Industrial Persistence in the U.S. and the 1918 Influenza Pandemic

by

Conor Harty

Dr. Maggie Jones
Adviser

Economics

Dr. Maggie Jones
Advisor

Dr. Jagdish Sheth
Committee Member

Dr. Kelli Lanier
Committee Member

Dr. Shomu Banerjee
Committee Member

2023

Industrial Persistence in the U.S. and the 1918 Influenza Pandemic

By

Conor Harty

Dr. Maggie Jones
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2023

Abstract

Industrial Persistence in the U.S. and the 1918 Influenza Pandemic

By Conor Harty

The 1918 flu was one of the deadliest pandemics in history, killing over 50 million people worldwide, and over 600,000 in the U.S. alone (Taubenberger and Morens 2006). In contrast to other illnesses, which typically kill very young and elderly individuals, this flu disproportionately killed working age people between the ages of 20 to 40. Therefore, this begs the question: How did high mortality amongst prime working age people impact industries across the U.S over time and did these effects persist? Using census data from the Integrated Public Use Microdata Series (IPUMS) database as well as U.S. city-wide excess influenza data from Beach et al. (2022b), this paper evaluates whether the Spanish flu was associated with different industrial trajectories in the decades that followed. Specifically, I examine whether industries that employed large fractions of the affected age group, in cities with high levels of excess mortality, experienced greater industrial decline in the post-influenza period. The goal is to understand which industries flourished and declined across U.S. cities and see whether 1918 flu deaths impacted the share of people in industries over time. My results shows that higher excess influenza did not appear to substantially impact the industrial composition of prime age working individuals in the decades after the flu. At the end of the paper, there is also a discussion relating my work to the economic effects of the COVID-19 pandemic and highlighting future research opportunities. My findings may provide insights surrounding its effects as more data becomes available, considering that both the 1918 flu and COVID-19 had remarkable epidemiological similarities.

Industrial Persistence in the U.S. and the 1918 Influenza Pandemic

By

Conor Harty

Dr. Maggie Jones
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2023

Acknowledgements

I'm incredibly grateful to all of my committee members. Thank you Dr. Jones for your patience, guidance, and constant encouragement throughout the research process. Dr. Sheth, thank you for your mentorship throughout my time at Emory as well as the opportunity to work as a research assistant; this equipped me with vital skills to write this paper. Last but not least, I want to thank Dr. Lanier and Dr. Banerjee for expanding my horizons in economics and for always being there to listen throughout the brainstorming process.

Contents

Industrial Persistence in the U.S. and the 1918 Influenza Pandemic

Conor Harty

April 2023

Introduction

The 1918 flu, or "Spanish flu," left an indelible mark worldwide because of its brutality and ravenous spread. About a third of the world's population at the time, translating to about 500 million people, had the virus in only a year and experienced "clinically apparent illnesses," but these numbers could be even higher. While other influenza pandemics had a case-fatality rate of less than 0.1 percent, this flu had a rate of over 2.5 percent, 25 times higher, demonstrating its relative severity (Taubenberger and Morens 2006). In fact, the labor force was "more affected by the lethality" than in any other influenza season, and pneumonia mortality in 1918 was 300,000, whereas U.S. war casualties were only 116,000, or slightly more than a third (Velde 2022). This is quite surprising considering that historical foci in U.S. is typically on the war.

The 1918 flu also affected future influenza pandemics. In fact, all influenza A pandemics since the Spanish flu and most cases of Influenza A worldwide originated from the 1918 flu and subsequent influenzas have overlapping genes from the Spanish flu. Therefore, it is unsurprising that researchers deem the 1918 flu the "mother" of all pandemics (Taubenberger and Morens 2006).

There are a number of aspects of the 1918 flu that made it unique from other pandemics. The first is the impact it had on immigrant populations. They were hit hard across a number of different states, including Hispanics and Japanese in California and Oregon, and Mexicans in New Mexico and Texas (Fanning 2010). The pandemic was also noteworthy for the significant impact it had on young, working age people between the ages of 20 and 40 and how significant the short-term labor supply shock was due to high death rates (Beach, Clay, et al.) Industries across the U.S. employed high numbers of young working age people and immigrants and therefore this paper seeks to understand if there was a strong relationship between industrial composition and influenza deaths.

Significant research and comments about the pandemic's effects in the U.S. have been made on a nationwide scale. However, the reality is that it varied regionally, city-wide, and across industries. Through an analysis that is both city specific and industry specific, this paper hopes to demonstrate the nuances of the pandemic's economic effects across the country. My work also uses econometric analysis to understand the relationship between the flu mortality rate and the share of the prime age group in industries in the short-term, medium-term, and long-term by looking the interaction term between excess influenza and the share of prime working age individuals in 1910 for each decade studied: 1920, 1930, and 1940.

Looking at the effects of flu deaths on industrial persistence as opposed to other dependent variables is important for a few reasons. Firstly, industries that are more resilient to pandemics help sustain economic growth and development and understanding this can inform better policy and investment decisions aimed at supporting and sustaining these stable industries (Brende and Sternfels 2022). Moreover, industries that struggle to sustain themselves during pandemics can be identified and developed to become more resilient during future crises. The COVID-19 pandemic showed that industries were impacted disproportionately based on how conducive they were to remote work and how successfully they digitally transformed, Looking back at the Spanish flu could give us further insights regarding what makes industries more susceptible to lethal pandemics.

In summary, the interaction terms between excess mortality and 1910 industrial composition for multivariate regressions with dependent variables for the proportions in 1940 and 1920 including controls appear statistically significant but the interaction term between excess influenza mortality and 1910 industrial composition for the share in 1930 with controls is not. This suggests that the flu affected the industrial composition changes in the short term between 1910 and 1920 and in the long term between 1910 and 1940, but this relationship does not seem as strong in the medium term between 1910 and 1930. However, looking beyond statistical significance and at the economic significance as well shows that the interaction term coefficients are all positive and minuscule. This suggests that the 1918 flu mortality did not have a strong impact on industrial persistence, regardless of statistical significance.

Literature Review

Extensive research has been conducted on the 1918 flu and its economic impact, particularly after the COVID-19 pandemic, because people drew comparisons between the two. Velde (2022) displays the numerous impacts that the pandemic had on the economy and compares the effects to the 1920-1921 recession shortly after. He argues that the recession after the pandemic was relatively mild and short-term, but the 1920-1921 recession was more severe and long-lasting. Velde (2022) also identifies an industry that was affected significantly from flu mortality: coal. While Velde (2022) makes significant contributions in his paper including how industries, like coal, were impacted and he contextualizes the pandemic by comparing it to the 1920 recession, there are some gaps in his work. For one, he does not show how many industries, besides coal, were affected, or how industrial composition changed over time. When he does, the effects are only described in the short-term, and an industry could suffer temporarily but quickly recover. Moreover, his statistical analysis focuses primarily on the relationship between the pandemic and production levels but does not look at whether the compositional changes of various industries were affected by the pandemic itself.

Other literature on the 1918 flu has underscored the effects of the flu on pregnant women, a particularly vulnerable group during pandemics, and their children. Bloom-Feshbach et al. (2011) underscores how the flu caused first tremester miscarriages for about 10 percent of pregnant women, and Ogasawara (2017) highlights how children born between 1919 to 1920 were significantly shorter than those in other cohorts. Moreover, Beach et al. (2022a) and Almond (2006) look at how the flu affected those in utero during the pandemic. Almond (2006) describes how the flu may have led to lower educational attainment, income levels, and socioeconomic status compared to other birth cohorts. However, Beach et al. (2022a) found that in utero exposure did not significantly affect economic outcomes for this birth cohort including socioeconomic status and income. Helgertz and Bengtsson (2019) similarly looked at the impact of fetal exposure during the pandemic on health, future income, and occupation and found a relationship with higher morbidity in late adulthood, between the ages of 57-87, but not significant effects on socioeconomic outcomes in the long-term. Vollmer and Wojcik (2017) also studied the effects of the pandemic on in-utero exposure on a global level and did not find a significant effect on educational attainment, employment share, and disability levels.

While conclusions about the impact of the flu on birthing women vary, my paper connects to the existing literature and builds upon it in a few ways. Firstly, many working age individuals were pregnant and giving birth during or in close temporal proximity to the pandemic. Therefore, having parents who were affected economically by the pandemic or displaced could have led to poorer outcomes for them and their families, which could have indirectly and adversely affected their children. Moreover, while socioeconomic effects have been studied repeatedly and some have found insignificant long-term effects, industrial persistence has yet to be studied by these authors, and industrial instability could have heavily influenced future generations and where they worked.

Aassve et al. (2021) explores the long-term impacts of the flu from another angle: social trust. Using data from the General Social Survey, they found that the pandemic significantly decreased social trust and that this consequence was passed down from survivors of the flu to their offspring. Social trust is a key component of economic development because it contributes to cohesion and lower levels of conflict. It may have played a part in how industrial composition changed, and vice versa, and therefore understanding how flu mortality changed the workforce could provide further insights on this topic.

In other work, Beach et al. (2022b) notes how the pandemic caused a negative labor supply shock and how nations with higher flu mortality rates had deeper recessions. They highlight that lost production due to labor shortages was higher in states with more excess influenza mortality which suggests that the flu may have impacted industrial composition, at least temporarily.

Beach et al. (2022a) also highlights the weak mortality data due to political pressures during the war and inconsistent reporting and testing across nations, including the U.S. As a result, scholars have used excess influenza mortality, or the difference in mortality with and without a pandemic, but estimates across countries are inconsistent. In fact, Beach et al. (2022a), compiled city-level mortality which was an invaluable resource for this paper as a means of gathering sufficient variation for statistical analysis.

As mentioned, the flu had a varied impact in cities across the U.S, and showing this is a prime objective of my work. Clay et al (2018) demonstrates that there is a strong relationship between flu mortality and air pollution by looking at panel data on infant mortality and all-age mortality and cities that burned large amounts of coal. Their work shows that cities with poor air quality had over tens of thousands more deaths. Clay et al (2019) also showed that there was a relationship between high mortality rates and cities with large illiterate populations and poor prepandemic health. Grantz et al. (2016) similarly found that those in cities with high illiteracy at time, like Chicago, were at higher risk of influenza and pneumonia mortality. My paper may provide future insights about how industries located in cities with high pollution, poor health, and high pollution fared over time.

Extensive research has also been conducted on non-pharmaceutical interventions (NPIs) during the flu including social distancing measures. Bootsma and Ferguson (2007) and Markel et al. (2007) showed that there was a correlation between early intervention and lower flu mortality. However, cities across the U.S. had varied levels of success. Bootsma and Ferguson (2007) show that San Francisco, Milwaukee, and St. Louis had relatively successful measures. My work may build on these findings and help researchers isolate industries in cities across the U.S. that effectively mitigated flu deaths through NPIs and those that did not.

Data and Methods

.1

3.1 Data Sources

There were two primary data sets used in this project. The first is from IPUMS USA. This database is incredibly valuable because it provides microdata, or data at the individual-level, with demographic information about U.S. citizens. The analysis would not have been feasible if I used data aggregated at the county level. The variables relevant for this analysis were age, race, sex, occupation, and the respective industries where the respondents worked. The races I chose to include were white, black, American Indian or Alaskan Native, Chinese, Japanese, and Other Asian or Pacific Islander.

The full data set included information from 1910 to 1940 and was a one percent sample, comprising over four million observations. While the full count was available for these years, the one percent sample was large enough to provide valid statistical inferences and easier to conduct data analysis.

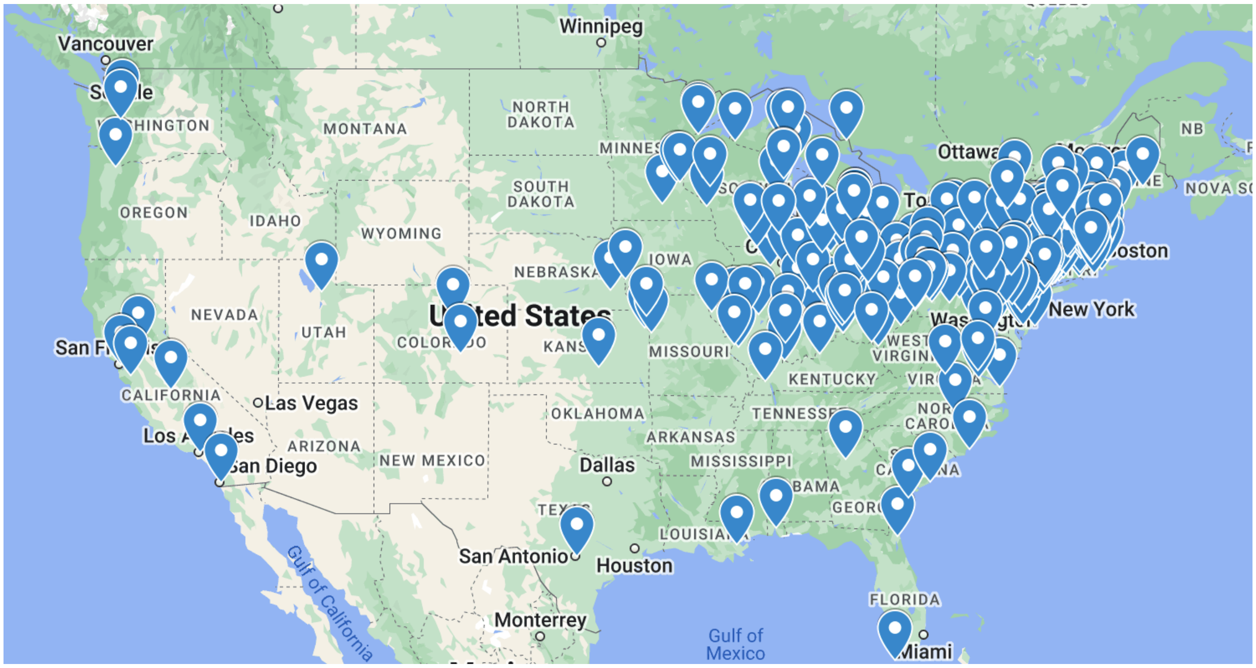


Figure 1: U.S Map of Cities with Excess Mortality Data

The map above shows the cities across the U.S. that had excess mortality data available and could be studied. The cities are primarily located on the East Coast, West Coast, and Southeast, with less concentration in the Midwest and central United States.

The other important data set used was the excess influenza mortality from Beach et al. (2022a) with excess mortality from approximately 310 cities. Excess mortality measures both indirect and direct flu deaths which captures a more realistic flu impact because many of those who died may have had underlying health conditions which were not officially attributed to the flu in death certificates and other official documents. It uses past historical flu seasons as a reference point and measures the difference between the season of interest and the past season to isolate the effects of the Spanish flu, which would presumably lead to a greater level of mortality than what is typically observed.

3.2 Data Cleaning and Processing

During the first part of the data analysis, I constructed the total number of the affected age group by city and year. Then, I constructed the total count of the relevant age group by city and industry.

Finally, I constructed city-industry-year shares of the fraction of those in the "affected" age group who worked in a specific industry and city in a specific year. Moreover, I calculated the means of the sociodemographic variables, which were coded as dummy variables, to understand the proportion of each demographic in each city and industry. While the data set contained a wide range of age groups, I wanted to explore the specific impact on 20-to-40-year olds, so I filtered the data to only include this target group. Once this portion was completed, the IPUMS data set was merged with the excess mortality data set, and the corresponding city codes, which were coded the same in both data sets, were matched. Throughout this entire process, there were some missing values (NA) for cities that were not in both data sets, and these were removed to allow for accurate data analysis.

To illustrate the goal of the empirical analysis, I chose to identify 10 cities that were representative of a few large regions of the U.S and track how their five top industries changed between 1910 and 1940. Therefore, I chose cities on the West Coast, East coast, and in the South. While it would have been ideal to view a larger range of regions, and perhaps have more representation from the Midwest, I was limited because not all cities had available data between 1910 and 1940 or excess mortality data available. The 10 cities I chose were Albany, NY, San Francisco, CA, Seattle, WA, Denver, CO, Los Angeles, CA, Washington, D.C, Waterbury, CT, New York City, NY, Atlanta, GA, and Philadelphia, PA. After choosing these locations, I identified the top five industries based on those employing the largest proportions of working age people in the respective cities. It is important to note that for many cities, the code zero, or unidentified industries, employed a a large fraction of working age people. Therefore, I chose to exclude this from the top five industries because it was not particularly informative, and regardless, the graphs illustrate my point which is that industrial composition did indeed change dramatically after the flu.

In the following portion of the code, summary statistics were created for the entire data set from all years, as well as for the subsets of 1910, 1920, 1930, and 1940. These were produced for age, sex, excess flu mortality, and the aforementioned race variables; these will be discussed in further detail later in this paper.

3.3 Descriptive Statistics

The table below shows the means and standard deviations for the full data set with all variables and years, and the specific data sets for 1910 through 1940. For instance, based on the full data set, about 49.5 percent of those in the sample are male and 94.1 percent of them are white because these were coded as dummy variables and therefore their interpretations are proportions. Furthermore, the mean age for the same data set is about 28.61 and the mean excess influenza mortality is about 38.38. The values below the means in parenthesis are the standard deviations. For illustrative purposes, the standard deviation of the sex variable is 0.0277 and the standard deviation of the white variable is 0.0944. Therefore, the values for the sex variable are clustered more closely around the mean, as indicated by the smaller standard deviation, whereas those for the white variable are more heavily spread around the mean.

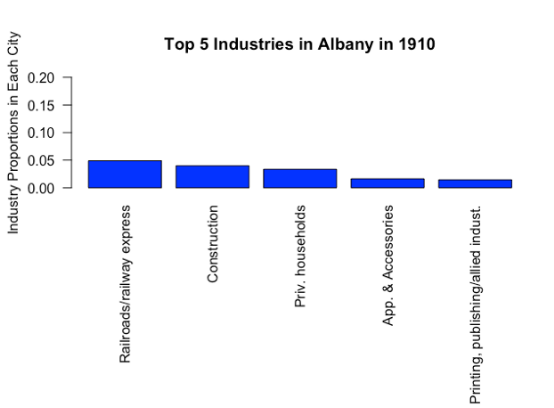
3.3.1 Summary Statistics

Table 1: Summary Statistics for Correlates

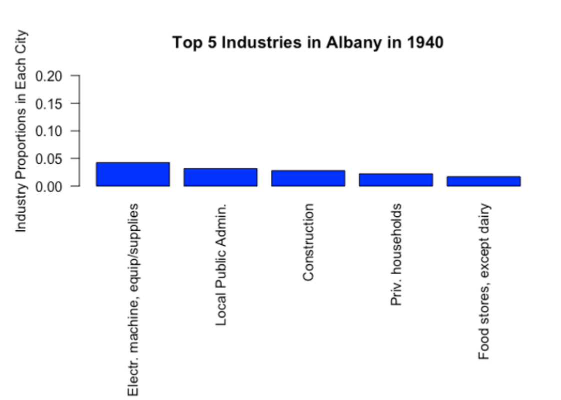
	Full Data Set	**1910**	**1920**	**1930**	**1940**
Sex (male)	0.495	0.499	0.497	0.4927	0.4879
	(0.0277)	(0.033)	(0.026)	(0.025)	(0.014)
White	0.941	0.949	0.942	0.945	0.910
	(0.0944)	(0.099)	(0.095)	(0.0855)	(0.10)
Other Asian or Pacific Islander	0.0002	0	0.001	0.0002	0.0005
	(0.0007)	(0)	(0.006)	(0.0009)	(0.001)
Black	0.05	0.048	0.054	0.051	0.085
	(0.0007)	(0.09)	(0.09)	(0.085)	(0.101)
Japanese	0.001	0.0007	0.0014	0.001	0.0017
	(0.005)	(0.003)	(0.005)	(0.005)	(0.006)
Chinese	0.001	0.001	0.001	0.008	0.0014
	(0.005)	(0.007)	(0.004)	(0.0035)	(0.0039)
Native American/Alaskan	0.0001	0.0005	0.00008	0.0002	0.0001
	(0.0012)	(0.0003)	(0.0004)	(0.002)	(0.0003)
Age	28.61	26.66	27.497	28.80	30.62
	(19.72)	(19.07)	(19.479)	(19.749)	(20.14)
Excess influenza	38.38	37.32	39.70	37.58	40.546
	(4.32)	(27.12)	(28.10)	(26.82)	(30.69)

3.3.2 Graphs

These graphs illustrate how the industries employing the top five largest proportions of people in the prime age group in 10 cities across the U.S. changed in the subsequent decades after the flu. These graphical representations make it easier to visualize how industries changed over 30 years. Looking at Albany, for instance, we can see that railroads and railway express service employed the largest proportion of this age group in 1910, whereas printing, publishing, and allied industries employed the smallest proportion that same year. However, by 1940, this changed entirely. Electrical machines, equipment and supplies eventually employed the largest share in 1940, whereas food stores excluding dairy employed the smallest fraction. While this was not necessarily caused by the flu, it will be interesting to explore whether flu deaths contributed to this, and if so, how significantly.

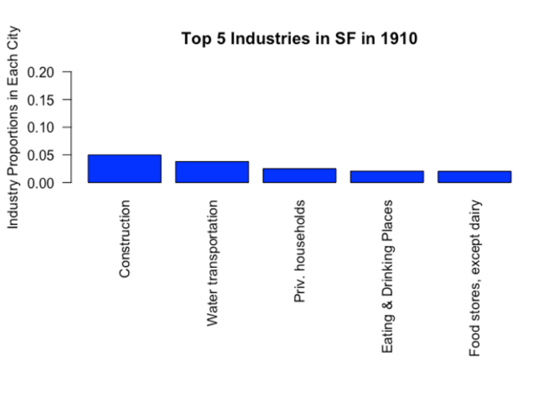


(a) Albany 1910

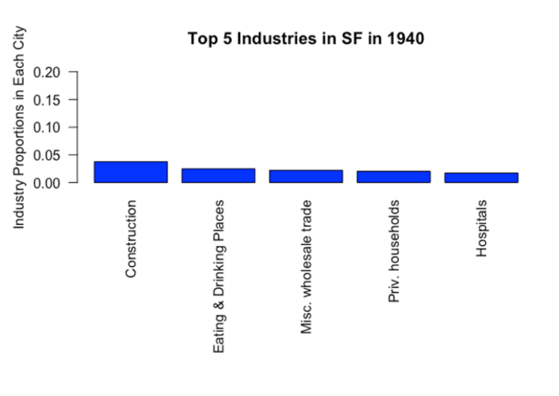


(b) Albany 1940

Figure 2: Industrial Composition Changes 1910-1940

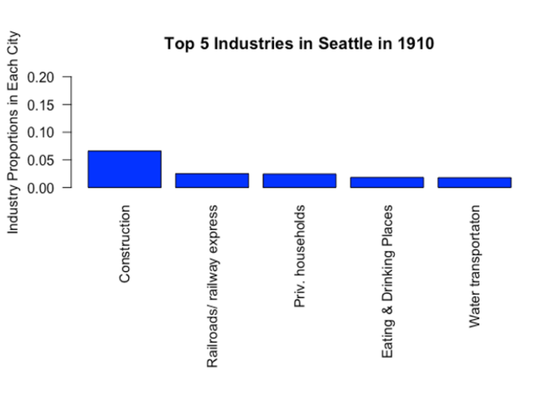


(a) SF 1910

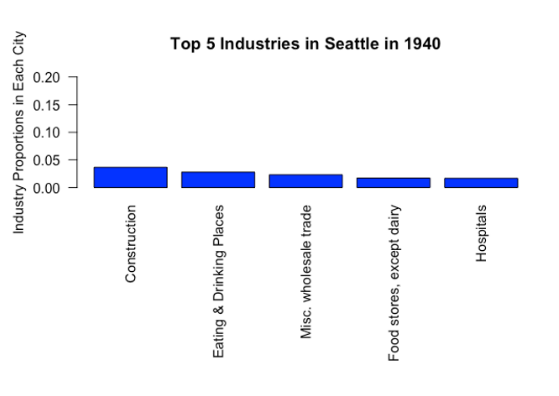


(b) SF 1940

Figure 3: Industrial Composition Changes 1910-1940

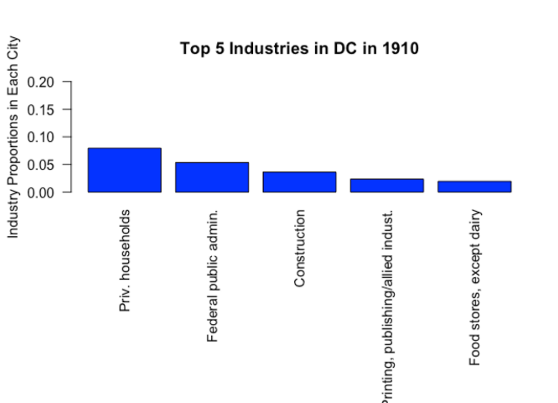


(a) Seattle 1910

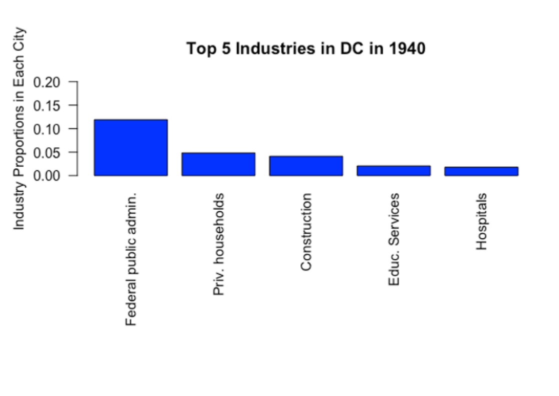


(b) Seattle 1940

Figure 4: Industrial Composition Changes 1910-1940

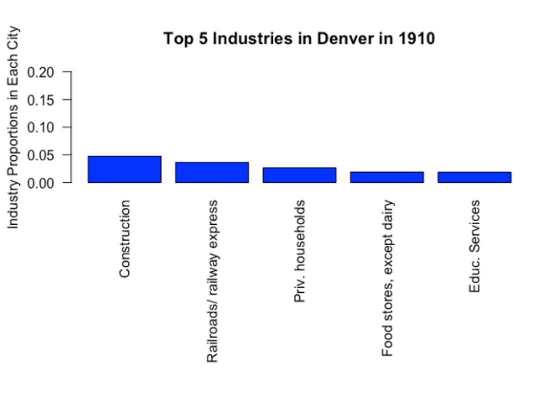


(a) Washington D.C. 1910

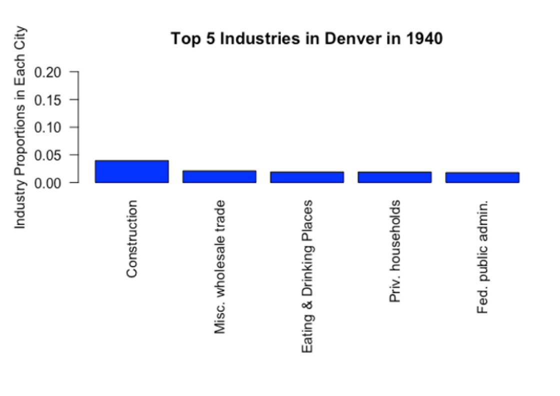


(b) Washington D.C. 1940

Figure 5: Industrial Composition Changes 1910-1940

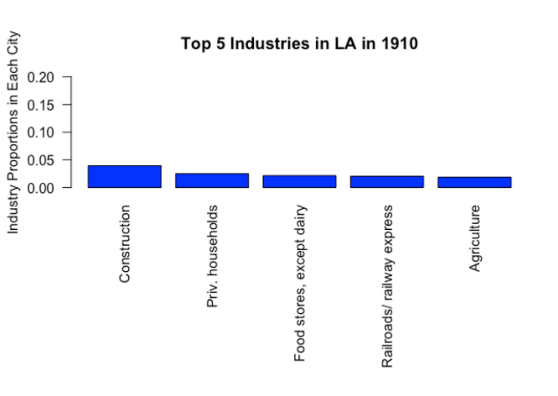


(a) Denver 1910

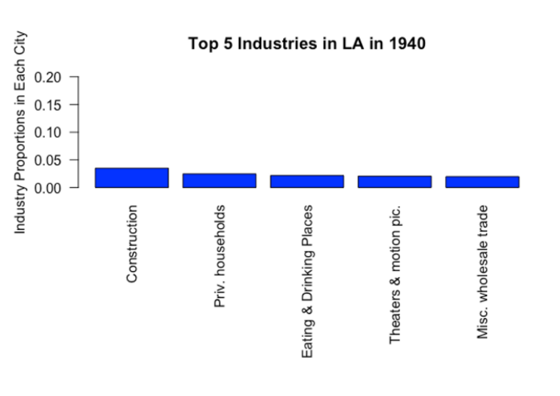


(b) Denver 1940

Figure 6: Industrial Composition Changes 1910-1940

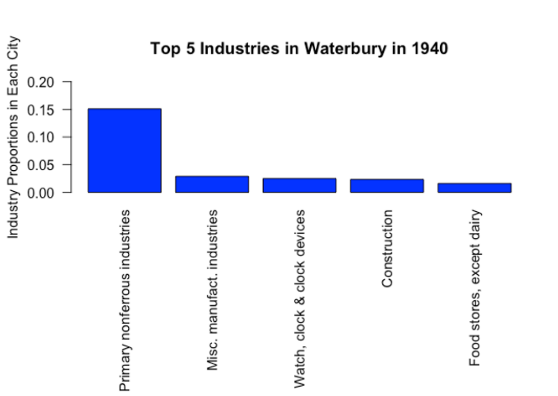


(a) LA 1910

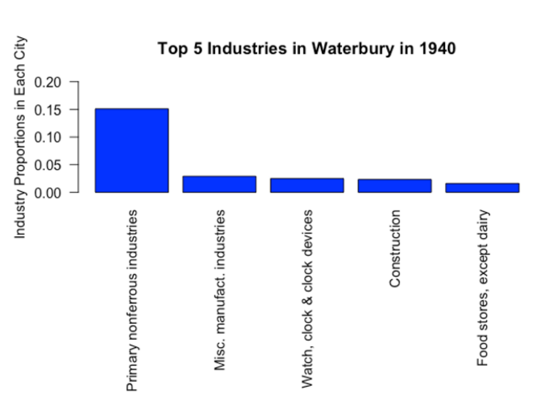


(b) LA 1940

Figure 7: Industrial Composition Changes 1910-1940

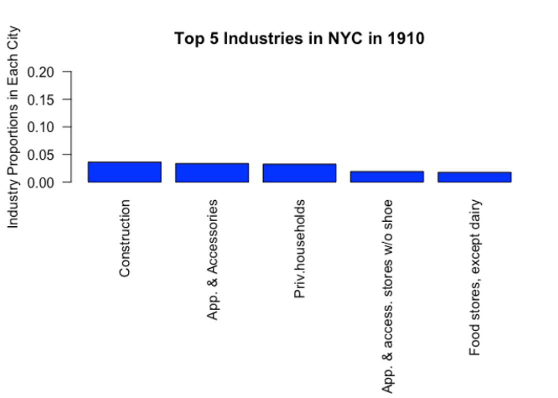


(a) Waterbury 1910

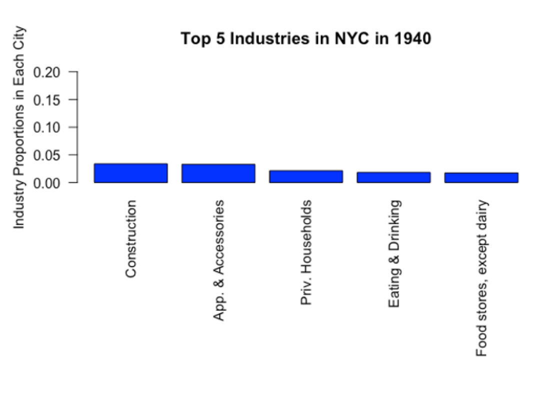


(b) Waterbury 1940

Figure 8: Industrial Composition Changes 1910-1940

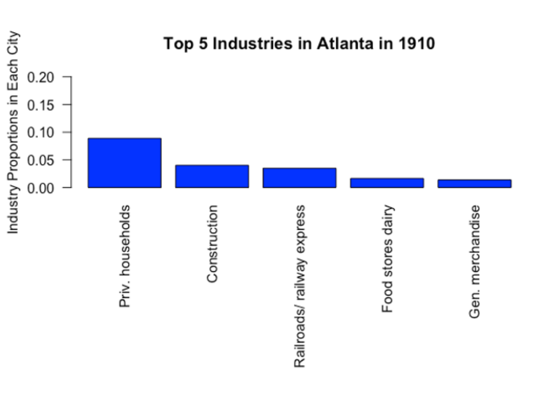


(a) NYC 1910

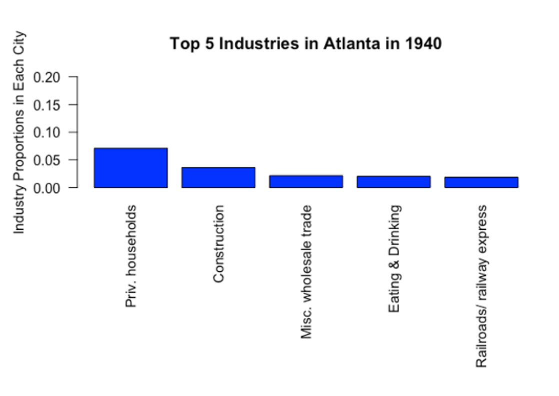


(b) NYC 1940

Figure 9: Industrial Composition Changes 1910-1940

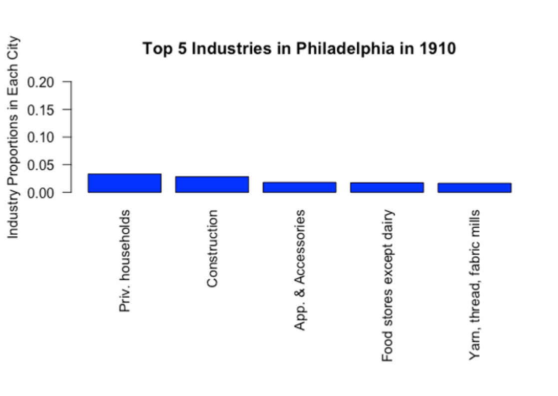


(a) Atlanta 1910

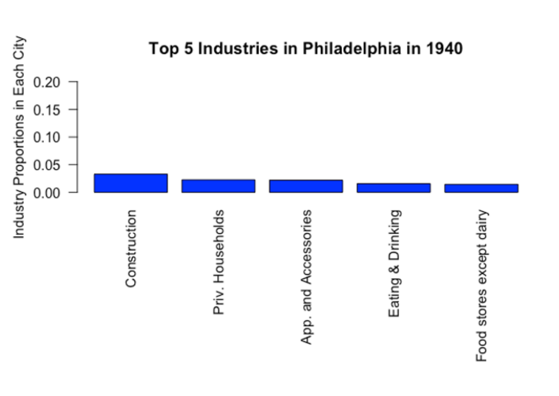


(b) Atlanta 1940

Figure 10: Industrial Composition Changes 1910-1940



(a) Philadelphia 1910



(b) Philadelphia 1940

Figure 11: Industrial Composition Changes 1910-1940

3.4 Methodology

The primary methodology used in this paper is an ordinary least squares (OLS) specification and I had two objectives. The first model explores the correlation between Spanish flu excess mortality and the baseline 1910 covariates including race and sex. While this does not show causation, it highlights relationships between the covariates and flu mortality.

The second model also used OLS and was created to understand the relationship between the influenza pandemic and the industrial composition of the labor market. Specifically, I estimate a number of controls, and dependent variables such as excess influenza mortality and the share of individuals working in specific cities and industries between 1910 and 1940. By including these controls, I am limiting the confounding effects of other factors and attempting to isolate the effects of individual variables. In reference to objective one, I estimated specifications that take the following form:

$$Excess\ Mortality_C = \alpha + X_C\beta + U_c \ (1)$$

$$\boxed{\beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}$$

Based on the equation above, $Excess\ Mortality_C$ is the excess mortality in city "c" during the pandemic. Moreover, X_C is a matrix of controls, including sex and race and U_c is the error term. All of the controls are computed as of 1910.

In the second model, I examine whether the pandemic affected the industrial composition trajectory of the workforce. Specifically, I estimate the following:

$$Share_{ic1920} = \alpha + \beta 1 share_{ic1910} + \beta 2 excessmortality_C + \beta 3 share_{ic1910} * excessmortality_C + X_c \mu + \varepsilon_c$$

Based on the second equation, $Share_{ic1920}$ is the share of prime working age people working in industry "I" in city "c" in 1920, $\beta 2 excessmortality_C$ is the excess mortality, and $\beta 3 share_{ic1910} * excessmortality_C$ is the interaction term of excess mortality and the share of the target age group in 1910. The remaining covariates, represented by the term $X_c \mu$, control for the aforementioned race and sex variables.

After calculating the equation for the share of the relevant age group in 1920, I calculated it for 1930 and 1940. All specifications were estimated with and without controls. Because my research question is how did the 1918 flu impact industrial persistence in the short-term and long-term, I specifically evaluated the coefficient on the interaction term $\hat{\beta}_3$. While other coefficients are informative, this one is relevant to my research question specifically because it tells us what the impact of one more unit of excess influenza is on the relationship between the share of the working age population in city c in 1910 and the industrial composition in each subsequent decade. I estimated models with and without controls to alleviate concerns that may arise from confounding factors.

Results

4.1 Correlates of Excess Mortality

Table 2: Demographic Variables and their Correlation with Excess Mortality

[-1.8ex]		Excess Influenza:				Final Model:
	(1)	(2)	(3)	(4)	(5)	(6)	
[-1.8ex] **Panel A: Correlates**							
Black.mean	-64.595^{***}						-27.202^{***}
	(0.249)						(3.849)
White.mean		64.605^{***}					35.574^{***}
		(0.250)					(3.848)
AmericanIndianandAlaskanNative.mean			$-6,286.114^{***}$				$-8,765.584^{***}$
			(84.837)				(84.747)
Chinese.mean				53.701^{***}			
				(3.719)			
Japanese.mean					171.685^{***}		3.568
					(7.909)		(11.003)
SEX.mean						78.950^{***}	66.975^{***}
						(0.814)	(0.870)
Constant	40.155^{***}	-24.330^{***}	37.246^{***}	36.867^{***}	36.825^{***}	-2.632^{***}	-28.570^{***}
	(0.027)	(0.238)	(0.025)	(0.025)	(0.025)	(0.409)	(3.865)
Observations	854,600	854,600	854,600	854,600	854,600	854,600	854,600
R^2	0.073	0.073	0.006	0.0002	0.001	0.011	0.089
Adjusted R^2	0.073	0.073	0.006	0.0002	0.001	0.011	0.089
Residual Std. Error	22.118 (df = 854598)	22.123 (df = 854598)	22.899 (df = 854598)	22.970 (df = 854598)	22.966 (df = 854598)	22.847 (df = 854598)	21.928 (df = 854594)
F Statistic	$67,338.280^{***}$ (df = 1; 854598)	$66,916.960^{***}$ (df = 1; 854598)	$5,490.227^{***}$ (df = 1; 854598)	208.469^{***} (df = 1; 854598)	471.251^{***} (df = 1; 854598)	$9,413.562^{***}$ (df = 1; 854598)	$16,677.150^{***}$ (df = 5; 854594)
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Table 2 shows the correlation between excess influenza for each of the correlates individually and then the final model controls for each of the demographic variables, excluding Chinese, to avoid collinearity. One noteworthy observation is that before anything is controlled for, all the independent variables are statistically significant at the 0.01 level, signified by the three asterix. This indicates that there is a strong correlation between various races, sex and excess influenza deaths.

While most variables have summary statistics available, the Other Asian or Pacific Islander predictor was removed from the model previously. This is because it is was perfectly collinear with the other race variables. Therefore, it is dropped and absorbed by the intercept term because it is not informative for the model.

Looking at the first row of the column, the slope is approximately -64.595 and the intercept is estimated to be about 40.155. This indicates that as you increase the proportion of black people by one percentage point, the number of excess influenza deaths decreases by an average of 64.595. Additionally, the intercept suggests that a city with no black people had an average number of influenza deaths of about 40.154. This is interesting because it indicates that there is a negative relationship between the number of black people and influenza cases in 1910 when there are not controls.

Moving to the second row and specifically the white coefficient value tells a different story. It has a coefficient value of 64.605 and an intercept equal to -24.330. This suggests that increasing the proportion of white people by one percentage point leads to an increase in the number of influenza deaths by an average of 64.605, and in cities with zero white people, the number of influenza deaths was about -24.33. A negative value is not possible in this context, and therefore the intercept cannot be interpreted. However, the slope value suggests that there is a strong positive correlation between the number of white people in a city and the level of excess influenza mortality.

The values associated with the race variable for American Indians and Alaskan Natives provides valuable insights regarding correlation as well. Looking at the third row, the intercept estimate is 37.246 and the slope value is -6,286.114. This indicates that a one percentage point increase of this racial group leads to a drop in influenza cases by 6,286.114, and that cities with zero American Indians and Alaskan Natives had an average of 37.246 excess influenza mortality cases. There is a similar negative relationship between this racial group and excess influenza and the relationship previously shown between excess influenza and black Americans. However, this slope coefficient is significantly larger than any of the coefficients studied

On the other hand, the Chinese race variable shows quite a different relationship from both of these racial groups and is more similar to the white variable. The slope estimate is 53.701 and the intercept is 36.867, indicating a positive correlation between the predictor and response variable, where a one percentage point increase in the number of Chinese people leads to an increase in the average number of excess influenza mortality cases by 53.701. Regarding the intercept, it demonstrates that cities with zero Chinese people had average excess influenza mortality of about 36.867 cases.

The estimates for the Japanese variable's slope coefficient and intercept are similar to the Chinese variable as well. With a slope estimate of about 171.685 and a starkly similar intercept of 36.825, this model conveys a positive relationship between the predictor and the response variable and suggests that a one percentage point increase in the number of Japanese people leads to an increase in number of excess influenza deaths by 171.685. It also suggests that cities with zero Japanese people had average excess flu mortality of about 36.825.

The last value in the table of correlates is for the sex variable. Its intercept is -2.632 and the slope is 78.95. This suggests that if the mean for sex is zero, or the number of females and males is equal, the average number of influenza cases is expected to be about -2.632, which is not able to be interpreted in this context. Moreover, the slope coefficient suggests that a one percentage point increase in the number of males leads to a decrease in excess influenza mortality by about 79 cases. These estimates reveal a negative relationship between men in cities in 1910 and the subsequent excess influenza mortality rate.

The final model looks at the relationship between each of the race and sex variables, save for the Chinese variable and Other Asian and Pacific Islander value, as previously mentioned. However, this time it controls for each to understand the relationship between influenza deaths and the proportion of people in industries in the decades after the flu. It appears that all coefficient estimates and intercepts are statistically significant at the 0.01 level, except for the Japanese variable. This is intriguing because the Japanese variable was statistically significant when the correlates were studied previously, but after controlling for other variables, it no longer is.

One can also look further into the coefficient and intercept interpretations for this model. The intercept value of -28.570 suggests that when all the values of the variables are 0 in 1910, the mean number of excess influenza deaths is approximately -28.570. However, this is not particularly informative or realistic because it describes a city where no individuals of any of the races in the model are present, but where there is an equal number of men and women.

I will now delve into the slope coefficient interpretations. The sex coefficient value is about 66.975, meaning that the predictor and response variable have a positive relationship with one another, and increasing the proportion of men in 1910 by one percentage point leads to an increase in the excess influenza mortality rate by 66.975 cases on average. Furthermore, the black coefficient is -27.202, showing a negative relationship between the predictor and response variable. This means that an increase in the number of black people by one percentage point in 1910 leads to a decrease in the average average number of excess influenza mortality cases by 27.202.

Furthermore, the American Indian and Alaskan Native coefficient value is -8,765.584. It was negative without controls, but this value is significantly larger than it was previously. This illustrates that the result of a one percentage point increase of this racial group drops excess influenza mortality by an average of almost 9000 cases; again, this value is dramatically higher than any of the others studied. Lastly, the slope coefficient for the white group is approximately 35.574, indicating that increasing the number of white people in 1910 by one percentage point leads to an increase in average excess influenza mortality by about 35.574 cases.

4.2 Main Results

Table 3: Excess Influenza and Share of Working Population in Industries

	Dependent variable:					
	Share 1920	Share 1930	Share 1940	Share1920 with Controls	Share1930 with Controls	Share1940 with Controls
	(1)	(2)	(3)	(4)	(5)	(6)
Share1910	0.992***	1.007***	0.996***	0.992***	1.007***	0.996***
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
excess influenza	−0.00000	−0.00000	−0.00000	−0.00000	−0.00000	−0.00001
	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00001)
SEX				−0.0004	0.004	−0.001
				(0.004)	(0.005)	(0.007)
White				−0.440	−0.475	−0.325
				(0.559)	(0.609)	(0.645)
Japanese				−0.437	−0.478	−0.315
				(0.568)	(0.618)	(0.653)
Black				−0.441	−0.474	−0.327
				(0.559)	(0.609)	(0.646)
Chinese				−0.454	−0.479	−0.348
				(0.560)	(0.610)	(0.646)
American Indian and Alaskan Native						
Share1910:excessinfluenza	0.0003***	0.00004	0.0004***	0.0003***	0.00004	0.0004***
	(0.00005)	(0.0001)	(0.0001)	(0.00005)	(0.0001)	(0.0001)
Constant	−0.0003*	−0.001***	−0.0001	0.440	0.472	0.326
	(0.0002)	(0.0002)	(0.0002)	(0.560)	(0.609)	(0.646)
Observations	9,079	8,072	3,830	9,079	8,072	3,830
R^2	0.984	0.984	0.985	0.984	0.984	0.985
Adjusted R^2	0.984	0.984	0.985	0.984	0.984	0.985
Residual Std. Error	0.010 (df = 9075)	0.010 (df = 8068)	0.009 (df = 3826)	0.010 (df = 9070)	0.010 (df = 8063)	0.009 (df = 3821)
F Statistic	185,766.900*** (df = 3; 9075)	166,736.100*** (df = 3; 8068)	82,030.820*** (df = 3; 3826)	69,643.380*** (df = 8; 9070)	62,505.170*** (df = 8; 8063)	30,755.030*** (df = 8; 3821)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 3 shows the relationship between excess influenza mortality and the proportions of the working age population employed in industries across the U.S. in the decades after the flu. First we look at three models without controls and the final models with controls to understand whether higher influenza led to a lower share of people working in these industries in the short-term, medium-term, and long-term. The independent variable on the left includes the share of people in the working age group in 1910, excess influenza, the interaction term between excess influenza and the share of people working in 1910, and the relevant demographic variables. The dependent variables on the top of each column, starting in the first three columns from the left, are the proportions of people in subsequent decades after the flu without controls. The last three columns are the same, but with added controls.

In column one, we see the relationship between working age individuals in 1920, excess influenza, those in industries in 1910, and the interaction of the two. The small positive coefficient of the slope for the share of those working in 1910 as well as the small positive interaction term of this variable and excess influenza, are statistically significant at the 0.01 level. This suggests that there is a positive relationship between the share of those employed in industries in 1910 and the share of those same people working in 1920. This is unsurprising considering that at least some of those employed in 1910 were likely working in the same industries in 1920. Moreover, the very small but statistically significant interaction term suggests that as the number of excess influenza mortality cases increases by one, the industrial composition increases by a proportion of 0.0003, or 0.3 percent, between 1910 and 1920. While the interaction term is statistically significant, it is very small and not economically significant.

The next two columns are very similar to the first including the same interaction term. However, instead, the response variable is the proportion of working age individuals in U.S industries in 1930 and 1940, respectively. The second column has only two statistically significant terms: the intercept and the share of working age people in industries in 1910. The interaction term is not statistically significant here, and similar to the previous column, neither is excess influenza mortality. This indicates that higher levels of influenza do not make the relationship between the share of people in industries in 1910 and the share in 1930 stronger.

Looking at the third column, the proportion of people in different industries in 1910 is statistically significant and the interaction term is as well. The interaction term has a positive, yet very small value, of approximately 0.0004. This indicates that as the number of excess influenza mortality cases increases by one, the industrial composition increases by a proportion of 0.0004, or 0.4 percent, between 1910 and 1940.

The last three columns show the final adjusted models for the first three and include controls for race and sex variables. I will now explore how the statistical significance and estimates changed as a result. Beginning with the fourth column, the only two statistically significant values are the variables for shares of people in 1910 and the interaction term. Therefore, the statistical significance of these two variables remained even when controls were added. This indicates that higher rates of influenza do indeed increase the strength of the relationship between the share of working age people in industries in 1910 and the share in 1920, and increasing excess influenza cases by one increases the industrial composition change between the same industries by 0.0003, or 0.3 percent, between 1910 and 1920. However, this is not intuitive, because one might expect an increase in influenza deaths to lead to a decrease in the industrial composition between the same industries. Regardless, it is important to note that the magnitude of this value is still relatively small (0.0003), and therefore it is not economically significant.

Looking at the penultimate multivariate regression model, the only statistically significant value is the proportion of people in industries in 1910, and with added controls, the interaction term is still not statistically significant. This suggests that higher levels of influenza do not strengthen the relationship between the share of people in industries in 1910 and the share in 1930.

Lastly, one can observe the final model looking at the share of people in industries in 1940 with controls. The only two variables that are statistically significant at the 0.01 level are the proportion of people in industries in 1910 and the interaction term. The interaction term stayed approximately the same as the model for 1940 without controls (0.0004). This value indicates that increasing excess influenza cases by one leads to an increase in the industrial composition between 1910 and

1940 by a proportion of 0.0004, or 0.4 percent. Similar to the first full model in column four, this interpretation is not intuitive and one might expect the interaction term to be negative. However, this value is incredibly small and not economically significant; this might explain why it has a positive sign.

One question that could be asked about these results is whether project-based industries, such as the railroad industry, which are inherently temporary, could have been significant in the short term but then dissipate when a project finished. This would lead to a large proportion of working age people in industries in 1910 and a small proportion in 1920, 1930, and 1940, or vice versa. This can be explored in the scatter plots below.

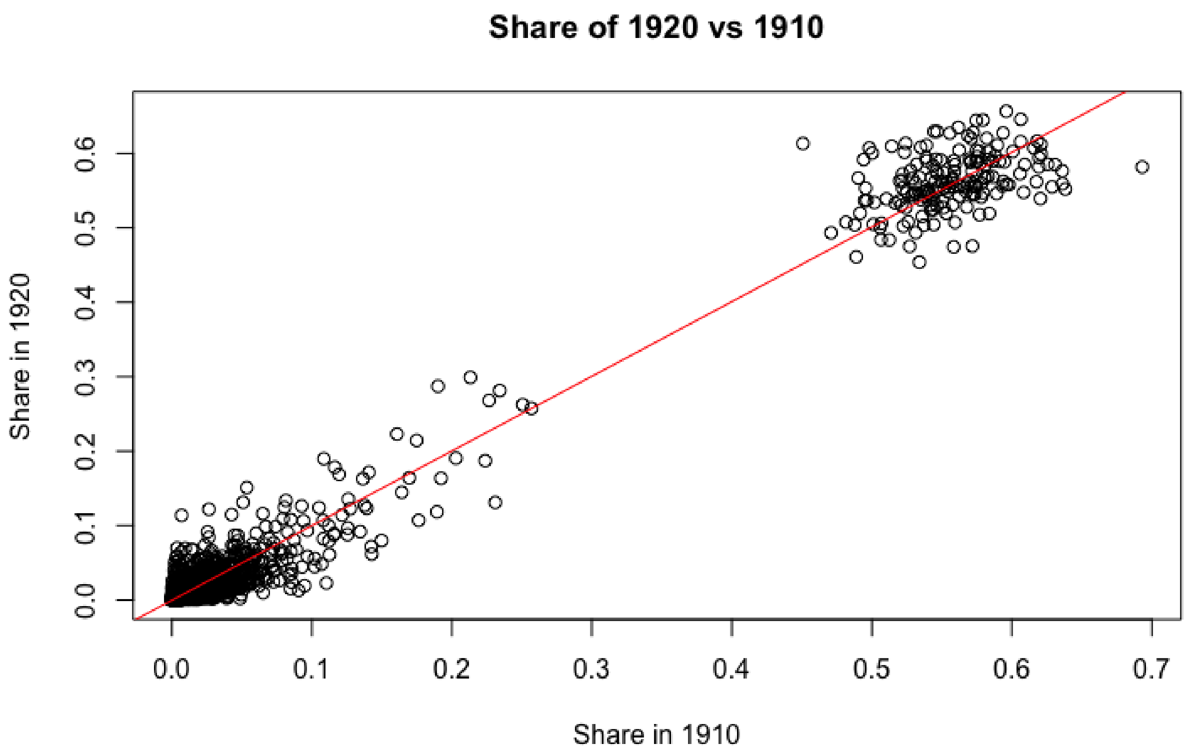


Figure 12: Scatterplot 1: 1910 and 1920 Shares

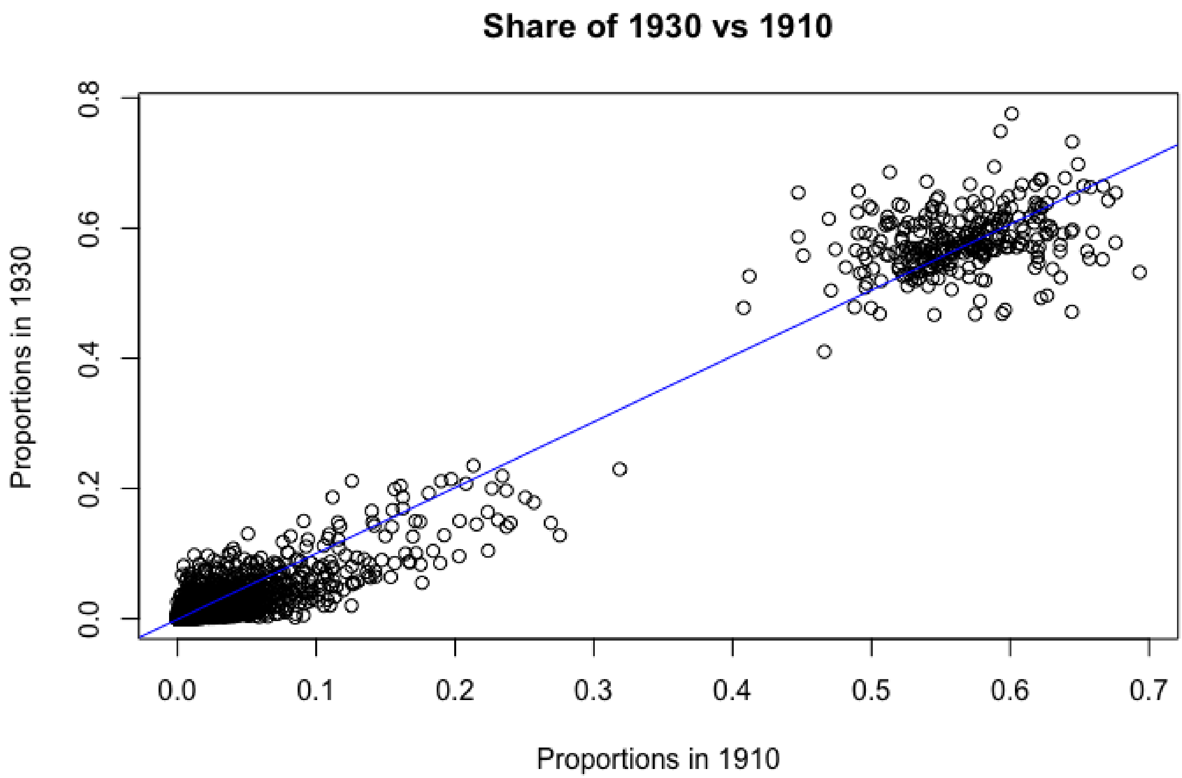


Figure 13: Scatterplot 2: 1910 and 1930 Shares

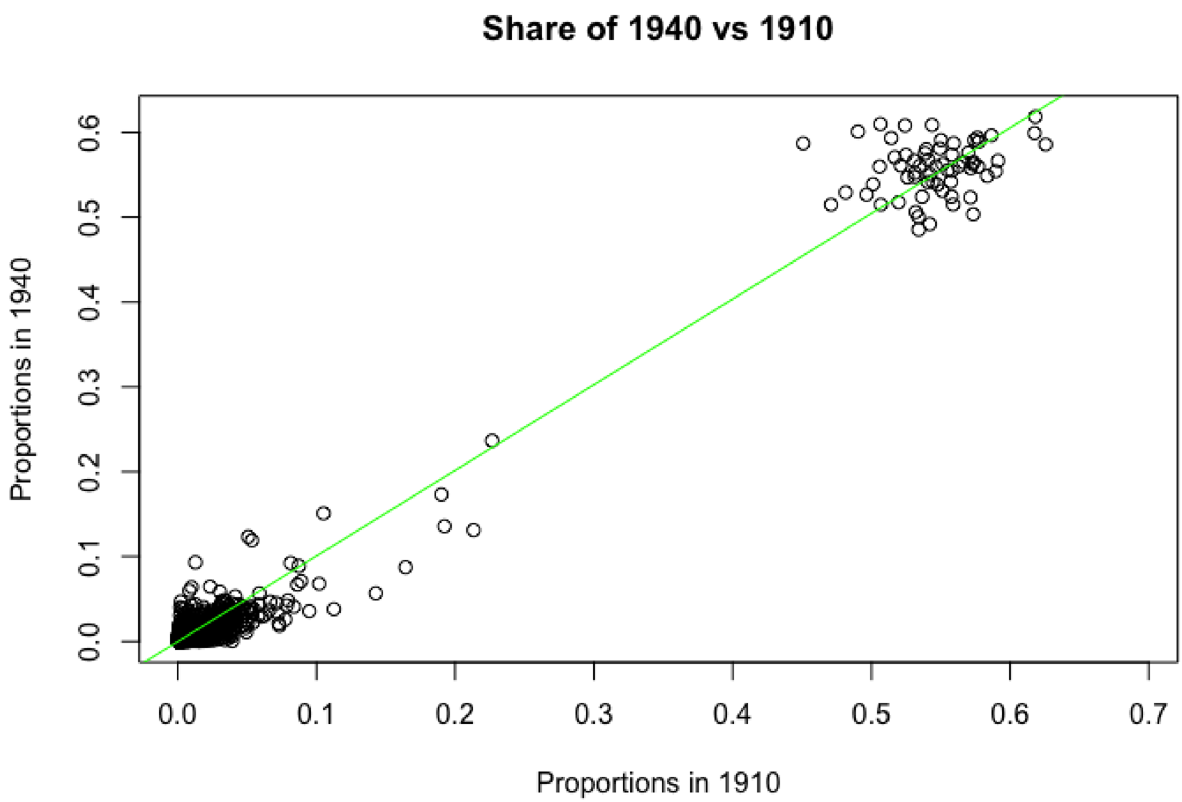


Figure 14: Scatterplot 3: 1910 and 1940 Shares

While there are some outliers in these graphs, there are relatively few. The plots show that the relationship between the share of industries in 1910 and the subsequent decades after the flu is a linear fit and that the majority of industries between 1910 and 1940 remained in existence. Therefore, while some industries may have disappeared or appeared abruptly between 1910 and 1940, it is not a strong concern.

Conclusion

The Spanish Flu was one of the most catastrophic events in U.S. history. It killed large numbers of immigrant populations and young people and had a significant economic impact. The goal of this paper, through a combination of econometric analysis and data visualization, is to look at how the flu impacted industrial persistence between 1910 and 1940. While there has been extensive research on this pandemic, especially after COVID-19, I saw a gap in the literature: researchers have neglected to use the city-wide data available to look at the relationship between flu deaths and industrial composition over time. Given that experts suspect pandemics will become more common as global warming increases and biodiversity wanes (Costley 2022), it is especially important to investigate the economic impacts of pandemics thoroughly. Moreover, as previously highlighted, there were striking similarities between the behavior of COVID-19 and the Spanish Flu. Therefore, my hope is that this paper and the corresponding research can be used in the future to look at the short-term and long-term economic effects of COVID-19 when more time has passed and more industry specific data is available.

My results show that although both the first and last multivariate models' interaction terms are statistically significant at the 0.01 level, they are not economically significant and therefore the 1918 flu did not appear to have a strong impact on industrial composition in the short-term, medium-term, or long-term. It is essential to not overly weight statistical significance and to look at economic significance as well. This likely explains why both statistically significant interaction terms for the first and last model had unexpected positive signs, suggesting that increased influenza mortality could lead to higher levels of the prime age group working in the same industries over time. The fact that the second full multivariate model's interaction term was not statistically significant also suggests that the flu was not a strong contributor to which industries flourished and which ones declined over the 30 year period between 1910 and 1940 because that would mean that the flu deaths were impactful in the short-term and long-term, but not in the medium term, which is illogical.

While my research provides valuable insights, it is important to acknowledge the limitations of the data and aspects of my research process that may undermine the conclusive nature of these results. The first is the lack of variables, coded as "N/A" throughout the full data set with demographic variables between 1910 and 1940. In order to create functioning statistical models, these needed to be removed which may have introduced bias into the results that is difficult to target and measure. Furthermore, a large proportion of industrial composition was coded as industry "0," which is a category for unknown industries that were not included in the top 5 industry graphs. Therefore, the tracking of the ten cities and how their top industries changed may have been slightly inaccurate because this category was removed and some of the values for this industry may have fit into other measured categories, but this information is unknown.

Another significant limitation of this study is the lack of mortality data available. Poor data tracking and a lack of technical sophistication prevents us from having access to precise, accurate 1918 flu deaths across the U.S. Therefore, excess influenza mortality, while it is the best measure we have, may not be measuring Spanish flu deaths precisely; it may pick up deaths from other similar illnesses and thus may improperly measure the impact of Spanish flu deaths.

Furthermore, as mentioned before, when the excess mortality data set was matched with the IPUMs primary data set, only cities with excess flu mortality data were matched and others were dropped. This limited my ability to graphically show the industrial persistence for various cities over time and also limited the sample size for cities on which I conducted econometric analysis. Having more city-wide mortality data available may have provided greater variation and stronger statistical analysis, and thus, more conclusive results.

Another limitation of this paper was the difficulty controlling for a multitude of variables that likely impacted industrial composition between 1910 and 1940. For instance, political events such as the 1920-1921 Recession and the Great Depression, as well as policies like the New Deal certainly impacted the workforce. It was especially challenging to study the effects of technological changes that could have benefited some industries and led to the decline of others. If these were controlled for, my model could have better isolated the effects of flu mortality and potentially showed different results.

There are numerous future research opportunities to build upon my findings in this paper. While I looked at the impact of influenza mortality on the U.S. workforce, one could also look at the impact in other countries that saw high deaths rates including Italy (Berbenni and Colombo 2021). Perhaps the relationship I studied is different and other countries potentially collected precise mortality data that could provide more robust results. Moreover, the 1918 flu and COVID-19 were similar in terms of their spread, but the age groups disproportionately affected were very different. While working age people suffered during the Spanish flu, during COVID-19, working age people were the most likely to survive during COVID-19 (Beach et al. 2022a). Instead, older individuals, particularly those in leadership and higher paying positions, suffered. Therefore, it would be intriguing to isolate the effects of COVID-19 on 50-65-year-olds over a few decades (McKinsey and Company 2023). Regardless of the direction taken, there are myriad methodological approaches that can be taken to expand on my paper and it is imperative that we thoroughly assess the economic impact of pandemics as they become the new reality.

The rise of remote work due to the COVID-19 pandemic also has enormous implications for industries and groups that may be more severely impacted during future pandemics. Barrero et al. (2021) show that an estimated 20 percent of full workdays will be remote when the pandemic ends compared to only 5 percent before. Technological advances have made working from home more feasible for some and social stigma around remote work has decreased. However, remote work also increased inequalities between groups. A recent survey shows that half of the top earners in 2019 always worked remotely, whereas only 20 percent of those in the bottom 20 percent of income said the same. Furthermore, those who could not work remotely were significantly more likely to report "deteriorating" working conditions, making them more susceptible to pandemic-related illnesses (Lavietes 2021).

Pandemic related layoffs also skewed more heavily toward low earning, minority groups. The same survey showed that Hispanics, African Americans, and those with lower income and education levels suffered higher job losses. Indeed, over 40 percent of Americans with incomes in the bottom 20 percent, Hispanic and multiracial individuals, and people without college degrees said they lost their jobs compared to 30 percent of those surveyed overall (Lavietes 2021).

Household technological barriers will also lead to unequal effects of future pandemics. A Pew Research survey from 2019 highlighted that 82 percent of White adults in the U.S. owned a computer, while only 58 percent of African Americans and 57 percent of Hispanics did. Additionally, 79 percent of White adults said they had broadband internet, while 66 percent of African Americans and 61 percent of Hispanics said the same (Johnson 2020).

Furthermore, COVID-19 showed that only some industries were well-suited for teleworking. Pew research showed that only four of the nine industries studied were able to perform most of their work from home. 84 percent of those in banking, finance, accounting, real estate and information and technology said they could work from home seamlessly. However, these numbers dropped off enormously when those in other industries were studied. A mere 33 percent of those in healthcare and social assistance, 23 percent in hospitality, service arts, entertainment and recreation, and 16 percent in retail, trade and transportation responded yes to the same question. (Minkin 2021).

It is clear that the effects of future pandemics will not be equal across industries, socioeconomic class, or race unless we make significant changes and minimize the effects for all groups. Therefore, policymakers must make drastic changes to address systemic inequalities and avoid feeding inequality through inaction.

Bibliography

References

[1] Aassve, Arnstein, et al. “Epidemics and Trust: The Case of the Spanish Flu.” *Health Economics*, vol. 30, no. 4, Apr. 2021, pp. 840–57, https://doi.org/10.1002/hec.4218.

[2] Almond, Douglas. “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*, vol. 114, no. 4, Aug. 2006, pp. 672–712, https://doi.org/10.1086/507154. Accessed 23 Mar. 2023.

[3] Barrero, Jose Maria, et al. “Why Working from Home Will Stick.” NBER Working Paper No. 28731, Apr. 2021, https://doi.org/10.2139/ssrn.3741644. Accessed 8 Apr. 2023.

[4] Beach, Brian, et al. “Re-Evaluating the Long-Term Impact of in Utero Exposure to the 1918 Influenza Pandemic.” *Journal of Political Economy*, vol. 130, no. 7, May 2022.

[5] —. “The 1918 Influenza Pandemic and Its Lessons for COVID-19.” *Journal of Economic Literature*, vol. 60, no. 1, Mar. 2022, pp. 41–84, https://doi.org/10.1257/jel.20201641. Accessed 17 Sept. 2022.

[6] Berbenni, Enrico, and Stefano Colombo. “The Impact of Pandemics: Revising the Spanish Flu in Italy in Light of Models’ Predictions, and Some Lessons for the Covid-19 Pandemic.” Journal of Industrial and Business Economics, vol. 48, no. 2, Feb. 2021, pp. 219–43, https://doi.org/10.1007/s40812-021-00182-1. Accessed 8 Apr. 2021.

[7] Bloom-Feshbach, Kimberly, et al. “Natality Decline and Miscarriages Associated with the 1918 Influenza Pandemic: The Scandinavian and United States Experiences.” The Journal of Infectious Diseases, vol. 204, no. 8, Oct. 2011, pp. 1157–64, https://doi.org/10.1093/infdis/jir510. Accessed 6 Apr. 2023.

[8] Bootsma, Martin C. J., and Neil M. Ferguson. "The Effect of Public Health Measures on the 1918 Influenza Pandemic in U.S. Cities." Proceedings of the National Academy of Sciences, vol. 104, no. 18, May 2007, pp. 7588–93, https://doi.org/10.1073/pnas.0611071104. Accessed 6 Apr. 2023.

[9] Brende, Borge, and Bob Sternfels. Resilience for Sustainable, Inclusive Growth. 7 June 2022, www.mckinsey.com/capabilities/risk-and-resilience/our-insights/resilience-for-sustainable-inclusive-growth. Accessed 27 Mar. 2023.

[10] Clay, Karen, et al. "Pollution, Infectious Disease, and Mortality: Evidence from the 1918 Spanish Influenza Pandemic." The Journal of Economic History, vol. 78, no. 4, Oct. 2018, pp. 1179–209, https://doi.org/10.1017/s002205071800058x.

[11] —. "What Explains Cross-City Variation in Mortality during the 1918 Influenza Pandemic? Evidence from 438 U.S. Cities." Economics and Human Biology, vol. 35, Dec. 2019, pp. 42–50, https://doi.org/10.2139/ssrn.3335921. Accessed 6 Apr. 2023.

[12] Costley, Drew. "Climate Change Will Result in New Viruses and Risk of New Diseases, Says Study." *PBS NewsHour*, 28 Apr. 2022, www.pbs.org/newshour/science/climate-change-will-result-in-new-viruses-and-risk-of-new-diseases-says-study. Accessed 22 Mar. 2023.

[13] Crosby, Alfred W. *America's Forgotten Pandemic : The Influenza of 1918*. Cambridge University Press, 2010.

[14] Fanning, Patricia J. *Influenza and Inequality: One Town's Tragic Response to the Great Epidemic of 1918*. University Of Massachusetts Press, 2010.

[15] Grantz, Kyra H., et al. "Disparities in Influenza Mortality and Transmission Related to Sociodemographic Factors within Chicago in the Pandemic of 1918." Proceedings of the National Academy of Sciences, vol. 113, no. 48, Nov. 2016, pp. 13839–44, https://doi.org/10.1073/pnas.1612838113. Accessed 6 Apr. 2023.

[16] Helgertz, Jonas, and Tommy Bengtsson. “The Long-Lasting Influenza: The Impact of Fetal Stress during the 1918 Influenza Pandemic on Socioeconomic Attainment and Health in Sweden, 1968-2012.” Demography, vol. 56, no. 4, Aug. 2019, pp. 1389–425, https://doi.org/10.1007/s13524-019-00799-x. Accessed 6 Apr. 2023.

[17] Johnson, Kelsey. “How Remote Work Risks a New Digital Jobs Divide for Minorities.” CNBC, 1 July 2020, www.cnbc.com/2020/07/01/remote-work-risks-creating-a-new-digital-jobs-divide-for-minorities.html. Accessed 8 Apr. 2023.

[18] Kent, Susan Kingsley. *The Influenza Pandemic of 1918-1919: A Brief History with Documents*. Bedford/St. Martin’s, 2013.

[19] Lavietes, Matthew. “Working from Home during Pandemic Found Widening U.S. Inequality.” Reuters, 7 Apr. 2021, www.reuters.com/article/us-usa-labor-market-gallup-trfn/working-from-home-during-pandemic-found-widening-u-s-inequality-idUSKBN2BU15R. Accessed 8 Apr. 2023.

[20] Markel, Howard, et al. “Nonpharmaceutical Interventions Implemented by US Cities during the 1918-1919 Influenza Pandemic.” JAMA, vol. 298, no. 6, 2007, pp. 644–54, https://doi.org/10.1001/jama.298.19.2261-a. Accessed 6 Apr. 2023.

[21] McKinsey and Company. One Billion Days Lost: How COVID-19 Is Hurting the US Workforce. 9 Jan. 2023, www.mckinsey.com/industries/healthcare/our-insights/one-billion-days-lost-how-covid-19-is-hurting-the-us-workforce. Accessed 26 Mar. 2023.

[22] Minkin, Rachel. “Which Industries Are Now Teleworking and What Are the Challenges?” World Economic Forum, 28 Jan. 2021, www.weforum.org/agenda/2021/01/teleworking-remote-working-pandemic-covid-19-industry/. Accessed 8 Apr. 2023.

[23] Ogasawara, Kota. “Persistence of Pandemic Influenza on the Development of Children: Evidence from Industrializing Japan.” Social Science and Medicine, vol. 181, May 2017, pp. 43–53, https://doi.org/10.1016/j.socscimed.2017.03.023. Accessed 6 Apr. 2023.

[24] Porras-Gallo, María-Isabel, and Ryan A. Davis. *The Spanish Influenza Pandemic of 1918-1919: Perspectives from the Iberian Peninsula and the Americas*. Rochester University Of Rochester Press Cop, 2014.

[25] Taubenberger, Jeffery K., and David M. Morens. "1918 Influenza: The Mother of All Pandemics." *Emerging Infectious Diseases*, vol. 12, no. 1, Jan. 2006, pp. 15–22

[26] Velde, Francois R. "What Happened to the US Economy during the 1918 Influenza Pandemic? A View through High-Frequency Data." The Journal of Economic History, vol. 82, no. 1, Feb. 2022, pp. 284–326, https://doi.org/10.1017/S0022050722000055. Accessed 1 Oct. 2022.

[27] Vollmer, Sebastian, and Juditha Wojcik. "The Long-Term Consequences of the Global 1918 Influenza Pandemic: A Systematic Analysis of 117 IPUMS International Census Data Sets." SSRN Electronic Journal, Jan. 2017, https://doi.org/10.2139/ssrn.3083584. Accessed 6 Apr. 2023.