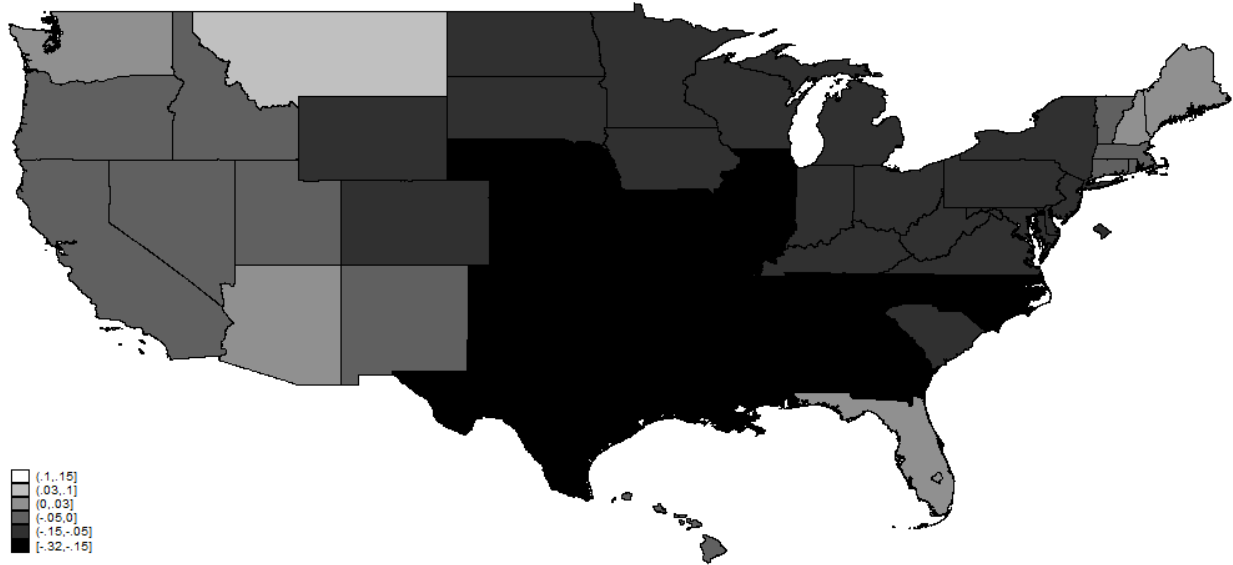
Notes: Differences from state-month mean for 2003-2007. Red dots are September and blue dots are October. Green line is linear best fit for the pooled data set of both months,

Surprisingly, the deviations in sunlight in 2010 are strongly negatively correlated with the deviations in 2009, the overwhelmingly relevant year. The correlation between the differences from the mean in the two years is -0.45, which is statistically significant at the 1% level. That implies that employing one-year ahead as a placebo year would be inappropriate.

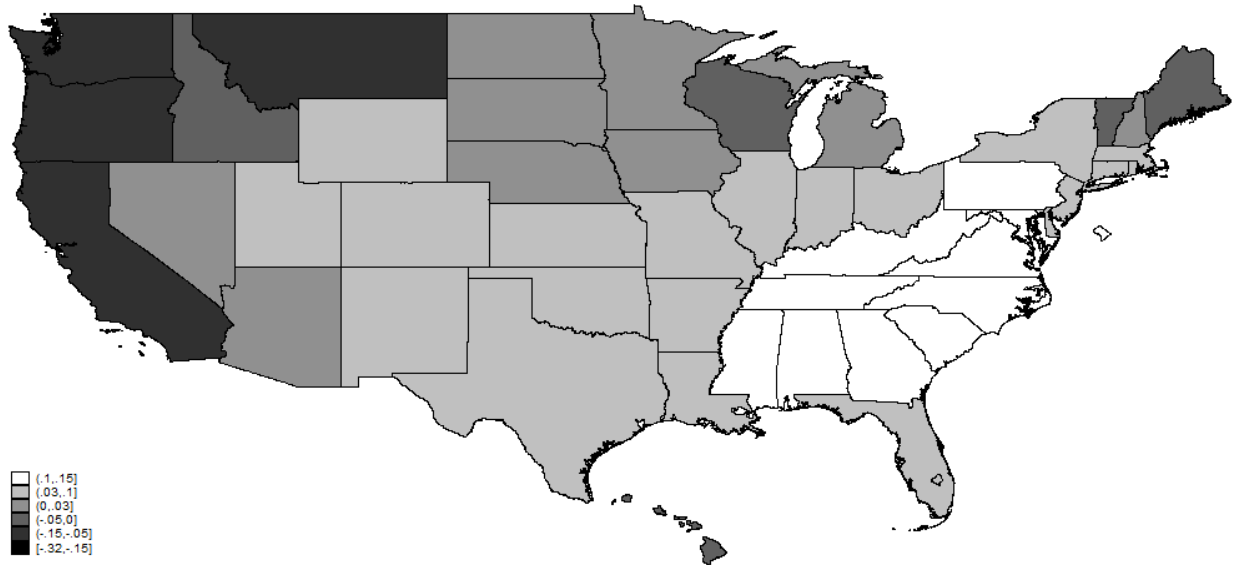
Appendix Figure F2 then shows the September and October average differences from the mean for each state and each year.

## Appendix Figure F2:

*Panel A: Sunlight in September & October 2009, Differences from the mean*



*Panel B: Sunlight in September & October 2010, Differences from the mean*



Notes: Lighter colors mean more relative sunlight, darker colors mean less relative sunlight. Scale is the same in both figures.

The figure shows the strong negative correlation between relative sunlight in the fall of 2009 and the fall of 2010. Given that the 2009 flu season was overwhelmingly responsible for our results, the one-year ahead lead is not an appropriate placebo test for our results.

In the final column of Appendix Table F1, we remove the one-year-ahead independent variables, and include only the previous five years and two and three years ahead. Here, our main coefficient is still highly statistically significant, and has a magnitude close to our main result. While we do see one other placebo result that is statistically significant, it is only at the 5% level and so unlike the main result would not survive an adjustment for testing multiple hypotheses.