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

RWP18-007

Sunlight and Protection Against Influenza

David J.G. Slusky
Richard J. Zeckhauser

February 9, 2018

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 within states and during the same calendar month, affect influenza incidence. We find that sunlight strongly protects against influenza. This relationship is driven by sunlight in late summer and early fall, when there are sufficient quantities of both sunlight and influenza activity. A 10% increase in relative sunlight decreases the influenza index in September by 3 points on a 10-point scale. This effect is far greater than the effect of vitamin D supplementation in randomized trials, a differential due to broad exposure to sunlight, hence herd immunity. We also find suggestive evidence, consistent with herd immunity theory, that the protective sunlight effect is strongest with a middle level of population density.

JEL codes: I10, I12, I18

Keywords: Seasonal Influenza, Sunlight, Vitamin D, Natural Experiment, Herd Immunity, Flu, Externalities, Epidemiology, Epidemic

---

Affiliations:

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

Zeckhauser: Harvard Kennedy School, Harvard University, 79 JFK St, Cambridge, MA, 02138, richard_zeckhauser@harvard.edu

Acknowledgements: We thank Nils C. Wernerfelt for exceptional advice and guidance throughout this project. We also thank Susan Graber for her comments and suggestions.

I. Introduction

Seasonal influenza has been with humans throughout history (Viboud and Epstein 2016). It imposes extreme costs on contemporary societies, with 2017-18 being a notable high outlier (CDC 2018). Beyond the significant discomfort to those it strikes, it 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 and 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 is a type of viral respiratory infection. Traditional public health measures to combat it include vaccination (Maurer 2009; and White 2016) and paid sick leave to keep contagious workers at home (Barmby and Larguem 2009; and Pichler and Ziebarth 2016). 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)].

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. 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.[1] Sunlight as a source has the added benefit that, unlike ingested vitamin D, which can become toxic at a certain concentration, the self-production mechanism does not generate toxic quantities (Holick 2007).

The 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). This paper is the first to estimate the relationship by calendar month and to find that the effect is largest in late summer and early fall, when there is both substantial sunlight and sufficient influenza activity. This paper is also the first to identify the relationship between population density and the strength of the sunlight-flu association.

Influenza is a contagious disease. Thus, one individual's protection will help reduce the risk to another, producing herd immunity when the level of protection in a community is sufficient. A theory developed in the mathematical biology literature indicates that there will be an intermediate optimal level of population density for herd immunity's greatest impact. Sparsely populated areas often cannot maintain sustained infection of a disease across a population (Bartlett 1957; Black 1996; and Anderson and May 1990), whereas highly dense populations require greater levels of immunization for the treatment to be effective (Arita, Wickett, and Fenner 1986). A more densely populated area is, therefore, both more susceptible to an epidemic and more able to benefit from herd immunity. Given this literature, we expect herd immunity to be strongest for the middle-density states, where there is significant influenza combined with

[1] The minimum amount of sunlight exposure (on head, neck, arm, and hands, without sunscreen) necessary to 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, whereas in the winter it can be over an hour. See http://nadir.nilu.no/~olaeng/fastrt/VitD-ez_quartMEDandMED_v2.html to calculate the minimum effective exposure time given a certain set of conditions.

sufficient density to make herd immunity meaningful. The analysis below confirms this hypothesis. In this context, we also note that sunlight-created vitamin D, as opposed to ingested-supplement vitamin D, automatically tends to enhance levels broadly within a community, thereby offering superior prospects for herd immunity.

II. Data

For influenza data, we used the CDC's flu index. The CDC index aggregates data reports from the individual state health department influenza surveillance points, and then harmonizes the aggregate to a consistent 10-point scale. Each point on the index represents an additional standard deviation above the mean for the ratio of visits to outpatient healthcare providers by those with symptoms of influenza, relative to all outpatient visits (regardless of symptoms). Weekly state-level data are available, from October 2008 to the present.[2] Some states, however, are missing individual weeks of data. Dropping the jurisdictions with missing flu data and the two states (Alaska and Hawaii) that are missing sunlight data leaves us with 36 contiguous states for our primary analysis sample (CDC 2017a).[3]

We combined this flu data with the North America Land Data Assimilation System's daily county-level sunlight data for 2003-2011, which covers the 48 contiguous states and the District of Columbia. This data represents the average of all observations (each a 1/8-degree grid cell, namely 14x14 square kilometers) of daily solar radiation (in kilojoules per square meter) in each county for each studied month from 2003-2011. While our influenza data only begins in 2008, we used earlier sunlight data for placebo tests (CDC 2016).

[2] See Appendix A for more details about the how the index is calculated.

[3] We drop 12 of the lower 48 states: Arkansas, Connecticut, Delaware, Idaho, Iowa, Louisiana, Maryland, Oklahoma Oregon, South Dakota, Virginia, and Washington. Washington, D.C. is also dropped due to missing data. As shown in Appendix Table 1, when we include these 13 jurisdictions, using whatever data is available for each month, we find consistent results.

For state-level population density, we used a variety of sources from the Census (2010a and 2010b) and also the SAS ZIP Code Database (2013). For several of our robustness checks, we also included weather station data 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 to the population-weighted centroid for each county (Census 2010c).

III. 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 approach follows an alternate path to methodological soundness. As an econometric study, it employs quasi-experimental variation to effectively create equivalent randomization. Implicitly, this approach controls for a wide number of variables. Moreover, it 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 month 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 month of the influenza report and the prior month. 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. We also calculate the monthly average flu index in each state from the weekly CDC data to get a monthly outcome variable.

We estimate the impact of the percent of deviation of sunlight (the change in log points) from its mean on deviations of the flu index from its mean as follows:

$$Flu_{smy} = \alpha + \boldsymbol{\gamma} \ln(sunlight_{smy}) + \mathbf{statemonth}_{sm} + \mathbf{year}_{y} + \varepsilon_{smy} .$$

Flu_{smy} is the flu index for state *s* in month *m* in year *y*. $Sunlight_{smy}$ refers to the average sunlight for month *m* and the prior month (as described above) for state *s* in year *y*. γ is our coefficient of interest. Our preferred specification includes interaction terms (**statemonth**) for state-month fixed effects (for example, October in Kansas) and year fixed effects (for example, 2009).[4] Robust standard errors are clustered at the state level. 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. There, we found a strong protective impact of a pregnant

[4] 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.

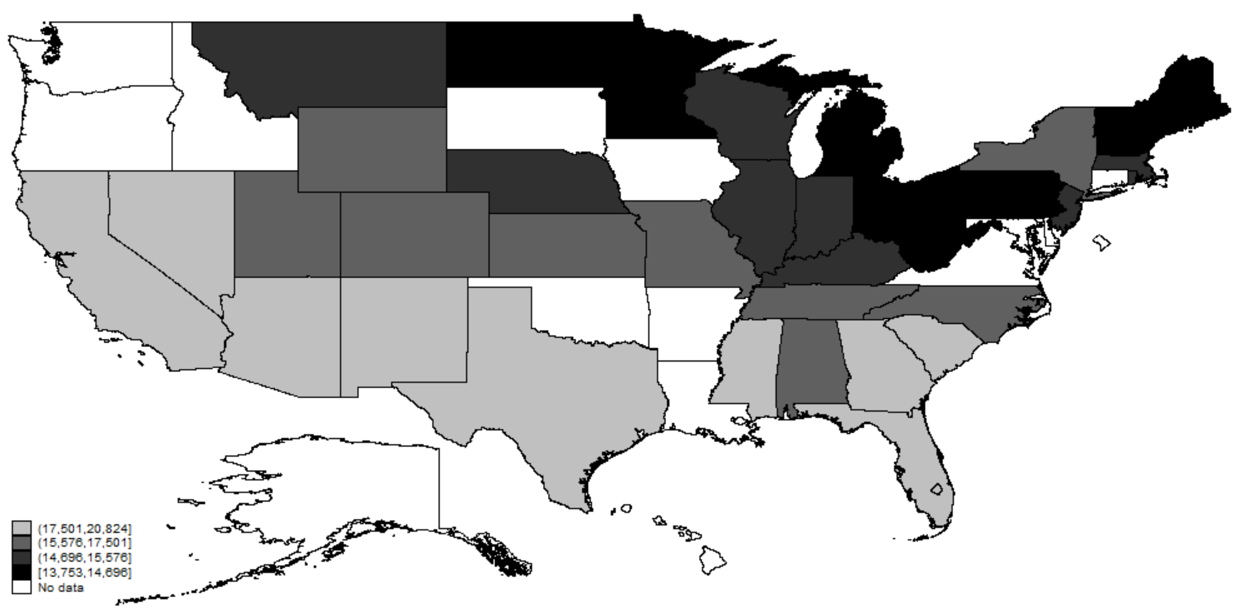
woman's exposure to sunlight on later-in-life asthma in her child (Wernerfelt, Slusky, and Zeckhauser 2017).

As described above, we stratify our current regressions by state population density. This enables us to test the hypothesis that there will be an optimal middle density, one dense enough for the virus to maintain itself but not so dense that externalities of protection are minimal. We also check that our results retain significance after adding a variety of weather controls, calculated analogously as county population-weighted averages, which others have found to have a significant impact on health in general and influenza in particular (including Barreca 2012; Barreca and Shimshack 2012; Deschenes 2013; Barreca, Deschenes, and Guldi 2015; Barreca et al. 2016; and Huetal, Miller, and Molitor 2017)

IV. Results

First, we examine variation in population-weighted sunlight averages (in kilojoules per square meter per day). Figure 1 shows the three-year (2009-2011) average.

Figure 1: Population-Weighted Geographic Sunlight Variation

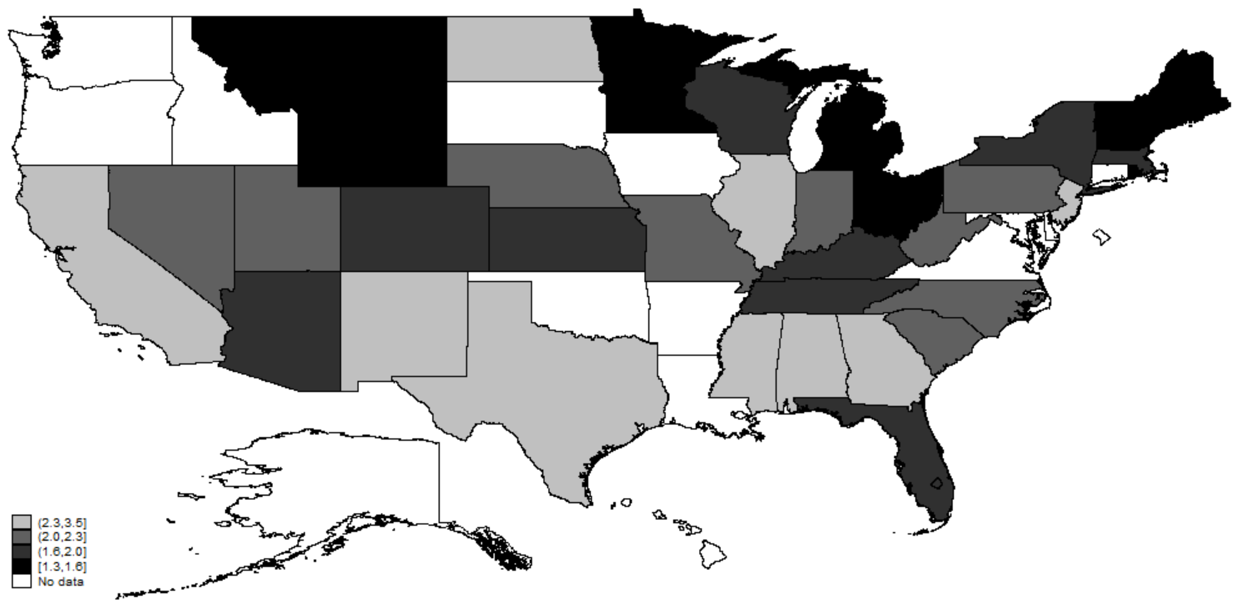


Notes: 3-year average (2009-2011) of daily county sunlight, weighted by county population.

We see the expected pattern, which is that the United States is sunnier in the south and west. One difference to note is that large states with populations concentrated in their southern regions (for example, New York) are sunnier by this metric than are nearby states with more evenly spread populations.[5]

Figure 2 then shows the variation by state in the average influenza index.

Figure 2: Geographic Flu Variation



Notes: 3-year average (2009-2011) of weekly state-level flu index.

Here we see a very different pattern than in Figure 1. 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 North Dakota and Illinois), and some have low flu levels (such as Maine and New Hampshire). This suggests that other state-specific factors strongly influence influenza levels, which makes controlling for state-specific fixed effects important.

Table 1 shows summary statistics for the flu index and population-weighted average

[5] The figures in Appendix B show the sunlight data in more detail, and investigates apparent puzzles, such as why New York gets more population-weighted sunlight than New Jersey.

sunlight levels, as well as for population density (used to stratify the results) and other weather variables (used as additional controls.)

Table 1: Summary Statistics

	(1) N	(2) Mean	(3) StDev	(4) Min	(5) Max
Flu index	1,404	2.000	2.139	1	10
Sunlight (kJ/m^2/day)	1,404	15,771	6,509	4,576	30,334
Population Density (individuals/mi^2)	1,404	197.2	269.5	5.8	1,195
Temperature (°F)	1,404	54.0	17.9	5.1	94.3
Days/month temp <15°F	1,404	2.0	4.7	0	29.8
Specific humidity (g water vapor / kg air)	1,404	10.8	6.4	1.8	29.7
Days/month specific humidity < 6 g/kg	1,404	9.8	10.5	0	31

Note: Unit of observation is a year-month for each of the 36 contiguous 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. Population density also varies substantially across states (see Figure 5 below). Finally, temperature and humidity vary extensively.

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

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

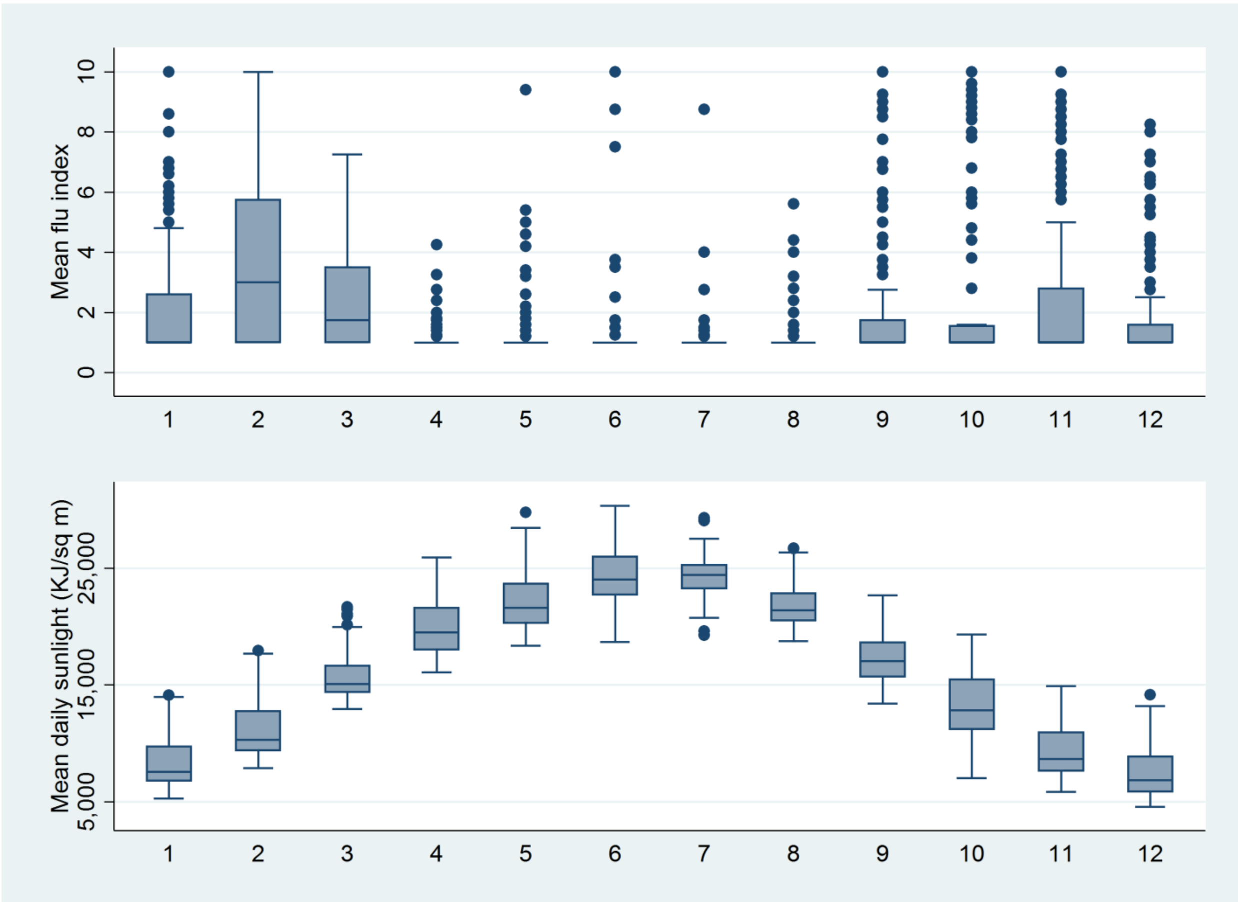
	(1)	(2)	(3)	(4)
Log sunlight for that month	-7.105***		-6.705***	
	(1.215)		(1.201)	
Log sunlight for the prior month		-4.942***	-4.321***	
		(1.009)	(0.964)	
Log sunlight for that month and the prior month				-10.57***
				(1.859)
Observations	1,404	1,404	1,404	1,404
R-squared	0.479	0.463	0.490	0.485

Notes: All regressions include month-state and year fixed effects. 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 relative sunlight for a month would lead to a 0.7-point decline in the influenza index for that month. In Column (2), we instead use the sunlight from the prior month (given the long half-life of vitamin D) and find a substantial effect as well. Column (3) includes each sunlight variable separately, and finds that each retains at least 85% of its magnitude when included alone. Column (4) confirms this by including instead a single variable for the average sunlight over the past two month, which as a coefficient is 95% of the sum of the two coefficients in Column (3). Given the uncertainty around the length of vitamin D's half-life (Mawer, Schaefer, Lumb, and Stanbury 1971; and Jones 2008), for the rest of the paper, we will include this broader two-month variable as our primary specification.

Table 2, however, includes months that have minimal influenza activity, and also months that have low levels of sunlight. Inclusion of either blunts the magnitude of the coefficients, and obscures any seasonality in the results. To address this, in Figure 3 we first plot the range of influenza and sunlight by month. The top half of the 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 extreme 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.

Figure 3: Box Plots of Average Flu and Sunlight by Month



Notes: Covers the 36 contiguous states that have full flu and sunlight data. Outliers are shown in blue dots.

Motivated by these plots, in Table 3 we re-estimate our model for each month of flu data after including the impact of that month and the prior month's sunlight. Here we include only state fixed effects. Given that each column includes data for only one calendar month of each year, adding month fixed effects would have no influence.

Table 3: Month by Month[6]

	(1) Jan	(2) Feb	(3) Mar	(4) Apr	(5) May	(6) Jun	(7) Jul	(8) Aug	(9) Sep	(10) Oct	(11) Nov	(12) Dec
Log sunlight for that month and the prior month	9.872	14.02	-5.661	-0.660	0.172	-3.526	-0.723	4.996	-30.64***	-7.315**	0.808	-1.622
	(7.170)	(11.70)	(7.560)	(1.226)	(2.346)	(4.051)	(1.138)	(3.480)	(7.299)	(3.061)	(4.579)	(2.539)
Observations	108	108	108	108	108	108	108	108	108	144	144	144
R-squared	0.679	0.698	0.524	0.402	0.514	0.379	0.404	0.431	0.749	0.922	0.816	0.642

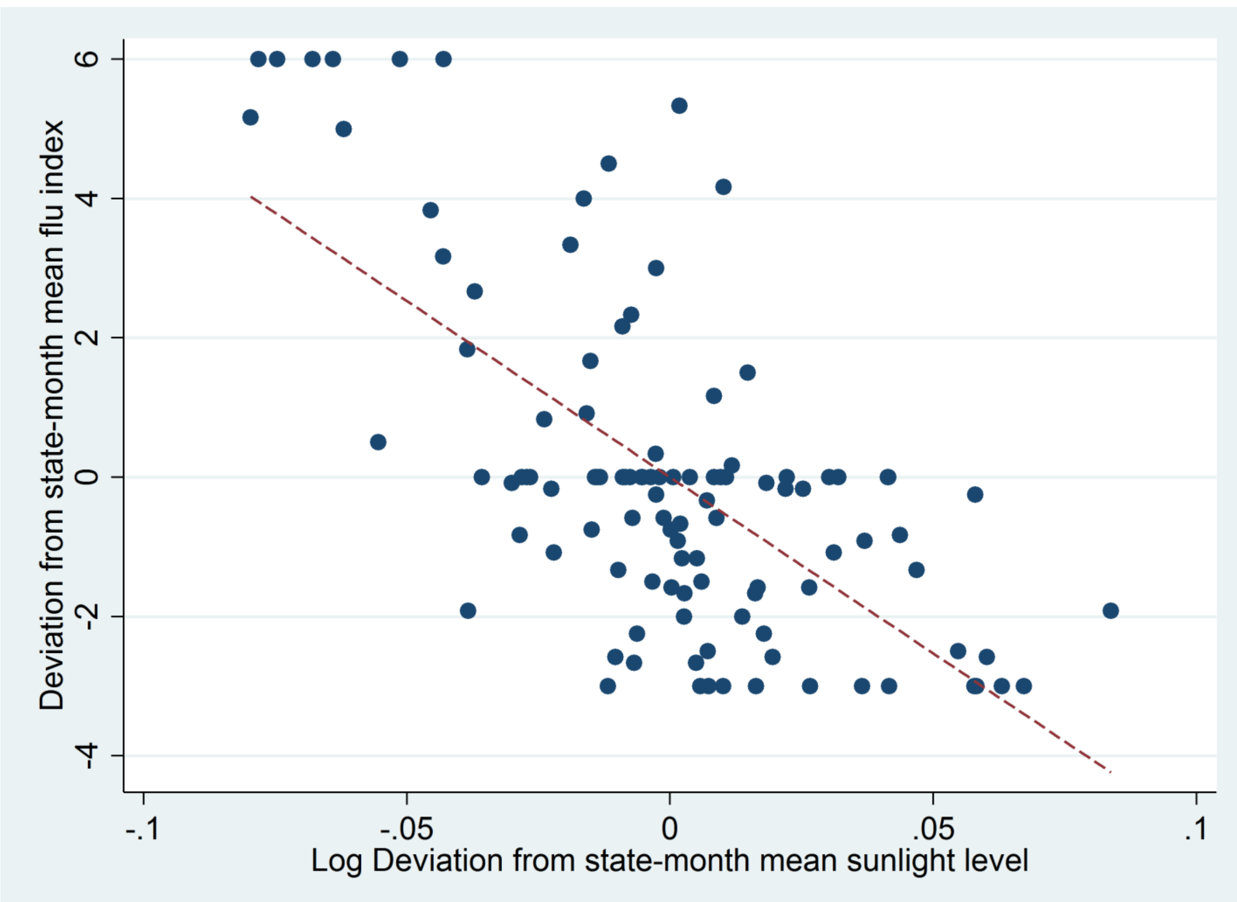
Notes: All regressions include state and year fixed effects. Robust standard errors clustered at the state level in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

[6] 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 36 states included in the primary analytic sample.

Table 3 shows that our results are being driven by September influenza (that is, August and September sunlight), and to a lesser extent by October influenza (that is, September and October sunlight). These months meet the overall requirements (as shown in Figure 3) of non-trivial level of influenza activity and still-substantial levels of sunlight. For these months, a 10% increase in relative sunlight levels leads to a 3-point decline in the influenza index.

We can also see this result in graphical form. Figure 4 graphs the deviations in the September influenza index and the log level of August and September sunlight from the mean for each state.

Figure 4: State-Month Deviations for Flu and Sunlight



Notes: Only for influenza data for September. Dashed line is linear best fit.

The horizontal axis displays our independent variable, the log of sunlight by date and month. The vertical axis graphs our dependent variable, flu index by state and month. Both variables are measured in deviations from the mean value, as log points[7] for sunlight, and by points on its index for flu incidence. 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. This relationship is clearly seen in this picture, confirming our results.

To additionally check the robustness of our results, Table 4 employs our preferred specification from column 9 in Table 3 and adds lagged sunlight for years before the treatment period. If the results are robust, such lagged variables should have little or no effect. Table 4 shows that our primary coefficient retains its statistical significance and 90% of its magnitude, despite the inclusion of multiple other independent variables. None of the coefficients on those added independent variables, which we take to be placebos, ever reaches statistical significance at even the 10% level. Moreover, the R-square value hardly changes as these variables are added.

[7] Given that these deviations are small, they can be interpreted as percentage points.

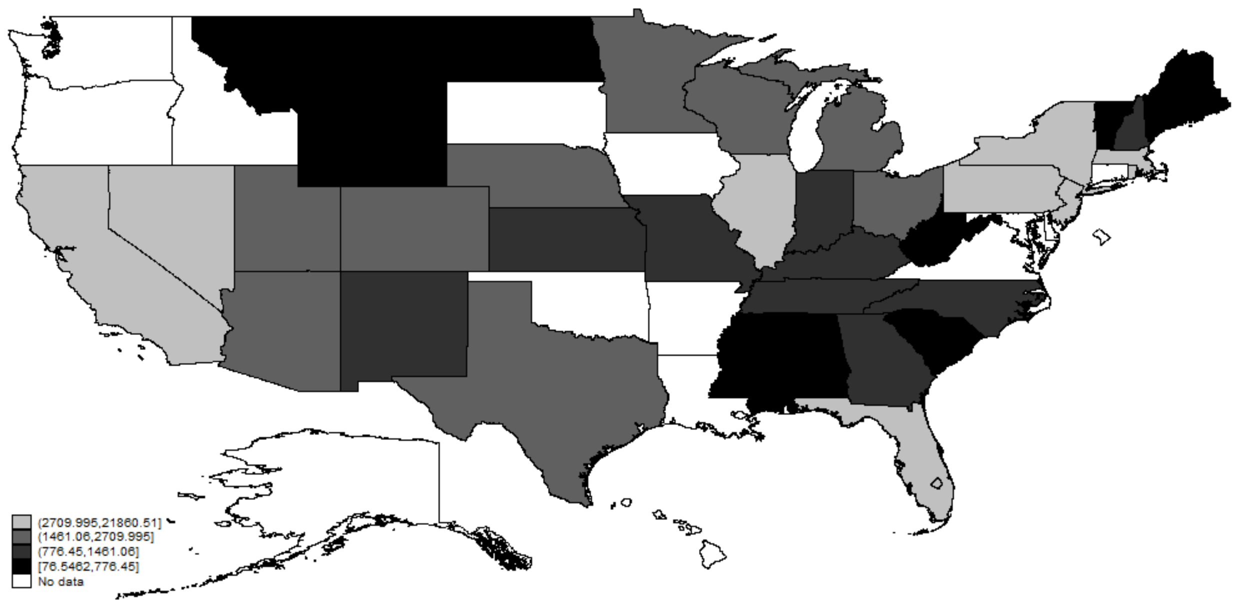
Table 4: Retrospective Placebo Results for September Flu

		(1)	(2)	(3)	(4)	(5)	(6)
Log sunlight for that month and the prior month	Treatment year	-30.64***	-32.45***	-29.73***	-29.62***	-29.38***	-27.83***
		(7.299)	(7.382)	(6.538)	(6.605)	(6.946)	(7.286)
	Year -1		-4.853	-0.532	3.209	3.216	4.406
			(7.906)	(9.730)	(11.53)	(11.65)	(11.16)
	Year -2			10.93	14.31	16.90	18.19
				(7.329)	(10.48)	(16.54)	(15.58)
	Year -3				7.252	9.337	12.57
					(12.06)	(16.36)	(15.39)
	Year -4					4.543	10.46
						(15.20)	(16.09)
	Year -5						8.593
							(16.04)
	Observations	108	108	108	108	108	108
	R-squared	0.749	0.751	0.763	0.765	0.766	0.768

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

We now turn to testing our hypothesis that there is an optimal, middle level of population density at which sunlight will have the strongest protective effect. First, we divide the 36 states in our sample into quartiles of population density, using data from the Census (2010a). Figure 5 shows which states lie in each quartile.

Figure 5: Population-Weighted Average Population Density[8]



Fortunately for this investigation, the states are reasonably distributed around the country, with each geographic region, as defined by average sunlight, having states in each density quartile.

We can also see this by looking at influenza and sunlight ranges by density quartile. Figure 6 repeats the box plots from Figure 3 for September only, by quartiles of population density.

8

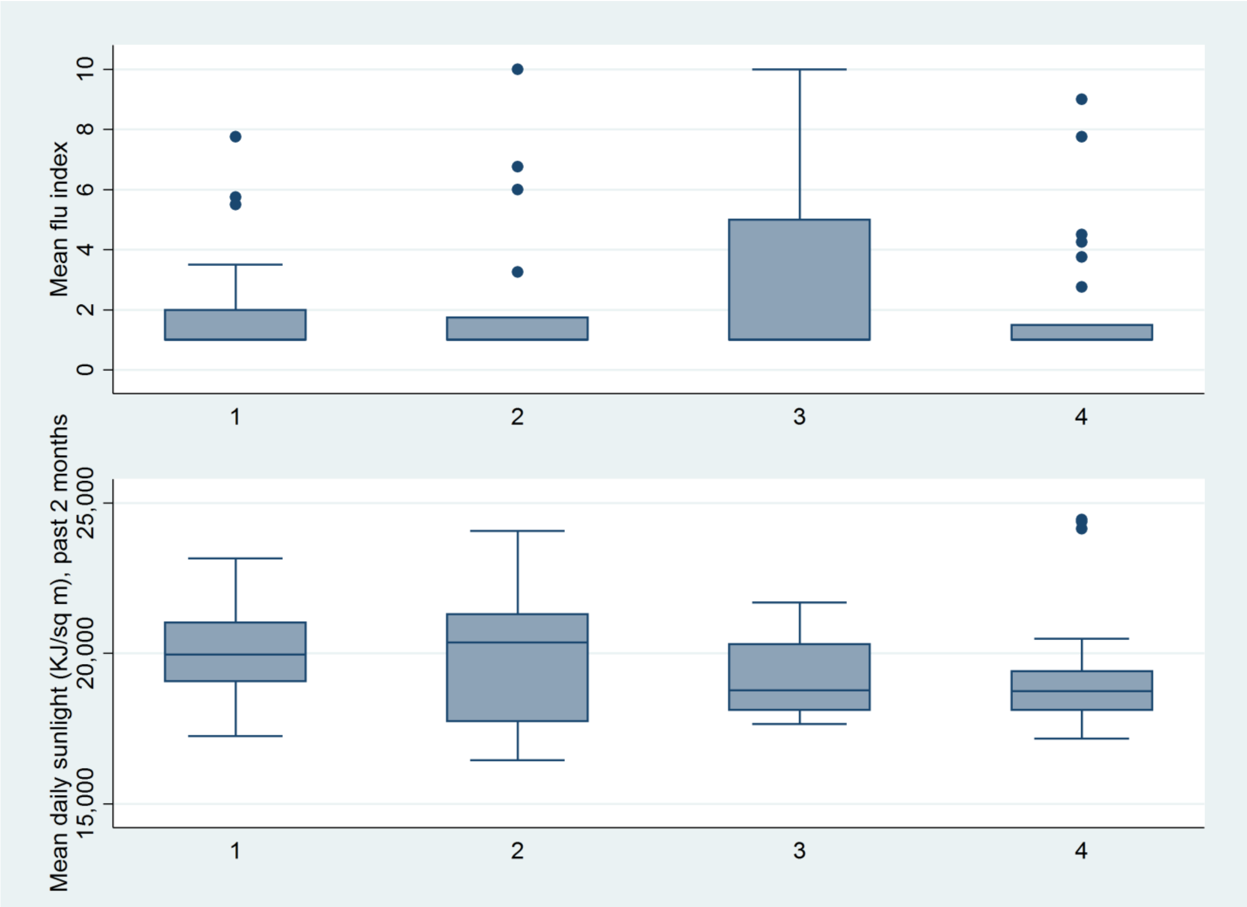
1st Quartile States: Alabama, Maine, Mississippi, Montana, North Dakota, South Carolina, Vermont, West Virginia, and Wyoming.

2nd Quartile States: Georgia, Indiana, Kansas, Kentucky, Missouri, New Hampshire, New Mexico, North Carolina, and Tennessee.

3rd Quartile States: Arizona, Colorado, Michigan, Minnesota, Nebraska, Ohio, Texas, Utah, and Wisconsin.

4th Quartile States: California, Florida, Illinois, Massachusetts, Nevada, New Jersey, New York, Pennsylvania, and Rhode Island.

Figure 6: Range of Flu and Sunlight in September by Quartiles of Population Density



Notes: Only August and September sunlight. Outliers are shown in blue dots.

Here we see vastly more influenza variation in the 3rd quartile of population density. This cannot be due simply to sunlight variability, since that variability is greater in the 2nd quartile of density than in the 3rd.

The box plots are suggestive, but a regression analysis may be more definitive. Table 5, therefore, stratifies our September regressions by population density.

Table 5: Results Stratified by Population Density

	(1) 1st Quartile	(2) 2nd Quartile	(3) 3rd Quartile	(4) 4th Quartile
ln_sun_0_1	-9.570	-28.70	-43.05***	-3.118
	(17.16)	(18.60)	(11.33)	(21.79)
Observations	27	27	27	27
R-squared	0.637	0.747	0.907	0.630

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

As expected, the effect is largest in magnitude in the middle quartiles of population density, it is particularly substantial and highly significant in the 3rd quartile.

This finding is strongly consistent with the joint hypothesis: 1. Herd immunity is a powerful part of the sunlight-vitamin D protective effect. 2. Herd immunity is most effective in locations with medium population density.

Even with this finding, strongly supported by both theory and empirical analysis, there could be a concern that our results are picking up some other kind of environmental variation. One possibility would be temperature, given that it is known to affect health (Deschenes 2013; Barreca, Deschenes, and Guldi 2015; Barreca et al. 2016; and Huetal, Miller, and Molitor 2017). Thus, following Barreca, Deschenes, and Guldi (2015) and Wernerfelt, Slusky, and Zeckhauser (2017), we now control for the number of days per month that a state experiences extreme cold (daily low temperature below 15°F). Such control is merited, because the influenza virus can survive better between hosts at lower temperatures (Polozov et al. 2008). Finally, 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 g of water vapor per kg of air had a substantial impact (per Barreca 2016; Barreca and Shimshack 2012).[9]

The results after adding these additional controls are shown in Table 6.

Table 6: Results Controlling for Other Weather Measures

	(1)	(2)	(3)	(4)
Log sunlight for that month and the prior month	-30.94*** (7.715)	-31.27*** (7.500)	-32.52*** (8.448)	-31.38*** (7.544)
Controls for log temperature for that month and the prior month	X			
Controls for the number of days per month below 15°F		X		
Controls for log specific humidity for that month and the prior month			X	
Controls for the number of days per month specific humidity is below 6 g/kg				X
Observations	108	108	108	108
R-squared	0.749	0.751	0.750	0.751

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

As shown in Table 6, adding these weather controls minimally affects our primary finding.

Additional Robustness Checks

The tables in Appendix A conduct additional robustness checks. Appendix Table C1 repeats the Table 2 analyses, but includes all contiguous states and D.C. (even those with missing influenza data). It finds a comparable result. Appendix Table C2 does the same, but now only for September influenza data, and again finds comparable results. Appendix Table C3 repeats Table

[9] Specific humidity is not directly provided in the NOAA data, so we calculated it using the available information on dew point and atmospheric pressure. See http://snowball.millersville.edu/~adecaria/ESCI241/esci241_lesson06_humidity.pdf for the necessary formulas.

1, but now employs a linear specification and finds strongly statistically significant results, though obviously at different coefficient magnitudes. Appendix Table C4 repeats the population density stratification but uses all months of data; it finds comparable results that show larger effects for the middle quartiles when compared to the top and bottom quartiles.

Discussion

We can attempt to estimate the impact of our results on welfare. As described above, each point on the influenza index represents an additional standard deviation above the mean of the non-flu week's ratio of outpatients presenting with symptoms of influenza to all outpatients (CDC 2017a). The data is also available on the actual outpatient counts, though not broken down at the state level (CDC 2017b).

As described above, the flu index indicates the number of standard deviations above the non-influenza mean of the share of outpatients who exhibit influenza symptoms. In the 2005-2008 "pre-period," this mean share is 1.03%, and the standard deviation is 0.39 percentage points.[10]

In Figure 4, we see that the range of relative sunlight levels for September within states across years is roughly plus or minus 0.05 log points, that is, 10 percentage points. Thus, our coefficient for log sunlight shown in Table 3 for September corresponds to a 3-point reduction in the influenza index, which can be interpreted as 3 standard deviations. Given that one standard deviation is 0.394 percentage points, 3 standard deviations would be 1.18 percentage points.

The average total number of all outpatients in September in our study years (2009-2011) (from CDC 2017b) is 2,504,249. A 1.18 percentage point reduction would produce 29,567 fewer cases.

[10] See Appendix A for additional calculation details.

To translate this into a dollar amount, we need two additional pieces of information. First, Molinaria et al. (2007) estimate that the total cost of seasonal influenza is $87 billion per year.[11] Secondly, again using the CDC (2017b) data, the average annual number of influenza patients for 2009-2011 is 718,285. Our reduction of 29,567 is 4.1%, which would give us approximate monetary equivalent savings of $3.6 billion.

It is worth considering why our results are so much larger in magnitude than the results secured in the randomized control trials of Martineau et al. (2017), which found an adjusted odds ratio of only 0.88 of acute respiratory tract infection for those receiving vitamin D supplements. The most likely explanation for the disparity relates to externalities promoting herd immunity when sunlight is the protective factor. Giving 100 people in a town of perhaps 10,000 people a vitamin D supplement will offer virtually no externalities of protection. But give that same town extra sunlight, and most of the community will produce vitamin D, thereby conveying an externality of protection that triggers herd protection against influenza, a highly communicable disease.[12] Positing that supplements and sunlight-produced vitamin D are equivalently powerful, that externality could massively increase the magnitude of the protective effect.

Finally, sunlight can also protect against influenza via a path apart from the production of vitamin D. Ultraviolet light deactivates the virus directly (Sagripanti and Lytle 2007). The data in this paper provides no direct way to assess the relative contributions of these two mechanisms. However, we can be confident that the vitamin D path is consequential. It has been demonstrated in a randomized control trial of supplements (Martineau et al. 2017).

[11] This estimate includes the cost of hospitalization and outpatient visits, lost earnings, and life-years lost. It does not include disutility from having the flu.

[12] 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 contagious.

Conclusion

Sunlight, likely operating through the well-established vitamin D channel, plays a significant role in flu incidence. A recent meta-analysis of 25 randomized controlled trials of vitamin D supplementation (Martineau et al. 2017) demonstrated significant benefits of vitamin D 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 reinforce the Martineau et al. evidence that vitamin D protects against such infections. The Martineau meta-analysis imposed stringent criteria for including a trial, thereby ruling out a variety of confounding factors, such as selection effects.

This analysis adds three findings to this literature. First, the relationship between relative sunlight and influenza is likely driven by late summer and early fall sunlight, and not by sunlight the rest of the year. Second, externalities of immunity contribute strongly to the protective effects of sunlight. Third, these externalities, the underpinnings of herd immunity, operate most forcefully at a middle level of population density.

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 sunny day.

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.

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, R.M., and R.M. May. 1990. Immunisation and Herd Immunity. The Lancet 335(8690): 641-645.

Arita, I., J. Wickett, and F. Fenner. 1986. Impact of population density on immunization programmes. J. Hyg., Camb. 96:459-466.

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. 2015. Maybe Next Month? The Dynamic Effects of Ambient Temperature on Fertility. NBER Working Paper No. w21681.

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.

Black, Francis L. 1996. Measles endemicity in insular populations: Critical community size and its evolutionary implication. Journal of Theoretical Biology 11(2): 207-211.

CDC 2016. North America Land Data Assimilation System (NLDAS) Daily Sunlight (KJ/m²) (1979-2011. https://wonder.cdc.gov/nasa-insolar.html

CDC. 2017a. Influenza-Like Illness (ILI) Activity Level Indicator Determined by Data Reported to ILINet https://gis.cdc.gov/grasp/fluview/main.html

CDC 2017b. National, Regional, and State Level Outpatient Illness and Viral Surveillance. https://gis.cdc.gov/grasp/fluview/fluportaldashboard.html

CDC 2017c. Overview of Influenza Surveillance in the United States. https://www.cdc.gov/flu/weekly/overview.htm

CDC 2018. Transcript for CDC Update on Flu Activity. https://www.cdc.gov/media/releases/2018/t0126-flu-update-activity.html

Census 2010a. State-level population density. https://factfinder.census.gov/bkmk/table/1.0/en/DEC/10_SF1/GCTPH1.US05PR

Census 2010b. Gazetteer Files for Zip Code Tabulation Areas. https://www.census.gov/geo/maps-data/data/gazetteer2010.html; http://www2.census.gov/geo/docs/maps-data/data/gazetteer/Gaz_zcta_national.zip

Census 2010c. US Population Weighted Center. https://www2.census.gov/geo/docs/reference/cenpop2010/CenPop2010_Mean_ST.txt

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.

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. Forthcoming at Health Economics.

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. 2017. Adaptation and the Mortality Effects of Temperature Across U.S. Climate Regions. NBER Working Paper No. w23271.

Jones, G. 2008. Pharmacokinetics of vitamin D toxicity. Am J Clin Nutr, 88(2):582S-586S.

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. Forthcoming at BMJ.

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.

Molinaria, NA.M., I.R. Ortega-Sanchez, M.L. Messonnier, W.W. Thompson, P.M. Wortley, E. Weintraub, C.B.Bridges. 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs, Vaccine 25(27):5086-5096.

NOAA. 2017. ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt

Pichler, S., and N.R. Ziebarth. 2016. The Pros and Cons of Sick Pay Schemes: Testing for Contagious Presenteeism and Noncontagious Absenteeism Behavior. NBER Working Paper No. w22530.

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.

Sagripanti, J.L. and C.D. Lytle. 2007. Inactivation of Influenza Virus by Solar Radiation. Photochemistry and Photobiology 83: 1278–1282.

SAS Zip Code Data Set. 2013. https://support.sas.com/downloads/download.htm?did=104285#.

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.

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. 2016. Measuring the Social and Externality Benefits of Influenza Vaccination. *Mimeo*, available at https://econ.ucsb.edu/docs/default-source/default-document-library/white-corey---jmp.pdf?sfvrsn=0

ZIP Code Database. 2013 https://www.unitedstateszipcodes.org/zip-code-database/

Appendix A

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 perform 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 differs when compared to all non-influenza weeks. A “non-influenza week” is defined as a week during which that week and its preceding week had fewer than 2% of all outpatient visits to healthcare providers indicating influenza.

As our study period is 2008-2011, 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 perform 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:

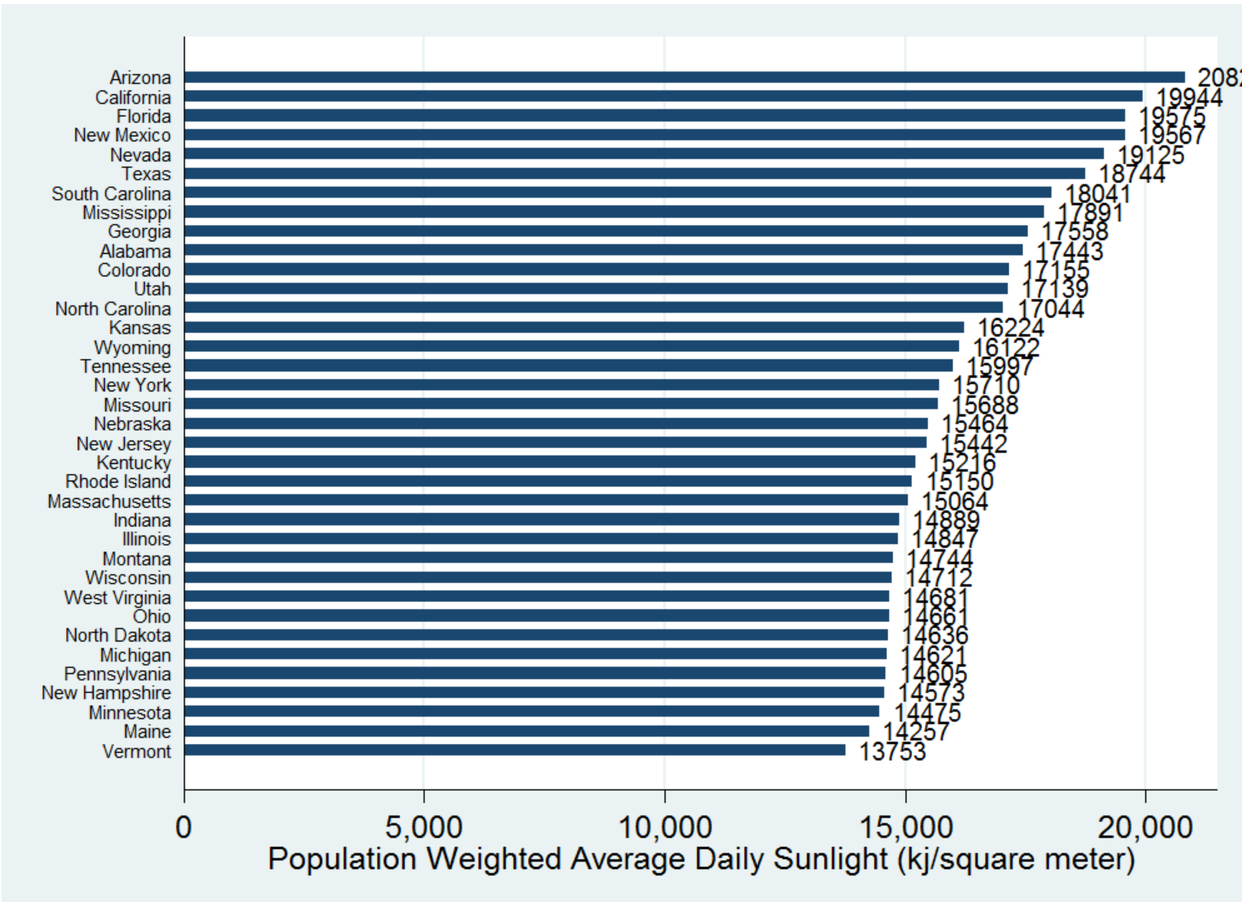
Flu index =

1	if	$Z < 0$
int(Z) +2	if	$0 < Z < 8$
10	if	$Z > 8$

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

Appendix B

Figure B1: Population-Weighted Average Daily Sunlight by state, 2009-11



Some of these rankings may seem surprising, such as New York ranking ahead of more southern nearby states of New Jersey and Pennsylvania. Figure A2 shows that New York City and Long Island (where much of New York State's population is) are substantially sunnier than New Jersey. Pennsylvania is further darkened by Pittsburgh, which is substantially less sunny than the East Coast. Figure A3 shows a similar situation when comparing Wisconsin and West Virginia, where Milwaukee is substantially sunnier on average than Charleston. Finally, Figure A4 shows the monthly averages for six populous counties across these states in question. For example, although Milwaukee is less sunny in the winter, it is sunnier than much of the East Coast in the summer, leading to Wisconsin's higher overall average.

Despite all of this evidence that our population-weighted state-level sunlight averages are valid, we also perform our main regressions below without New York State; we find that the results are robust.

Figure B2: Average Daily Sunlight and Population for Counties in Select States, 2009-11

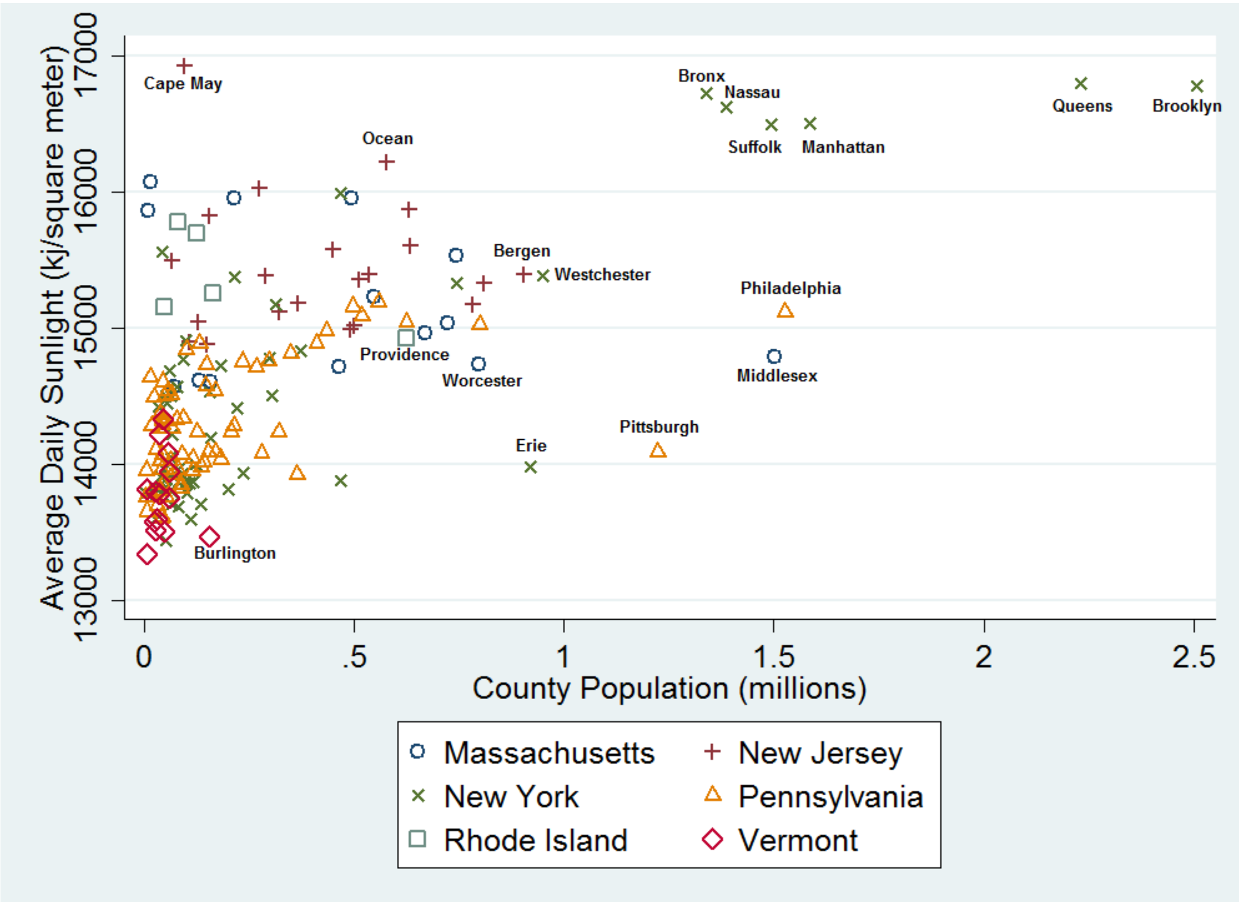


Figure B3: Average Daily Sunlight and Population for Counties in Select States, 2009-11

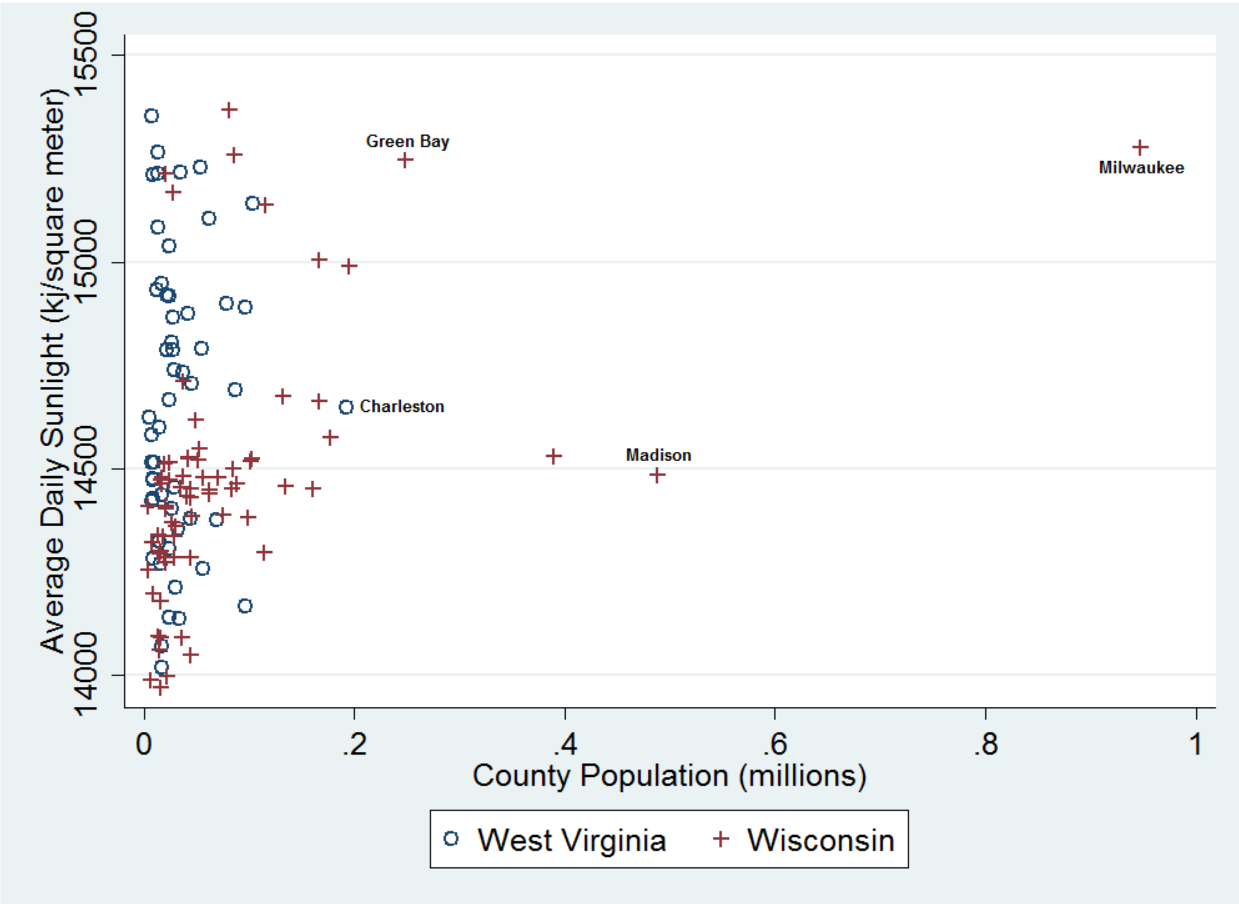
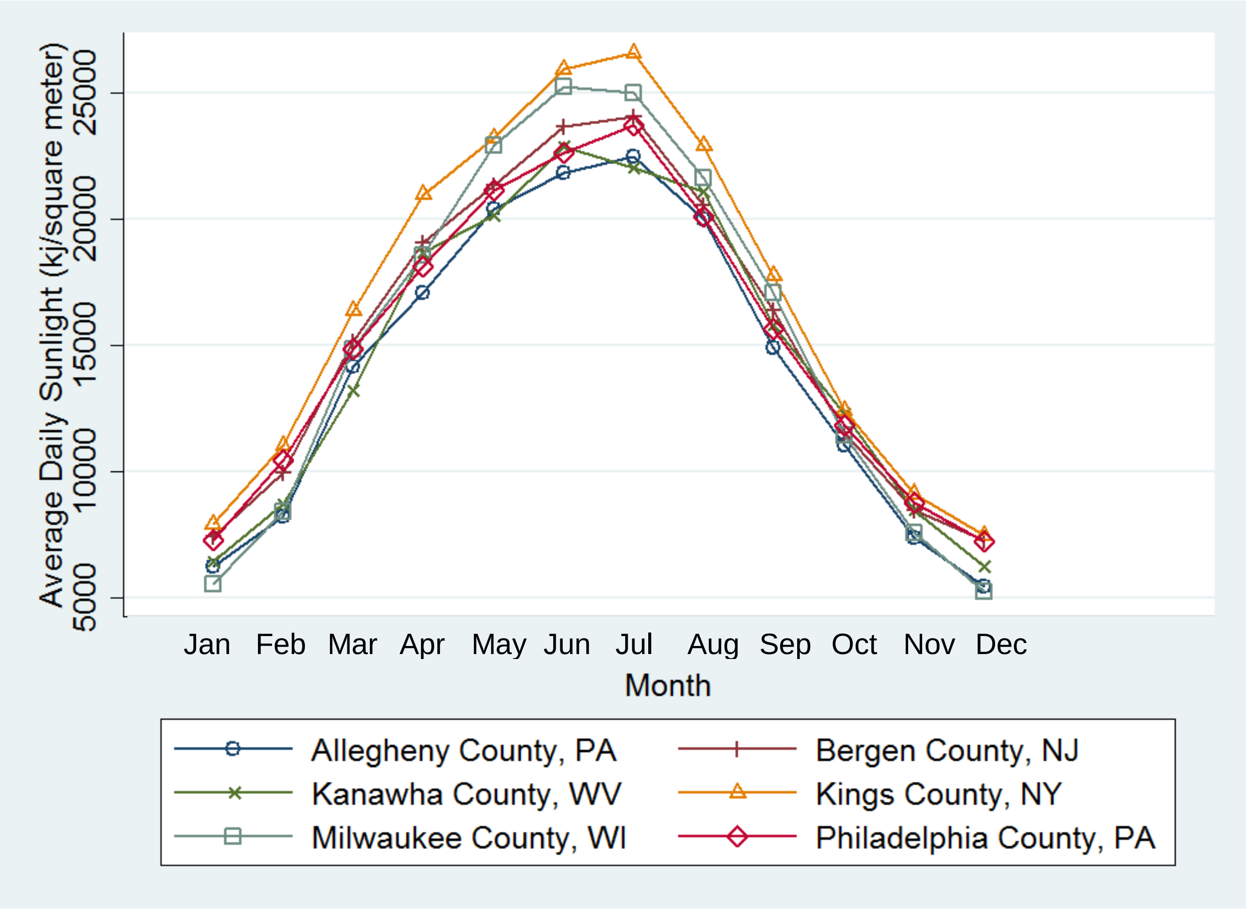


Figure B4: Average Daily Sunlight and Population for Select Counties, By Month, 2009-11



Appendix C

Appendix Table C1: Results Including States with Missing Flu Values, All Months

	(1)	(2)	(3)	(4)
Log sunlight for that month	-7.484***		-6.951***	
	(0.973)		(0.937)	
Log sunlight for the prior month		-5.087***	-4.201***	
		(0.872)	(0.803)	
Log sunlight for that month and the prior month				-10.76***
				(1.488)
Observations	1,875	1,875	1,875	1,875
R-squared	0.501	0.482	0.512	0.507

Notes: Includes all 48 contiguous states and D.C. Robust standard errors clustered at the state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C2: Results Including States with Missing Flu Values, September Only

	(1)	(2)	(3)	(4)
Log sunlight for that month	-13.33***		-14.03***	
	(2.876)		(3.007)	
Log sunlight for the prior month		-3.139	-7.538	
		(5.968)	(5.809)	
Log sunlight for that month and the prior month				-27.41***
				(6.183)
Observations	141	141	141	141
R-squared	0.741	0.650	0.749	0.740

Notes: Robust standard errors clustered at the state level in parentheses. September only. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C3: Linear specification

	(1)	(2)	(3)	(4)
Sunlight for that month	-0.000364*** (7.23e-05)		-0.000330*** (6.97e-05)	
Sunlight for the prior month		-0.000260*** (6.79e-05)	-0.000208*** (6.42e-05)	
Sunlight for that month and the prior month				-0.000537*** (0.000118)
Observations	1,404	1,404	1,404	1,404
R-squared	0.468	0.459	0.474	0.473

Notes: Robust standard errors clustered at the state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix Table C4: Results Stratified by Population Density, All Months

	(1) 1st Quartile	(2) 2nd Quartile	(3) 3rd Quartile	(4) 4th Quartile
Log sunlight for that month and the prior month	-11.05** (4.268)	-12.44** (4.807)	-14.55*** (3.192)	-5.658* (2.826)
Observations	351	351	351	351
R-squared	0.457	0.524	0.511	0.446

Notes: Robust standard errors clustered at the state level in parentheses. *** p<0.01, ** p<0.05, * p<0.1