Investigating the Relationship between Immigration Status and Mortality during the 1918 Influenza Pandemic in Cleveland, Ohio

Honours Thesis

Jodi Overwijk

ISCI 4A12: Thesis

Dr. Amanda Wissler

April 11, 2025

Abstract

The 1918 Influenza Pandemic stands as one of the most notable pandemics to date. Not only for its global death toll, but for its unique age-mortality curve, showing unusually high mortality among young adults (approximately ages 20 to 40) (Gagnon et al., 2013; Noymer & Garenne, 2000). While current literature explores various characteristics of this Pandemic, it is limited on the impact immigration status may have had during this time. As such, the objective of this study was to assess how death counts differed across immigration statuses during the 1918 Influenza Pandemic when compared to the year prior. Using the genealogy website, *Family Search*, 2% of death certificates, per month, from Cuyahoga County in Cleveland, Ohio, were collected. This information was utilized to categorize individuals into the Control or Flu timelines (i.e., September 1917 to April 1918, and September 1918 to April 1919, respectively), immigration status (i.e., domestic, first-generation, or second-generation), age of death, and cause of death. A saturated generalized linear model assessed status and age combinations between the Control and Flu timelines using an analysis of deviance and chi-squared analyses. Results suggest that no specific immigration status group was associated with disproportionate mortality. However, when specifically assessing individuals ages 20 to 39, analyses suggest that first-generation individuals had a higher mortality relative to domestic-status individuals. This may illuminate additional characteristics from the Pandemic, suggest an additional avenue for research when assessing its unique age-mortality curve, and create an area of consideration for future pandemic research. Future studies may continue studying this relationship on larger spatial, temporal, and data scales.

Introduction

The 1918 Influenza Pandemic was the result of an avian, type-A influenza (H1N1) virus, and stands as one of the most notable pandemics to date (Worobey et al., 2014). Generally viewed in three distinct waves, the second, occurring during the fall of 1918, has been established as the most lethal (Patterson & Pyle, 1991; Shanks & Brundage, 2012). Not only did this pandemic reach global death toll estimates of over 50 million people, with no regard for class, sex, or race (Murray et al., 2006; Noymer & Garenne, 2000; Reid et al., 2004), but it also had a uniquely lethal effect among the most typically immuno-resilient age groups. Typically with other historical influenza-type viruses, when visualizing age-mortality curves, they produce a distinct U-shape, where the highest mortality is prevalent among very young (i.e., less than 5 years of age) and old (i.e., above the age of 74) populations, with young adults (approximately ages 20-40) being most resilient to illness (Gagnon et al., 2013; Noymer & Garenne, 2000; Simonsen et al., 1998). However, the 1918 virus reimagines this curve and shows a notably higher death rate among the young adult population, resulting in a unique W-shaped curve, where the mortality for this demographic is higher than any historical record of influenza viruses (Nickol & Kindrachuk, 2019; Noymer & Garenne, 2000; Viboud et al., 2013).

The ongoing discourse surrounding the likely causes of this abnormal curve has proposed various hypotheses. The most researched suggestions include tuberculosis (Noymer & Garenne, 2000), cytokine storms (King & Londrigan, 2021), or previous exposure to a similar viral strain, such as the Russian flu of 1889, which would cause immune dysregulation when exposed to the pandemic-type virus (Gagnon et al., 2013). While plausible arguments are made for each case, there is no academic consensus to deduce the exact cause(s). As such, the ongoing debate has shifted towards historical records and societal factors for further explanation of the death toll.

A specific factor of consideration for the time period is the ending of the First World War (WWI). Initially, it was suspected that travel associated with military efforts would increase the young adult demographic's exposure to the virus, serving as a contributing factor to the age discrepancy seen in the mortality curve (Viboud et al., 2013). However, when comparing male to female deaths within the age demographic, the two groups succumbed to illness at similar rates. This discredits the proposed theory, as a disproportionate effect for males would be expected due to their direct involvement with deployment (Viboud et al., 2013). Considering this, another notable factor during this time period was the mass immigration to the United States (USA) that occurred, with more than 10% of the USA being foreign-born in 1900 (US Census Bureau, 1901). This was a major wave of immigration and was regarded as the largest wave in the USA's history at the time (Kraut, 2010). In search of more promising opportunities, the rate of immigration increased in the early 1900s, especially from Southern and Eastern Europe to Cleveland, Ohio, due to its proximate location to water and rail routes (Bronner, 2006). For those arriving in Cleveland, many individuals remained close to the Cuyahoga River and Lake Erie, increasing and diversifying the population of Cuyahoga County (Bronner, 2006). At this time, Cuyahoga County had the highest overall population compared to any other county in Ohio, regardless of the counties sharing a similar spatial size (US Census Bureau, 1920).

In response to the mass immigration experienced by the USA, the US Government implemented various immigration laws throughout the late 1800s and early 1900s. These laws were used to limit who was granted entry into the country, such as the law in 1890, which made it so that only robust and healthy individuals were allowed to immigrate (Kraut, 2010). This led to the US Government enforcing inspections of each individual seeking to enter the USA, where one could be denied entry if not deemed physically and mentally competent (Kraut, 2010). Mandatory quarantine periods for newcomers were also implemented to prevent potential disease spread to the American population, especially during the pandemic. The implementation of the 1917 Immigration Act added further regulations and required those over the age of 16 to pass a literacy test, where one needed to display basic reading comprehension in any language (Office of the Historian, n.d.).

Regardless of the mass wave of immigration to the USA, especially within Cuyahoga County, there is limited literature which investigates its relevance in conjunction with mortality during the 1918 Influenza Pandemic. As such, the objective of this study is to investigate the relationship between immigration status and mortality during the 1918 Influenza Pandemic, specifically within Cuyahoga County in Cleveland, Ohio. To guide this research, two main research questions are established:

Research Question 1: Did a specific immigration status group (i.e., domestic, first-generation, and second-generation) experience disproportionate mortality during the 1918 Influenza Pandemic in Cleveland, Ohio?

Research Question 2: When assessing mortality across immigration status groups, did a specific age group, holding a particular immigration status, experience disproportionate mortality during the 1918 Influenza Pandemic in Cleveland, Ohio?

Materials and Methodology

To investigate the role that immigration status may have had on mortality during the 1918 Influenza Pandemic, the following materials and methodologies were used:

Data Collection

Death certificates from Cuyahoga County in Cleveland, Ohio, were obtained from the free, publicly accessible genealogy website, Family Search ("Ohio Death Records, 1917-1919", n.d.).

A sample of 2% of all death certificates from the County, per month, were randomly collected between the years of 1917 to 1919, amounting to approximately one in every fifty death certificates (range was dependent on the total number of certificates in a given month). This ensured that a representative sample was obtained, which limited selection bias and yielded a proportional data set between months, and between the years preceding and during the pandemic. Once collected, pertinent information from the documents were transcribed into a database using Microsoft (MS) Excel. This information included county, township, address, ward, name, sex, race, marital status, date of birth, age, occupation, industry, birthplace, names of parents, birthplace of parents, date of death, and cause of death (COD).

Following data collection, the data was cleaned utilizing various analytical techniques in both MS Excel and OpenRefine, an open-source data cleaning application. Specifically, data was cleaned to isolate information pertaining to date of birth, age, birthplace, birthplace of parents, date of death, and COD. Date of birth was refined to month and year, to easily determine deaths that occurred preceding, and during the pandemic. Ages were categorized into four age bins, including 0-19, 20-39, 40-59, and 60+ years. These age groups were chosen to ensure an adequate number of death certificates were in each group for statistical analysis and that each age range encompassed approximately the same number of years. Birthplace and parent birthplaces were used to categorize individuals by immigration status. “Domestic” was assigned to individuals who themselves and both parents were born within the USA, “First Generation” was assigned to individuals who themselves and both parents were born outside of the USA, and “Second Generation” was assigned to individuals who themselves were born within the USA, but one or both parents were born outside of the USA. COD was categorized into two groups: 'PI' (COD due to pneumonia/influenza-related causes, see Appendix A for details on COD classification) and 'Other' (all remaining CODs). It should be noted that any death certificates where birthplaces or CODs were listed as “unknown” or were illegible were not included in the dataset to ensure accurate interpretations of any data analyses.

Using the date of death information, death certificates were further separated into two timelines: “Flu” which included individuals who died during the pandemic (i.e., September 1918 - April 1919), and “Control” which included individuals who died one year prior to the Pandemic within the same month range (i.e., September 1917 - April 1918). The control timeline was chosen to assess the typical number of monthly deaths preceding the pandemic, rather than spanning the complete year prior to ensure proportionality among the datasets.

Statistical Analysis

Once the dataset was cleaned and properly categorized, the computer programming application RStudio was used for analysis.

Ensuring the data meets the assumptions of the Pandemic

Given that 2% of the available death certificates were collected from Cuyahoga County, data visualization and analyses were used to ensure that particular trends of the pandemic were seen within the dataset. This included that the dataset had an increase in overall deaths and that those aged 20-39 had a higher mortality. This analysis utilized all individuals irrespective of immigration status.

From the collected death certificates, the monthly death toll between the Control and Flu was used to visualize the influenza wave and compare the differences between the two timelines.

Additionally, COD was used to compare the number of PI-related deaths to all other CODs between the timelines, with special considerations being taken for proportionality. A chi-square test was used to determine if this difference in PI-related deaths, when compared to all other CODs, was significant between the Control and Flu, to align with the aforementioned expectations of the pandemic.

The other assumption of the pandemic that was assessed was that the age group of 20-39 were disproportionately affected by the pandemic than what is typically expected for other known historical influenza viruses. This was accomplished via chi-squared analysis, assessing the proportion of deaths across the four established age groups, between the Control and Flu, for both all CODs and PI-related CODs.

Question 1: Assessing the effect of immigration status without age considerations

To determine if a specific immigration status group experienced disproportionate mortality during the pandemic when compared to the prior year, a chi-squared test was used to assess the difference in the total number of deaths between the Control and Flu, across the three statuses, for all CODs. This test was repeated specifically for PI-related deaths.

Question 2: Assessing the impact of immigration status and age distributions

Age group was assessed to determine if individuals of a particular age group and of a particular immigration status, experienced disproportionate mortality when compared with other age-status combinations. This was determined by assessing the change in death counts from the year prior to during the pandemic.

In order to limit misinterpretation from visualizing the raw death counts, data visualization focused primarily on the proportion of deaths occurring during the pandemic relative to the overall number of deaths occurring within a subgroup (i.e., a certain age and status group). As each respective group had varying death counts during the Control, the impact of the pandemic on the death counts is better interpreted via proportions rather than the raw numbers between the Control and Flu.

The data variables were fitted to a saturated generalized linear model in its hierarchical order given its nested characteristics, and were set as a function of death count. Specifically, data pertaining to whether an individual died during the Control or Flu, their immigration status, and their age group were fitted to the model in that hierarchical order. An analysis of deviance was performed, specifically using a chi-squared test, to assess if there was a significant three-way interaction between the variables of timeline (i.e., Control and Flu), status (i.e., domestic, first-generation, and second-generation), and age (i.e., 0-19, 20-39, 40-59, and 60+).To assess if there was a significant interaction between the variables of timeline and age, the interaction between timeline and age was removed from the original model. This also helped to determine if the interaction between these two variables significantly impacted the fit of the model. Post-hoc analysis was conducted via pairwise testing, with a Bonferroni adjustment, to determine if the marginal means of a specific age and status combination differed significantly between the Control and Flu groups from the other combinations. This methodology was repeated specifically for data where the COD was due to pneumonia and/or influenza-related causes, to determine if the results changed when focusing only on those who died from the influenza virus rather than all CODs. However, it should be noted that data visualization in this respect is less encapsulating due to its lower total death counts in each respective group.

Results

Table 1. Total number of death certificates collected with the cause of death being pneumonia/influenza (PI)-related, compared to all other causes of death, between the Control and Flu timelines (Appendix B).

Timeline	Other	PI	Unknown/Illegible*
Control	114	32	4
Flu	125	125	2

**N.B. Unknown and illegible causes of death variables are removed for analysis.*

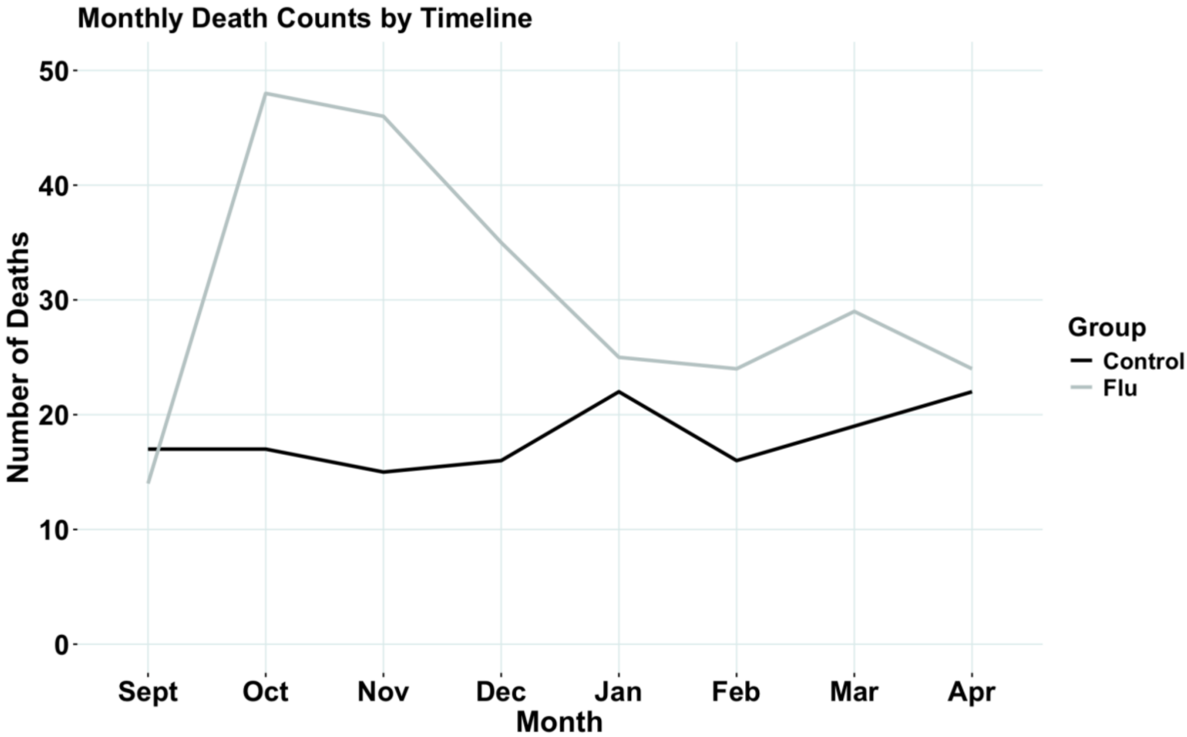


Figure 1. Line graph showing the total number of death certificates collected for all causes of death (n=402) between the Control (black) and Flu (grey) timelines for each month within the respective timeline (Control: 1917-1918; Flu: 1918-1919).

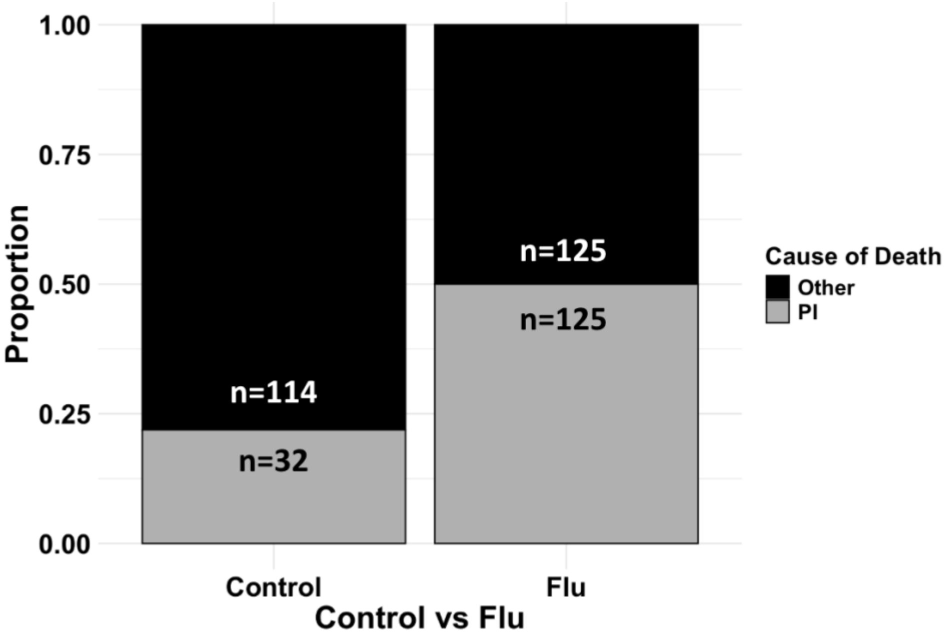


Figure 2. Mosaic plot displaying the proportion of pneumonia/influenza (PI)-related deaths (grey) to all other CODs (black) between the Control (left) and Flu (right) timelines, with the total number of death certificates collected in each respective category indicated by the n-values. PI-related deaths increased from 22% to 50% from the Control to Flu.

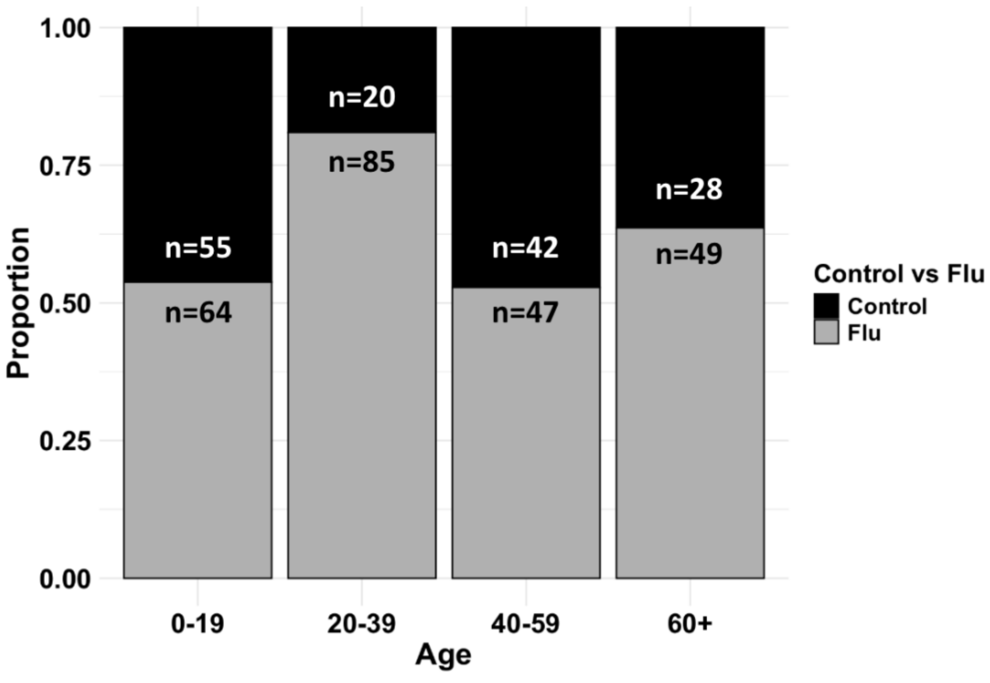


Figure 3. Mosaic plot displaying the proportion of deaths from all causes across age groups (i.e., 0-19, 20-39, 40-59, and 60+) between the Control (black) and Flu (grey) timelines. The total number of death certificates collected in each respective category is indicated by the n-values. The highest proportion of deaths occurring during the Flu is within the 20-39 age group, accounting for 81% of deaths in this age group.

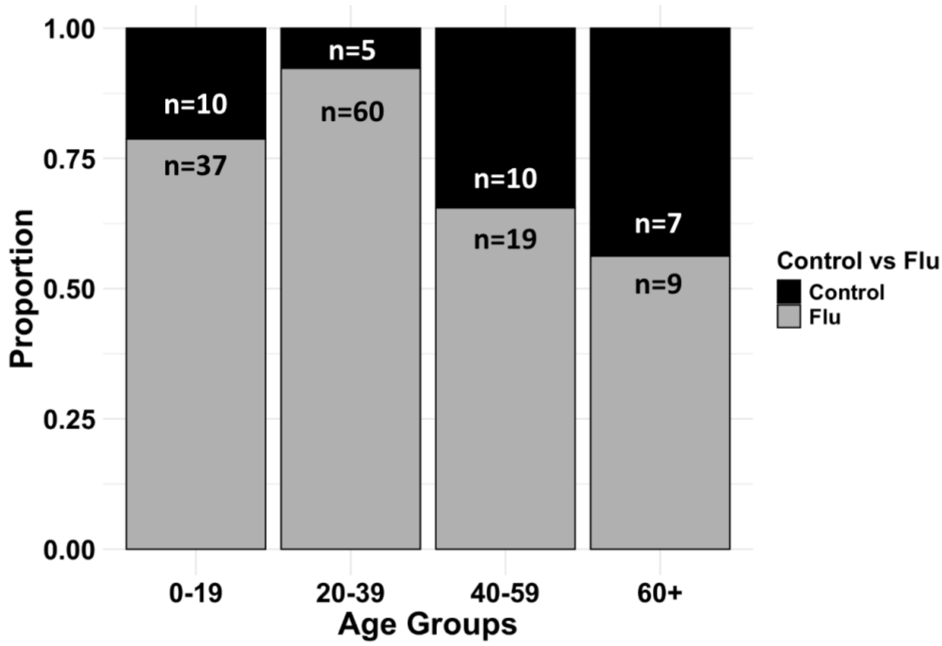


Figure 4. Mosaic plot displaying the proportion of pneumonia/influenza (PI)-related deaths across age groups (i.e., 0-19, 20-39, 40-59, and 60+) between the Control (black) and Flu (grey) timelines. The total number of death certificates collected in each respective category is indicated by the n-values. The highest proportion of PI-related deaths occurring during the Flu is within the 20-39 age group, accounting for 92% of deaths in this age group.

Table 2. Total number of death certificates collected for all causes of death during the Control and Flu, across immigration statuses (Appendix H).

Timeline	Domestic	1st Gen	2nd Gen	Unknown/Illegible*
Control	30	52	63	5
Flu	62	90	93	7

**N.B. Unknown and illegible causes of death variables are removed for analysis.*

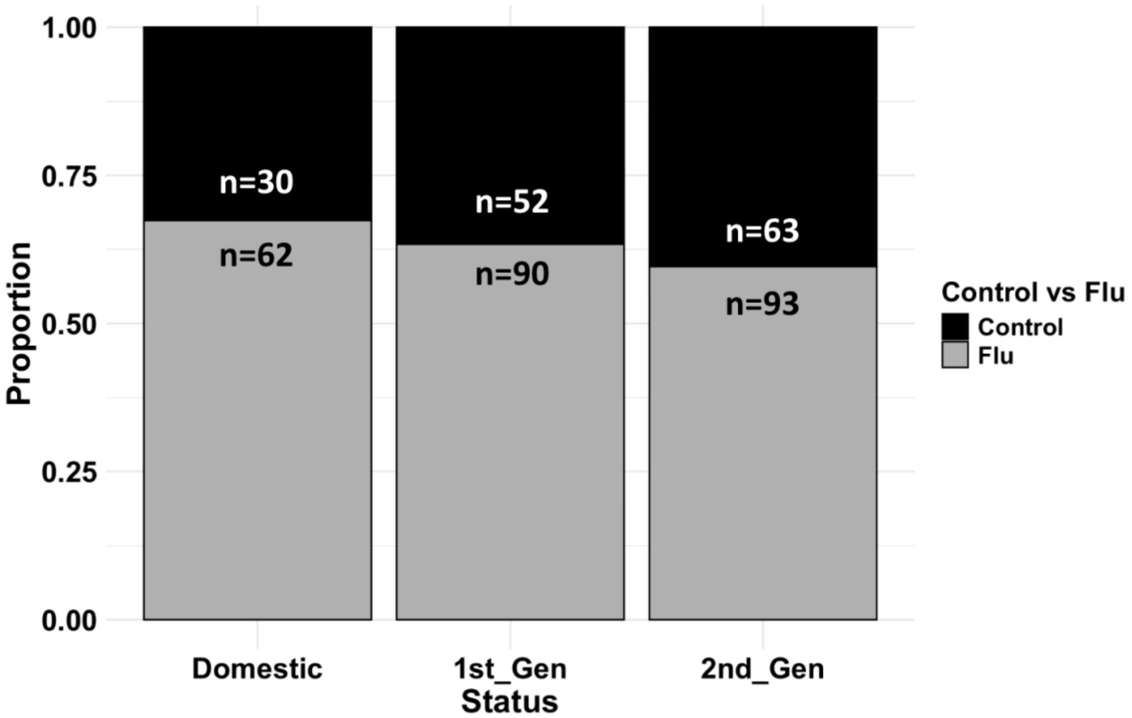


Figure 5. Mosaic plot displaying the proportion of deaths, by immigration status (i.e., domestic-status, first-generation, and second-generation), between the Control (black) and Flu (grey), for all causes of death. The total number of death certificates collected in each respective category is indicated by the n-values.

Table 3. Total number of death certificates collected during the Control and Flu, where the cause of death is pneumonia/influenza (PI)-related, by immigration status (Appendix J).

Timeline	Domestic	1st Gen	2nd Gen	Unknown/Illegible*
Control	6	12	13	5
Flu	32	40	49	7

**N.B. Unknown and illegible causes of death variables are removed for analysis.*

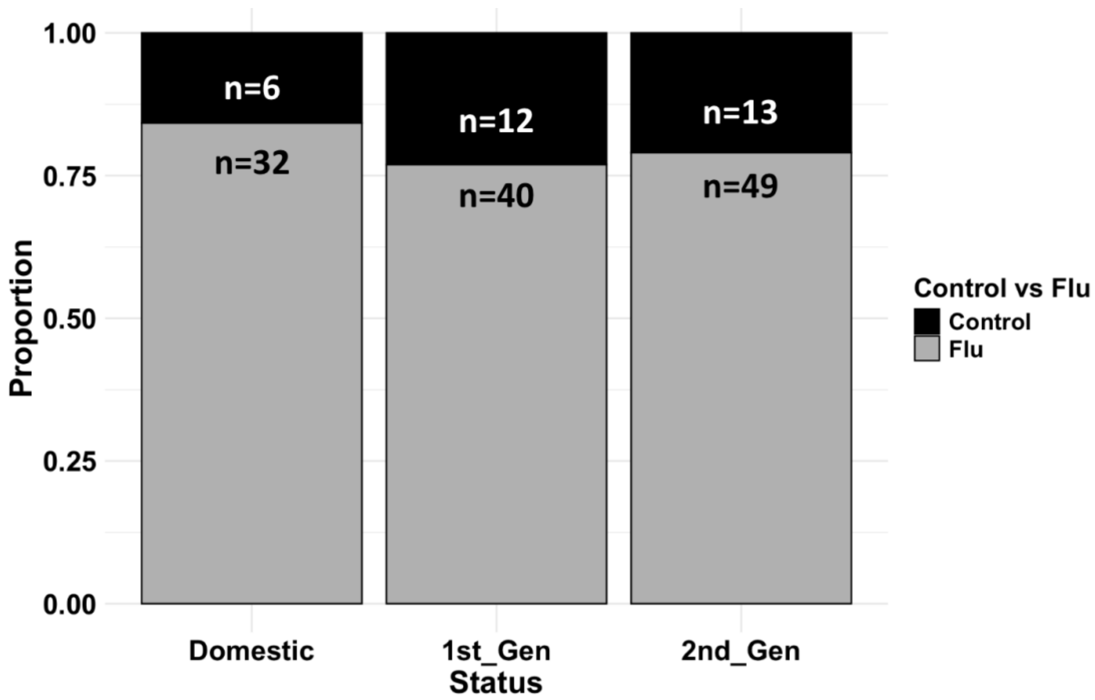


Figure 6. Mosaic plot displaying the proportion of pneumonia/influenza (PI)-related deaths by immigration status (i.e., domestic-status, first-generation, and second-generation) between the Control (black) and Flu (grey) timelines. The total number of death certificates collected in each respective category is indicated by the n-values.

Table 4. Total number of death certificates collected for all causes of death during the Control (A) and Flu (B), by immigration status and age groups.

A: Control

Age (years)	Domestic	1st Gen	2nd Gen
0-19	16	2	37
20-39	3	10	7
40-59	8	22	12
60+	3	18	7

B: Flu

Age (years)	Domestic	1st Gen	2nd Gen
0-19	23	1	40
20-39	17	40	28
40-59	13	16	18
60+	9	33	7

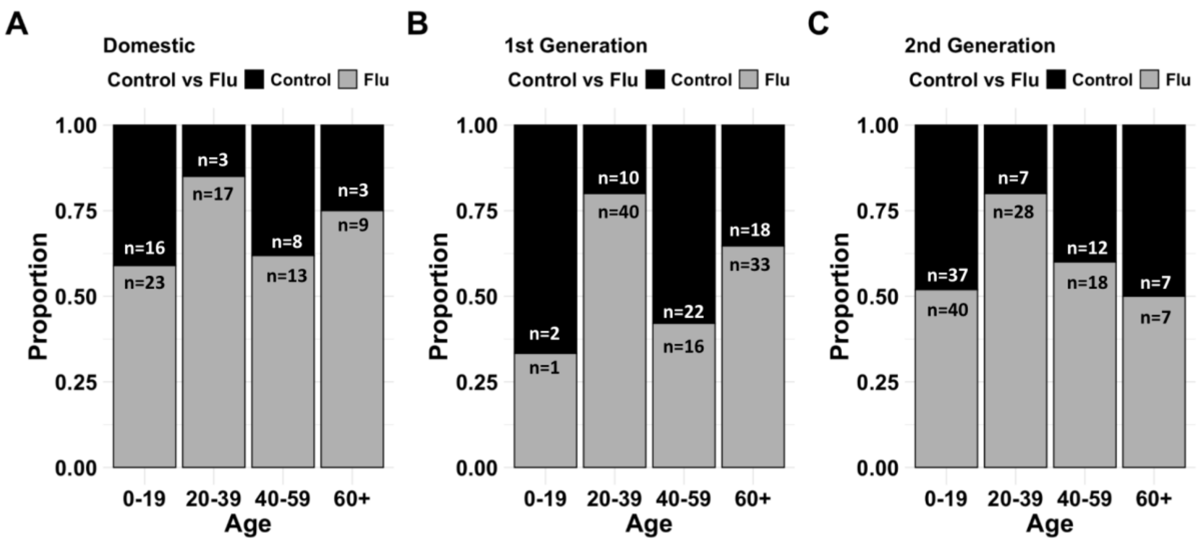


Figure 7. Proportion of deaths between the Control (black) and Flu (grey) timelines, for all causes of death, by age group (i.e., 0-19, 20-39, 40-59, and 60+ years) for each assessed immigration status (n=390). The total number of death certificates collected in each respective category is indicated by the n-values. Across all immigration statuses, the highest proportion of deaths that occurred during the Flu, relative to that of the Control, occurred within the 20-39 age group. **A)** Proportion of deaths across age groups for domestic-status individuals (n=92). **B)** Proportion of deaths across age groups for first-generation individuals (n=142). **C)** Proportion of deaths across age groups for second-generation individuals (n=156).

Table 5. Total number of death certificates collected during the Control (A) and Flu (B), where the cause of death is pneumonia/influenza (PI)-related, collected across immigration statuses and age groups.

A: Control

Age (years)	Domestic	1st Gen	2nd Gen
0-19	3	0	7
20-39	2	2	1
40-59	1	6	2
60+	0	4	3

B: Flu

Age (years)	Domestic	1st Gen	2nd Gen
0-19	14	1	21
20-39	12	32	16
40-59	2	5	10
60+	4	2	2

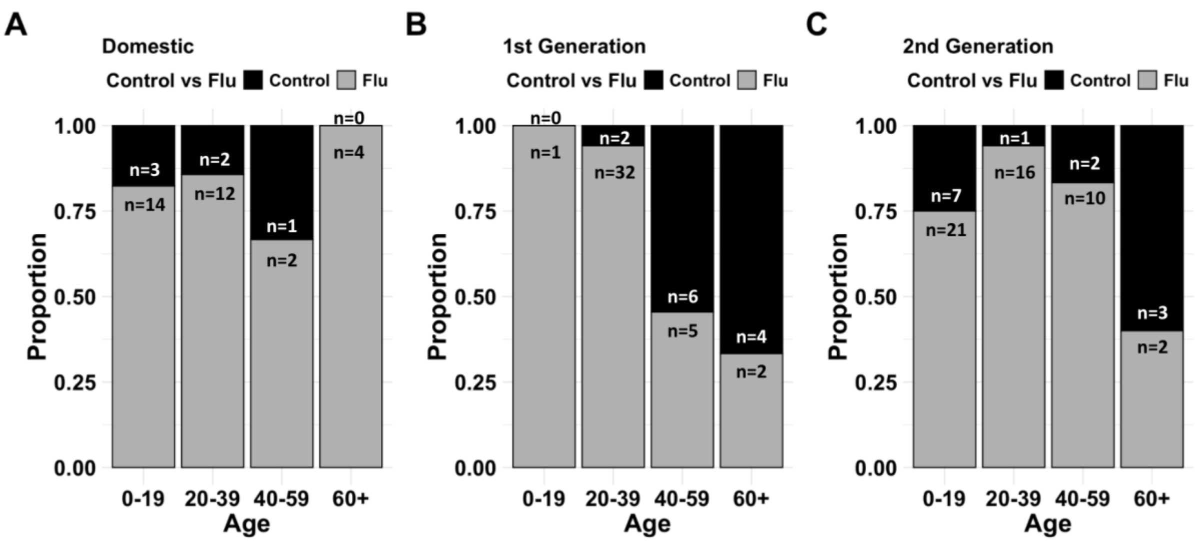


Figure 8. Proportion of deaths from pneumonia/influenza (PI)-related causes between the Control (black) and Flu (grey) by age group (i.e., 0-19, 20-39, 40-59, and 60+ years) for each assessed immigration status (n=152). The total number of death certificates collected in each respective category is indicated by the n-values. **A)** Proportion of deaths across age groups for domestic-status individuals (n=38). **B)** Proportion of deaths across age groups for first-generation individuals (n=52). **C)** Proportion of deaths across age groups for second-generation individuals (n=62).

Count data was collected from 402 death certificates from Cuyahoga County in Cleveland, Ohio, to assess the potential impact that an individual's immigration status had on the proportional difference in death counts between the 1918 Influenza Pandemic and the preceding year (September 1918 - April 1919 and September 1917 - April 1918, respectively). Individuals were categorized into the Control or Flu timelines (n=145 and n=245, respectively) based on their date of death, immigration status, and age. Following the removal of death certificates with incomplete or illegible pertinent information, 390 death certificates were used for analysis.

Data fitting the assumptions of the 1918 Influenza Pandemic

When comparing the total death counts between the Control and Flu for all CODs, per month, all months displayed a higher death count during the Flu, with the exception of September and April (Figure 1). September showed a slightly higher death count in the Control (i.e., Control: n=17 and Flu: n=14), while April had the same number of deaths for both timelines (n=24). The largest discrepancies between the Control and Flu were seen in October, November, and December, displaying an approximate 194%, 220%, and 131% increase from the Control to Flu, respectively (n= 17 to 50, n= 15 to 48, and n= 16 to 37).

Statistical analyses determined that the change in death counts for PI-related deaths, relative to all other CODs, significantly differed between the two timelines (χ^2=29.215, df=1, p<0.001) (Table 1, Appendix C). PI-related deaths accounted for approximately 50% of all deaths during the Flu period, but only 22% of deaths in the Control period (Figure 2).

Assessing the change in death counts across the age groups for all CODs, there is a statistically significant difference between age groups (χ^2=22.784, df=3, p<0.001) (Figure 3, Appendix E). Those aged 20-39 displayed the largest change between the Control and Flu, increasing 325%

(n=20 to 85, respectively), compared to 16%, 12%, and 75% increases for age groups 0-19, 40-59, and 60+, respectively.

These results are consistent when assessing for only PI-related deaths, where there is a statistically significant difference between death counts from the Control to Flu, across all age groups (χ^2=15.41, df=3, p=0.001) (Figure 4, Appendix G). Death count changes from the Control to Flu, in the age range 20-39, increased 1100% (n= 5 to 60, respectively), which is notably higher than the 270%, 90%, and 29% increases for age groups 0-19, 40-59, and 60+, respectively.

Change in death counts across immigration statuses

To assess changes in death counts between the Control and Flu, across immigration statuses, for all CODs, individuals are categorized as domestic-status (Control: n= 30; Flu: n= 62), first-generation (Control: n= 52; Flu: n= 90), or second-generation (Control: n= 63; Flu: n= 93) (Table 2). When comparing the proportion of deaths that occurred during Flu to the total number of deaths in each immigration status group, domestic-status holders had approximately 67% of all deaths occurring during the Flu, first-generation status holders had approximately 63%, and second-generation had approximately 60% (Figure 5). While domestic-status holders had the highest proportion of deaths occurring during the Flu, there is no significant difference between groups (χ^2=1.528, df=2, p=0.466) (Appendix I).

These proportions are then assessed specifically for PI-related CODs for domestic-status (Control: n= 6; Flu: n= 32), first-generation (Control: n= 12; Flu: n= 40), and second-generation individuals (Control: n= 13; Flu: n= 49) (Table 3). Approximately 84% of domestic-status deaths occurred during the Flu, approximately 77% for first-generation, and approximately 79% for second-generation (Figure 6). While domestic-status holders had the highest proportion of deaths occurring during the Flu timeline, there is no significant difference between groups (χ^2=0.739, df=2, p=0.691) (Appendix K).

Change in death counts between timelines, across immigration statuses and age groups

Assessing the change in death counts between the Control and Flu, for each age group (i.e., 0-19, 20-39, 40-59, and 60+), across immigration statuses for all CODs, there is no statistically significant interaction between the three variables (Dev=3.284, df=6, p=0.772) (Figure 7, Appendix L). However, when removing the interaction between the categorization of Control/Flu and age from the model, there is a statistically significant result (Dev=25.533, df=3, p<0.001) (Appendix M). This suggests that there is a significant interaction between the established timelines and age. Thus, removing this interaction would greatly impact the model of the relationship between timeline, status, and age. Ultimately, this indicates that the difference in death counts between the Control and Flu differs significantly for at least one age group. Following pairwise post-hoc analysis contrasting the marginal means of all the status and age group combinations, significant results were found for the following combinations: domestic-status with the age group of 20-39 (z-ratio=25.533, SE=0.626, p=0.006), first-generation with age group of 20-39 (z-ratio=-3.921, SE=0.354, p<0.001), first-generation with the age group of 60+ (z-ratio=-2.069, SE=0.293, p=0.039), and second-generation with the age group of 20-39 (z-ratio=-3.281, SE=0.423, p=0.001) (Appendix N).

Given the significant interactions for all status-age interactions for the 20-39 age group, a pairwise post-hoc analysis contrasting the Control and Flu, with status combinations, specifically for the age group of 20-39, was conducted. Results show a significant p-value when contrasting domestic and first-generation statuses during the Flu for this age group (z-ratio=-2.955, SE=0.290, p=0.009) (Appendix O). This suggests that the difference in death counts between the Control and Flu differs significantly between domestic-status and first-generation status for those aged 20-39, indicating that one of these groups experienced higher mortality than what is expected. There is no significant p-value for this age group when contrasting domestic-status and second-generation status (z-ratio=-1.623, SE=0.307, p=0.314), nor first-generation and second-generation status holders (z-ratio=1.448, SE=0.246, p=0.443).

These interactions were also assessed for PI-related CODs specifically, although no statistically significant interaction between the three variables were found (Dev=-9.903, df=6, p=0.129) (Figure 8, Appendix P). When removing the interaction between the categorization of Control/Flu and age from the model, there is a statistically significant result (Dev=17.307, df=3, p<0.001) (Appendix Q), indicating an interaction between the established timelines and age. Thus, removing this interaction would greatly impact the model of the relationship between timeline, status, and age. Ultimately, this indicates that the difference in death counts between the Control and Flu differs significantly for at least one age group.

Following pairwise post-hoc analysis contrasting the marginal means of the status and age group combinations, significant results are shown for multiple combinations: domestic-status with the age group of 0-19 (z-ratio=-2.421, SE=0.640, p=0.016), domestic-status with the age group of 20-39 (z-ratio=-2.346, SE=0.760, p=0.019), first-generation status with the age group of 20-39 (z-ratio=-3.804, SE=0.730, p<0.001), second-generation with age group 0-19 (z-ratio=-2.517, SE=0.440, p=0.012), second-generation with the age group 20-39 (z-ratio=-2.690, SE=1.030, p=0.007), and second-generation with the age group 40-59 (z-ratio=-2.078, SE=0.770, p=0.038) (Appendix R).

Given the significant interactions for all status-age combinations for the age group of 20-39, a pairwise post-hoc analysis contrasting the Control and Flu with status combinations for the age group of 20-39 was conducted. Results show a significant p-value when contrasting domestic-status and first-generation status counts during the Flu for this age group (z-ratio=-2.898, SE=0.339, p=0.011) (Appendix S). This suggests that the difference in death counts between the Control and Flu differs significantly between domestic-status and first-generation status for those aged 20-39, indicating that one of these groups experienced higher mortality than expected. There is no significant p-value when contrasting domestic-status and second-generation status (z-ratio=-0.753, SE=0.382, p=1.000), nor when contrasting first-generation and second-generation status holders (z-ratio=2.264, SE=0.306, p=0.071) during the Flu timeline for this age group.

Discussion

This study aimed to evaluate the impact of an individual's immigration status on mortality during the 1918 Influenza Pandemic using death certificates collected from Cuyahoga County in Cleveland, Ohio.

Data meeting the assumptions of the Pandemic

Given the relatively small size of the dataset, both numerically and spatially, it was important to ensure that the dataset aligned with certain demographic trends that have previously been explored in the literature to enhance the validity of the findings pertaining to this study.

Specifically, this pandemic was associated with high mortality rates across all demographics, with those aged approximately 20-40 dying at significantly higher rates than what is typically expected for other historical influenza viruses, presenting a unique characteristic of this pandemic (Gagnon et al., 2013; Noymer & Garenne, 2000). Therefore, these two trends were assessed on the dataset used for this study.

Death counts per month and increase in pneumonia/influenza (PI)-related deaths

Current literature suggests that the pandemic specifically hit Cleveland in October 1918, with citizens receiving their first official warning of the pandemic on September 22, 1918, and reaching peak death rates during that Fall (University of Michigan, n.d.). Come March and April 1919, mortality rates from the virus were declining. This aligns with the observed distribution of death counts seen in Figure 1, where September had a slightly higher death count during the Control than the Flu, as the pandemic had not yet hit Cleveland, and the highest death count discrepancies were observed during October, November, and December, followed by more comparable death counts between the two timelines for January to April.

The difference in the number of PI-related deaths between pandemic and non-pandemic times was also assessed in contrast to all other CODs (Figure 2). This highlighted the significant increase in PI-related deaths from the Control to Flu (χ^2=29.215, df=1, p<0.001)(Appendix C), also aligning with the expectation that there would be a higher number of PI-related deaths occurring during the pandemic.

Distribution of death counts across age groups

The dataset was then assessed to ensure it aligned with previous literature demonstrating that those ages approximately 20-40 had significantly higher mortality compared to other age groups (Gagnon et al., 2013; Noymer & Garenne, 2000). When assessing all CODs across the four age groups (i.e., 0 to 19, 20 to 39, 40 to 59, and 60+), the age range of 20-39 displayed both the highest death counts during the Flu (Appendix D) and the largest proportion of the deaths occurring during the Flu (Figure 3). Additionally, there was a significant difference between the increases of death counts from the Control to Flu between the age groups (χ^2=15.41, df=3, p=0.001) (Appendix E), suggesting that the 20-39 age group had significantly higher mortality in this dataset and ultimately aligning with current literature.

This assessment was repeated, including only death counts where the COD was due to PI-related causes (Figure 4). Ultimately, this resulted in a similar outcome from the assessment that included all CODs, where the age group of 20-39 displayed the highest overall death counts during the Flu (Appendix F) and the highest proportion of the death counts occurring during the Flu across all age groups (Figure 4). When assessing the change in death counts between the two timelines, there was a significant difference between age groups (χ^2=22.784, df=3, p<0.001) (Appendix G). This aligns with current literature and further suggests that those aged 20-39 experienced higher mortality.

Given the increase in total death counts and PI-related deaths during the Flu, and the disproportionate mortality for the 20-39 age group, this dataset aligns with the trends of the 1918 Influenza Pandemic as determined by current literature.

<u>Research Question 1:</u> Did a specific immigration status group experience disproportionate mortality during the 1918 Influenza Pandemic in Cleveland, Ohio?

<u>Null Hypothesis:</u> There is no specific immigration status group that experienced disproportionate mortality between the Control and Flu, relative to the other status groups. This suggests that holding a specific immigration status <u>would not</u> make someone more likely to die during the 1918 Influenza Pandemic.

<u>Alternative Hypothesis:</u> There is at least one immigration status group that experienced disproportionate mortality between the Control and Flu, relative to the other status groups. This suggests that holding a specific immigration status <u>would</u> make an individual more likely to die during the 1918 Influenza Pandemic.

To determine whether a specific immigration status group experienced disproportionate mortality during the 1918 Influenza Pandemic, total death counts for each status group were collected for both the Control and Flu.

Total death counts were collected for all CODs (Table 2), where although second-generation status holders held the highest number of deaths during the Flu (n= 93), when comparing the proportion of deaths occurring during the Flu to the total number of deaths collected for each respective status group, the proportions are relatively similar (Figure 5). This suggests that there was no one specific immigration status group that had a disproportionate mortality during the 1918 Flu in Cleveland. This was further confirmed using a chi-squared test, where I failed to reject my null hypothesis, suggesting that a specific immigration status would not make one more likely to die during the 1918 Influenza Pandemic (χ^2=1.528, df=2, p=0.466) (Appendix I).

This analysis was repeated, using only death counts where the COD was due to PI-related causes (Figure 6). Similar results followed this analysis, with second-generation status holders having the largest number of deaths occurring during the Flu (Table 3). However, when assessing the proportion of deaths occurring during the Flu to the total number of deaths for each status, the proportions across the status groups are relatively similar (Figure 6). Again, this suggests that there was not one specific status group that experienced disproportionate mortality during the Pandemic. Chi-squared analysis was performed to assess the change in death counts from the Control to the Flu. Following this, I failed to reject my null hypothesis, further suggesting that a specific immigration status would not make one more likely to die during the 1918 Influenza Pandemic (χ^2=0.739, df=2, p=0.691) (Appendix K).

Irrespective of whether one assesses all CODs or specifically PI-related causes, based on these results, there is no indication that a specific immigration status group experienced disproportionate mortality during the 1918 Influenza Pandemic, when only considering immigration status. These findings generally align with current literature, as it is accepted that this Pandemic affected all individuals, regardless of various demographics, including class and race (Kraut, 2010).

Regardless of this Pandemic occurring amidst a large wave of immigration into the USA and with immigrants being a historic scapegoat for disease spread blame, no specific status group was blamed during the 1918 Influenza Pandemic (Bronner, 2006; Irwin, 2008; Kraut, 2010). While there are cases of prejudice towards racialized groups, particularly towards African-Americans, many places, such as areas in Connecticut, also had a lack of hostility towards immigrant groups during the Pandemic (Irwin, 2008; Kraut, 2010). This was due to the response of the Italian community, one of the largest immigrant groups. Following the end of WWI, Italian business-owners and professionals used the Pandemic to show their integration and patriotism of the USA through their attempts at compliance with public health efforts, reshaping their public standing to be that of cohesion with the USA community (Irwin, 2008). Furthermore, when considering cases of medicalized prejudice, it should be noted that during the Pandemic, hospitals on the front lines were typically those that were supported by various religious and ethnic groups, and that these hospitals held comparable mortality rates of the flu as other hospitals at this time (Kraut, 2010).

While this speaks largely to those who would have already immigrated to the USA prior to the Pandemic, the travel required for immigration should also be noted as a possible contributing factor to the disease spread. Regardless of the immediate assumption that travel to the USA could cultivate disease amidst the long journey, between the years of 1891 and 1930, anyone who wished to immigrate to the USA was required to undergo a medical inspection and quarantine period (Bateman-House & Fairchild, 2008; Fairchild, 2003; Kraut, 2010). Although the quarantine period was not standardized, those who arrived at the border with any impaired mental or physical capabilities, or developed indications of disease or illness during the quarantine period, would be denied entry into the USA (Bateman-House & Fairchild, 2008; Fairchild, 2003; Kraut, 2010). Ultimately, this mitigated the risk of disease spread to the American population via travel conditions that immigrants faced during this time.

Given the current literature supporting the ideas that immigration status may not have played a crucial role in an individual's mortality, irrespective of additional demographic information, and the analysis conducted within the scope of this study, when only considering the immigration status of those who perished between the years of 1917-1919 (within the relevant months), there was no singular immigration status group that exhibited a higher mortality during the 1918 Influenza Pandemic.

Research Question 2: When assessing mortality across immigration status groups, did a specific age group, holding a particular immigration status, experience disproportionate mortality during the 1918 Influenza Pandemic in Cleveland, Ohio?

Null Hypothesis: There is no specific age and immigration status combination that experienced disproportionate mortality between the Control and Flu, relative to the other age-status group combinations. This suggests that holding a specific immigration status at any age would not make someone more likely to die during the 1918 Influenza Pandemic.

Alternative Hypothesis: There is at least one age and immigration status combination that experienced disproportionate mortality between the Control and Flu, relative to the other age-status group combinations. This suggests that holding a specific immigration status within a specific age range would make an individual more likely to die during the 1918 Influenza Pandemic.

Given the known relationship between mortality during the 1918 Influenza Pandemic and age, with those aged approximately 20-40 years old succumbing to illness at much higher rates than is typically expected for other historical influenza viruses (Gagnon et al., 2013; Nickol & Kindrachuk, 2019; Noymer & Garenne, 2000), this study then assessed this characteristic in conjunction with immigration status.

Specifically, this portion of the study assessed whether a specific immigration status and age group combination experienced disproportionate mortality during the 1918 Influenza Pandemic, when compared with the year prior. From each immigration status group, the data were further categorized into the four age groups (i.e., 0 to 19, 20 to 39, 40 to 59, and 60+). Across all immigration statuses for all CODs, the highest proportion of deaths occurring during the Flu timeline occurred within the 20-39 age group (Figure 7). This aligns with the expectation that those within this age range would experience a higher mortality when compared to other age groups. There was a significant difference in the proportion of deaths occurring during the Flu among first-generation individuals aged 20-39 relative to domestic-status individuals within the same age group (z-ratio=-2.955, SE=0.290, p=0.009), but not for any other status comparison (Appendix O).

To further investigate the result, this analysis was repeated, but isolated only for CODs that were a result of PI-related causes (Figure 8). This notably reduced the sample size for each age group (Table 5) and presented figures and analyses that were less encompassing of the data and trends. Regardless of this, similar insights emerged. Statistical analysis demonstrated a significant difference in the proportion of deaths occurring during the Flu between first-generation and domestic-status groups for the age group of 20-39 (z-ratio=-2.898, SE=0.339, p=0.011) (Appendix S). Therefore, the null hypothesis was rejected, as at least one age-status combination significantly differed in the proportion of deaths occurring during the Flu relative to other age-status combinations. More specifically, when assessing the age group of 20-39, first-generation individuals displayed a significantly higher proportion of deaths during the Flu than domestic-status individuals within the same age range. This result is not seen when contrasting first- and second-generation individuals, nor second-generation and domestic-status individuals. Overall, this suggests that first-generation individuals aged 20-39 experienced disproportionate mortality during the 1918 Influenza Pandemic when compared to domestic-status individuals within the same age range.

To investigate why first-generation individuals between 20-39 years of age experienced higher mortality during the pandemic relative to domestic-status individuals within the same age range, one must look past the initial assumptions of how this virus spread. As mentioned previously, this virus did not have a preference for a specific class or race (Bronner, 2006; Irwin, 2008; Kraut, 2010), nor are there mentions of hostility towards many immigrant groups, with certain exceptions for particular localized groups who held biases given the recent ending of WWI (Irwin, 2008; Kraut, 2010). In addition, those immigrating to the USA underwent strict medical inspections and quarantines to ensure only healthy and fit individuals were granted allowance into the country, and to prevent disease spread to the American population (Bateman-House & Fairchild, 2008; Fairchild, 2003; Kraut, 2010). Regardless of the literature supporting the notion that a specific immigration status would not be disproportionately affected in terms of mortality, considering additional underlying factors (e.g., literacy, occupations, and cultural shifts) highlights a greater discrepancy between the status groups that may have made first-generation individuals more vulnerable to the pandemic.

The realities for first-generation immigrants

This pandemic occurred amidst the largest wave of immigration to the USA in its history (Kraut, 2010). Once arriving in Cleveland, Ohio, many immigrants remained close to the Cuyahoga River (Bronner, 2006), increasing Cuyahoga's population by over 48% from 1910 to 1920, and making Cuyahoga the most populated County in Ohio, regardless of it sharing a similar spatial size to other counties in Ohio (US Census Bureau, 1920). Not only did this provide Cleveland and Cuyahoga County with a highly populated and diverse population, but it also made living conditions crowded relative to the towns most immigrants were familiar with within their home countries (Kraut, 2010).

Over 70% of those arriving to the USA during the early 1900's were between the ages of 18-40 years old and typically held occupations that were considered 'less skilled,' regardless of the strict immigration laws that almost exclusively only allowed skilled, healthy, and competent individuals into the country (Hirschman & Mogford, 2009; Office of the Historian, n.d.). Although those who immigrated were not less skilled compared to USA-domestic individuals, the industries with the highest proportions of immigrants employed were those of mining, manufacturing, and hospitals (Hirschman & Mogford, 2009).

In addition to the above considerations, there is also that of cultural differences. The highest proportions of immigrants, both in Cleveland as a whole, and specifically in Cuyahoga, arrived from Germany, Italy, Hungary, and Poland (US Census Bureau, 1920). Not only were immigrants navigating their new and crowded towns, they also had to balance the cultural shift to the USA, especially around that of public health regulations, and beliefs pertaining to disease spread and illness (Kraut, 2010). For example, while many immigrant workers became ill due to their labourious occupations and crowded towns, it is suggested that many Italian immigrants at the time of the Pandemic tended to avoid seeking medical treatment as they believed illness stemmed from the practice of sorcery, rather than biological means (Dundes, 1992; Kraut, 2010). This belief was especially common among Southern Italians who were accustomed to having limited accessible medical care and therefore turned to folk remedies when falling ill (Kraut, 2010).

Given the many challenges faced by immigrant groups, it was crucial that these individuals received adequate information on the pandemic, especially pertaining to disease spread. While many public health efforts were made to inform the public of the magnitude of the Pandemic and how to avoid falling ill, information that was accessible to immigrants, especially first-generation immigrants who were limited in their English skills and were still adjusting to the new country, was minimal (Kraut, 2010). Rather, these efforts were spearheaded by second-generation immigrants who were more educated and established in the USA (Kraut, 2010; Irwin, 2008).

The limited information accessible to first-generation immigrants becomes more apparent when considering the levels of illiteracy among this group relative to domestic individuals. According to the 1920 census in Ohio, specifically in Cuyahoga County, 13.7% of first-generation immigrants, aged 21 years or older, were illiterate, compared to only 0.1% for domestic and second-generation immigrant individuals (Appendix O) (US Census Bureau, 1920). While this census dates after the pandemic (illiteracy rates specific to Cuyahoga were not listed in the 1910 Ohio census), when comparing the illiteracy percentage from 1910 to 1920 for the entirety of Ohio, there remains a notably higher percentage of those illiterate for first-generation individuals (Appendix U) (US Census Bureau, 1910, 1920). This ultimately suggests that first-generation immigrants had considerably higher illiteracy rates relative to domestic and second-generation immigrant individuals throughout the Pandemic.

Impact of inaccessible information during pandemics

The impact of limited accessible information for marginalized groups can further be demonstrated in contemporary public health disasters. Specifically, this was highlighted during the SARS-CoV-2 (COVID-19) Pandemic of 2020, in Puerto Rico. Puerto Rico is a territory of the USA, with Spanish being the dominant language over English, which is primarily spoken among the majority of the USA population. Similar to the 1918 Influenza Pandemic, COVID-19 is a respiratory illness where high mortality was seen across racial, ethnic, and socioeconomic classes (Cox et al., 2023). During the 2020 Pandemic, the US Government provided crucial information and resources to help mitigate the major impacts of the virus was eliciting. However, given the language and cultural discrepancies between Puerto Rico and the majority of the USA, the government failed to provide linguistically and culturally accessible resources to this population (Cox et al., 2023). The lack of information about COVID-19, paired with the inequitable allocation of pertinent resources to Puerto Rico, resulted in the territory being particularly high-risk to COVID-19 (Cox et al., 2023; García et al., 2020; McSorley et al., 2024). In regards to the lack of equitable resources, this caused total COVID-19 deaths to be incorrectly reported in Puerto Rico, causing the impact to be misinterpreted by the Government. It is now estimated that the death toll was more than double what was originally reported during 2020 (Azofeifa et al., 2021). Additionally, Latin-Americans (which approximately 99% of Puerto Ricans identify as) were found to have been disproportionately impacted by COVID-19 (Rodriguez-Diaz et al., 2020). In this study, it was found that while only 14% of the counties in the USA held a larger Latin-American population, among these counties, 54.4% reported that a COVID-19-related death had occurred. Comparing this number to the remaining 86% of counties where the Latin-American is lower, only 49.4% of these counties reported that a COVID-19-related death had occurred (Rodriguez-Diaz et al., 2020).

This provides a contemporary example of the impact on public health when proper accessible information and resource allocation are not provided, especially among racial and ethnic minorities.

The combination of societal, cultural, and living condition differences experienced by first-generation individuals, along with the considerably higher levels of illiteracy within this group compared to the domestic and second-generation population, highlights the need for accessible information pertaining to the 1918 Influenza Pandemic. The lack of resources being allocated towards these efforts by the USA Government may have resulted in the disproportionate mortality experienced by first-generation individuals, and this is further supported by the contemporary example of the impact that the 2020 COVID-19 Pandemic had on the residents of Puerto Rico and Latin-Americans.

Additionally, this could also explain why only first-generation individuals experienced disproportionate mortality relative to domestic-status individuals, and not relative to second-generation, nor between domestic and second-generation. Given that this explanation refers to the limited accessible information, the largest discrepancy would be seen between those domestic to the USA and new immigrants, as second-generation individuals had lower illiteracy rates, likely due to their extended exposure in the USA.

Limitations of the study

Regardless of the considerations made during this study to ensure the validity of the findings, there are limitations that should be discussed, many of which pertain to the collected data. The information for this study was collected via death certificates from the early 1900s, posing issues regarding the accuracy of the information on the death certificates and during the transcription process, as well as introducing a certain level of human bias.

Furthermore, while an adequate number of death certificates were collected for the means of this study and the efforts made to mitigate any potential selection bias, the data is limited both quantifiably (since only 2% of death certificates, each month within the established time periods, were used) and spatially (given that all death certificates are specifically from Cuyahoga County in Cleveland, Ohio). Additionally, while the dataset used during the study was tested to ensure it aligned with the general trends of the pandemic as discussed in current literature (i.e., increase in the number of overall deaths and PI-related deaths, and the age-mortality distributions), due to the spatial limitations, any results produced may not be indicative of global or country-wide patterns. Despite this, I argue that an investigation of this temporospatial scale provides a unique perspective when assessing immigration, due to the high rates of immigration in Cleveland and allows one to assess local societal and cultural differences which could contribute to inequalities for certain demographics.

This study also did not assess differences in mortality among immigrants with regard to their country of origin, leading to potential discrepancies among a particular group, and limiting the findings' applicability to greater regional and global scales.

Conclusion

The aim of this study was to determine the impact that immigration status had on mortality during the 1918 Influenza Pandemic, specifically in Cuyahoga County in Cleveland, Ohio. Through count data collected via death certificates, death counts were determined during both a wave of the Pandemic and from one year prior, to provide a baseline of the expected number of deaths preceding the Pandemic. While no singular immigration status group was found to have experienced disproportionate mortality during the Pandemic, when further considering age distribution, first-generation immigrants between the ages of 20-39 were found to have disproportionate mortality, relative to domestic-status individuals within the same age range. This discrepancy could be due to the lack of linguistically and culturally accessible resources regarding the Pandemic provided by the public health efforts of the US Government. This would have limited first-generation individuals' information about the Pandemic and personal protection advisories, similarly seen in the contemporary example of the US Government's reaction when the 2020 COVID-19 Pandemic emerged in Puerto Rico. These findings highlight a unique attribute of the 1918 Influenza Pandemic and the role that immigration status held in terms of mortality. Future studies may wish to continue assessing this relationship on larger spatial, temporal, and data scales.

References

Azofeifa, A., Valencia, D., Rodriguez, C. J., Cruz, M., Hayes, D., Montañez-Báez, E., Tejada-Vera, B., Villafañe-Delgado, J. E., Cabrera, J. J., & Valencia-Prado, M. (2021). Estimating and Characterizing COVID-19 Deaths, Puerto Rico, March–July 2020. *Public Health Reports*, *136*(3), 354–360. https://doi.org/10.1177/0033354921991521

Bateman-House, A., & Fairchild, A. (2008). Medical Examination of Immigrants at Ellis Island. *AMA Journal of Ethics*, *10*(4), 235–241. https://doi.org/10.1001/virtualmentor.2008.10.4.mhst1-0804

Bronner, S. (2006). Cleveland. In *Encyclopedia of American Folklife*. M.E. Sharpe.

Cox, H., Gebru, Y., Horter, L., Palomeque, F. S., Myers, K., Stowell, D., Easterling, T., de Noguera, N. S., Medina-Forrester, A., Bravo, J., Pérez, S., Chaparro, J., Ekpo, L. L. P., Cranford, H., Santibañez, S., & Valencia, D. (2023). New York State, New York City, New Jersey, Puerto Rico, and the US Virgin Islands' Health Department Experiences Promoting Health Equity During the Initial COVID-19 Omicron Variant Period, 2021-2022. *Health Security*, *21*(S1), S25–S34. https://doi.org/10.1089/hs.2023.0001

Dundes, A. (1992). *The Evil Eye: A Casebook*. The University of Wisconsin Press.

Fairchild, A. L. (2003). *Science at the Borders: Immigrant Medical Inspection and the Shaping of the Modern Industrial Labor Force*. JHU Press.

Gagnon, A., Miller, M. S., Hallman, S. A., Bourbeau, R., Herring, D. A., Earn, D. JD., & Madrenas, J. (2013). Age-Specific Mortality During the 1918 Influenza Pandemic: Unravelling the Mystery of High Young Adult Mortality. *PLoS ONE*, *8*(8), e69586. https://doi.org/10.1371/journal.pone.0069586

García, C., Rivera, F. I., Garcia, M. A., Burgos, G., & Aranda, M. P. (2020). Contextualizing the COVID-19 Era in Puerto Rico: Compounding Disasters and Parallel Pandemics. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *76*(7), e263–e267. https://doi.org/10.1093/geronb/gbaa186

Hirschman, C., & Mogford, E. (2009). Immigration and the American Industrial Revolution From 1880 to 1920. *Social Science Research*, *38*(4), 897–920. https://doi.org/10.1016/j.ssresearch.2009.04.001

Irwin, J. F. (2008). An Epidemic without Enmity: Explaining the Missing Ethnic Tensions in New Haven's 1918 Influenza Epidemic. *Urban History Review / Revue d'histoire Urbaine*, *36*(2), 5–17.

King, P. T., & Londrigan, S. L. (2021). The 1918 influenza and COVID-19 pandemics: The effect of age on outcomes. *Respirology (Carlton, Vic.)*, *26*(9), 840–841. https://doi.org/10.1111/resp.14109

Kraut, A. M. (2010). Immigration, Ethnicity, and the Pandemic. *Public Health Reports*, *125*(Suppl 3), 123–133. https://doi.org/10.1177/00333549101250S315

McSorley, A.-M. M., Rivera-González, A. C., Mercado, D. L., Pagán, J. A., Purtle, J., & Ortega, A. N. (2024). United States Federal Policies Contributing to Health and Health Care Inequities in Puerto Rico. *American Journal of Public Health*, *114*(S6), S478–S484. https://doi.org/10.2105/AJPH.2024.307585

Murray, C. J., Lopez, A. D., Chin, B., Feehan, D., & Hill, K. H. (2006). Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: A quantitative analysis. *The Lancet*, *368*(9554), 2211–2218. https://doi.org/10.1016/S0140-6736(06)69895-4

Nickol, M. E., & Kindrachuk, J. (2019). A year of terror and a century of reflection: Perspectives on the great influenza pandemic of 1918–1919. *BMC Infectious Diseases*, *19*(1), 117. https://doi.org/10.1186/s12879-019-3750-8

Noymer, A., & Garenne, M. (2000). The 1918 Influenza Epidemic's Effects on Sex Differentials in Mortality in the United States. *Population and Development Review*, *26*(3), 565–581. https://doi.org/10.1111/j.1728-4457.2000.00565.x

Office of the Historian, Department of State. (n.d.). *Milestones in the History of U.S. Foreign Relations*. Retrieved March 30, 2025, from https://history.state.gov/milestones/1921-1936/immigration-act

Ohio Death Records, 1917-1919. (n.d.). In *FamilySearch*. Retrieved October 4, 2024, from https://www.familysearch.org/ark:/61903/3:1:33S7-9PJ1-9F1V?i=1249&owc=waypoints&wc=MD96-62S%3A287600101%2C287600102&cc=1307272

Patterson, K. D., & Pyle, G. F. (1991). The geography and mortality of the 1918 influenza pandemic. *Bulletin of the History of Medicine*, *65*(1), 4–21.

Reid, A. H., Taubenberger, J. K., & Fanning, T. G. (2004). Evidence of an absence: The genetic origins of the 1918 pandemic influenza virus. *Nature Reviews. Microbiology*, *2*(11), 909–914. https://doi.org/10.1038/nrmicro1027

Rodriguez-Diaz, C. E., Guilamo-Ramos, V., Mena, L., Hall, E., Honermann, B., Crowley, J. S., Baral, S., Prado, G. J., Marzan-Rodriguez, M., Beyrer, C., Sullivan, P. S., & Millett, G. A. (2020). Risk for COVID-19 infection and death among Latinos in the United States: Examining heterogeneity in transmission dynamics. *Annals of Epidemiology*, *52*, 46-53.e2. https://doi.org/10.1016/j.annepidem.2020.07.007

Shanks, G. D., & Brundage, J. F. (2012). Pathogenic Responses among Young Adults during the 1918 Influenza Pandemic. *Emerging Infectious Diseases*, *18*(2), 201–207. https://doi.org/10.3201/eid1802.102042

Simonsen, L., Clarke, M. J., Schonberger, L. B., Arden, N. H., Cox, N. J., & Fukuda, K. (1998). Pandemic versus epidemic influenza mortality: A pattern of changing age distribution. *The Journal of Infectious Diseases*, *178*(1), 53–60. https://doi.org/10.1086/515616

University of Michigan Center for the History of Medicine. (n.d.). *Cleveland, Ohio and the 1918-1919 Influenza Epidemic | The American Influenza Epidemic of 1918: A Digital Encyclopedia*. Retrieved March 30, 2025, from https://www.influenzaarchive.org/cities/city-cleveland.html#

US Census Bureau. (1901). *1900 Census: Volume I. Population, Part 1*. https://www.census.gov/library/publications/1901/dec/vol-01-population.html

US Census Bureau. (1910). *1910 United States Census- Statistics for Ohio*. https://www2.census.gov/library/publications/decennial/1910/abstract/supplement-oh-p1.pdf

US Census Bureau. (1920). *1920 United States Census- Ohio*. https://www2.census.gov/prod2/decennial/documents/06229686v32-37ch3.pdf

Viboud, C., Eisenstein, J., Reid, A. H., Janczewski, T. A., Morens, D. M., & Taubenberger, J. K. (2013). Age- and Sex-Specific Mortality Associated With the 1918–1919 Influenza Pandemic in Kentucky. *The Journal of Infectious Diseases*, *207*(5), 721–729. https://doi.org/10.1093/infdis/jis745

Worobey, M., Han, G.-Z., & Rambaut, A. (2014). Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proceedings of the National Academy of Sciences*, *111*(22), 8107–8112. https://doi.org/10.1073/pnas.1324197111

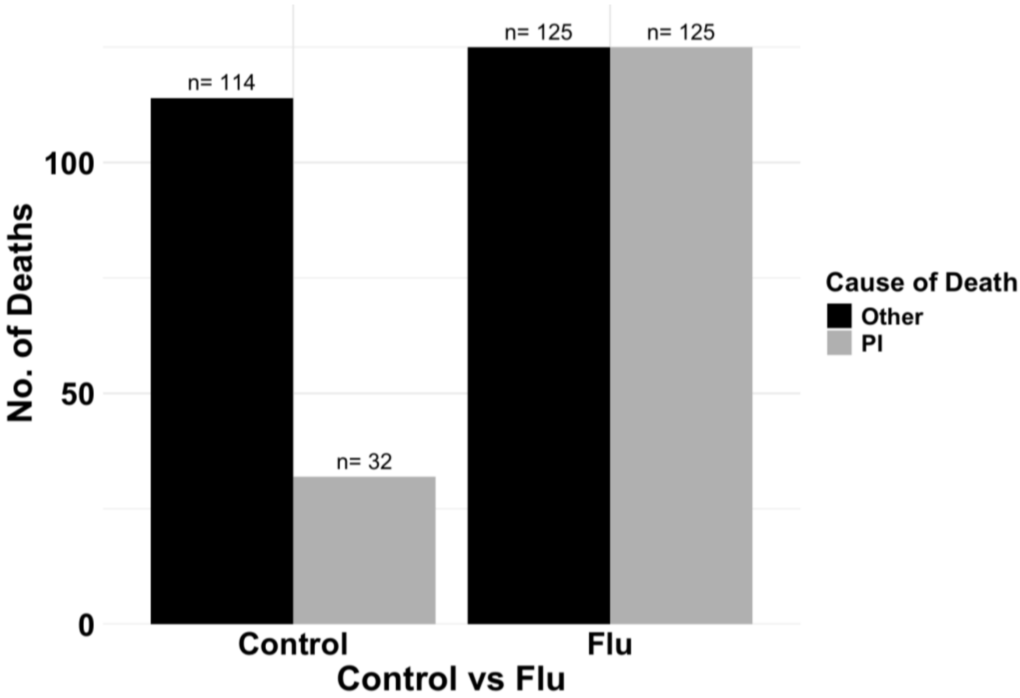
Appendix

Please note that all of the data that was collected for the means of this investigation, and the R Script code used during the statistical analysis, can be found in the following GitHub repository: https://github.com/JodiOv/iSci-Thesis-W2025

A) Categorization of causes of death during data cleaning.

Listed COD on Death Certificate	Corrected COD
PI	- Influenza - Pneumonia - Bronchopneumonia - Lobar Pneumonia - La Grippe
Other	- All other causes of death that were known and legible

B) Bar graph showing the total number of death certificates collected, comparing pneumonia/influenza (PI)-related deaths (grey) to all other causes of death (COD) (black) between the Control (left) and Flu (right) timelines. Across these timelines, PI-related deaths increased from a total of 32 to 125 individuals during the Pandemic, and all other CODs increased from a total of 114 to 125 individuals.



C) R output comparing the number of deaths between the Control and Flu timelines for pneumonia/influenza(PI)-related deaths and all other causes of death.

Data: data1

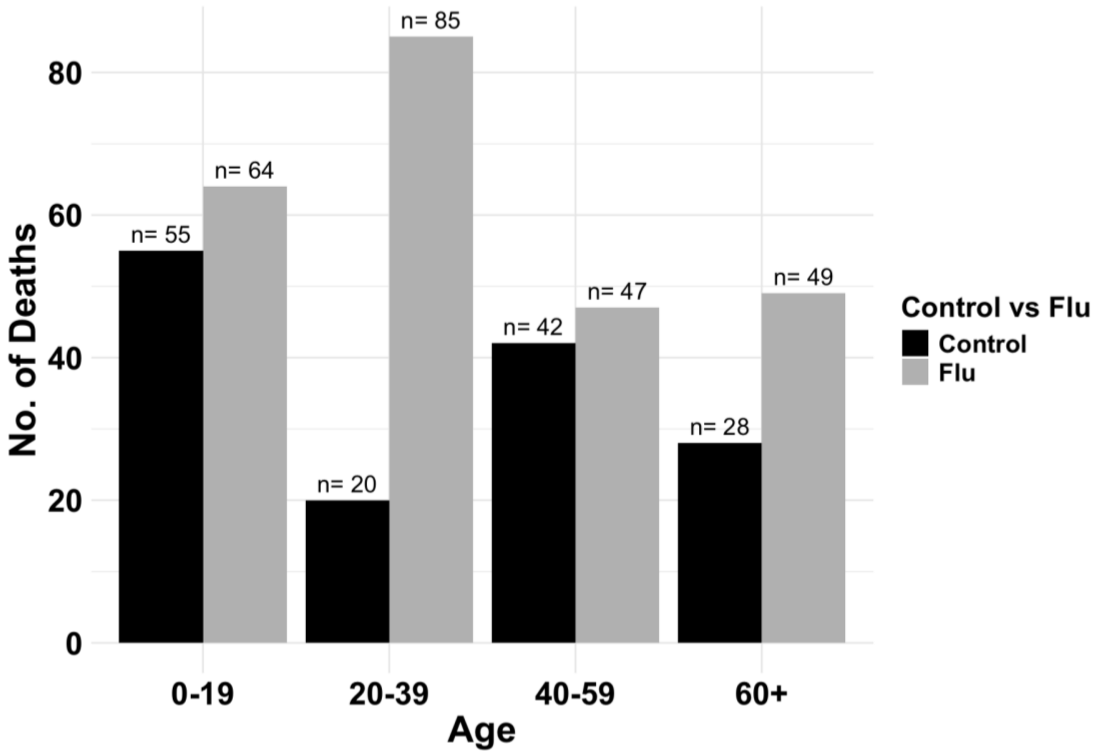
	PI	Other
Control	32	114
Flu	125	125

Call:
chisq.test(data1)

Pearson's Chi-square test with Yates' continuity correction:

χ^2	Degrees of Freedom	p-value
29.215	1	6.477×10^{-8}

D) Bar graph showing the total number of death certificates (n=390) collected between the Control (black) and Flu (grey) timelines across age groups (i.e., 0-19, 20-39, 40-59, and 60+), for all causes of death. The greatest discrepancy is seen within the 20-39 age group, which increased from 20 to 85 individuals, respectively.



E) R output comparing the number of deaths between the Control and Flu timelines, across age groups for all causes of death.

Data: data3

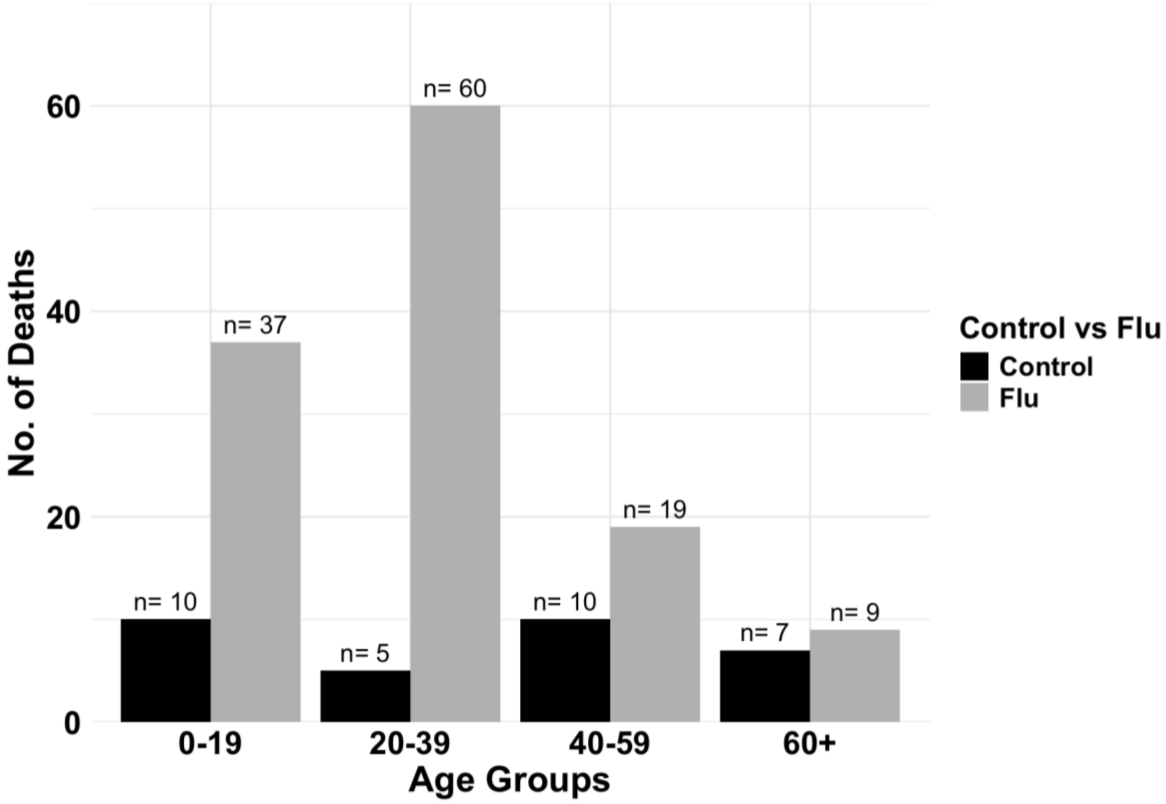
	0-19	**20-39**	**40-59**	**60+**
Control	55	20	42	28
Flu	64	85	47	49

Call:
chisq.test(data3)

Pearson's Chi-square test:

χ^2	**Degrees of Freedom**	**p-value**
22.74	3	4.48×10^{-5}

F) Bar graph showing the total number of death certificates (n=157) collected for pneumonia/influenza (PI)-related deaths, between the Control (black) and Flu (grey) timelines across the established age groups (i.e., 0-19, 20-39, 40-59, and 60+). The greatest discrepancy is seen within the 20-39 age group, increasing from 5 to 60 individuals, respectively.



G) R output comparing the number of deaths between the Control and Flu timelines across age groups for only pneumonia/influenza (PI)-related causes of death.

Data: data2

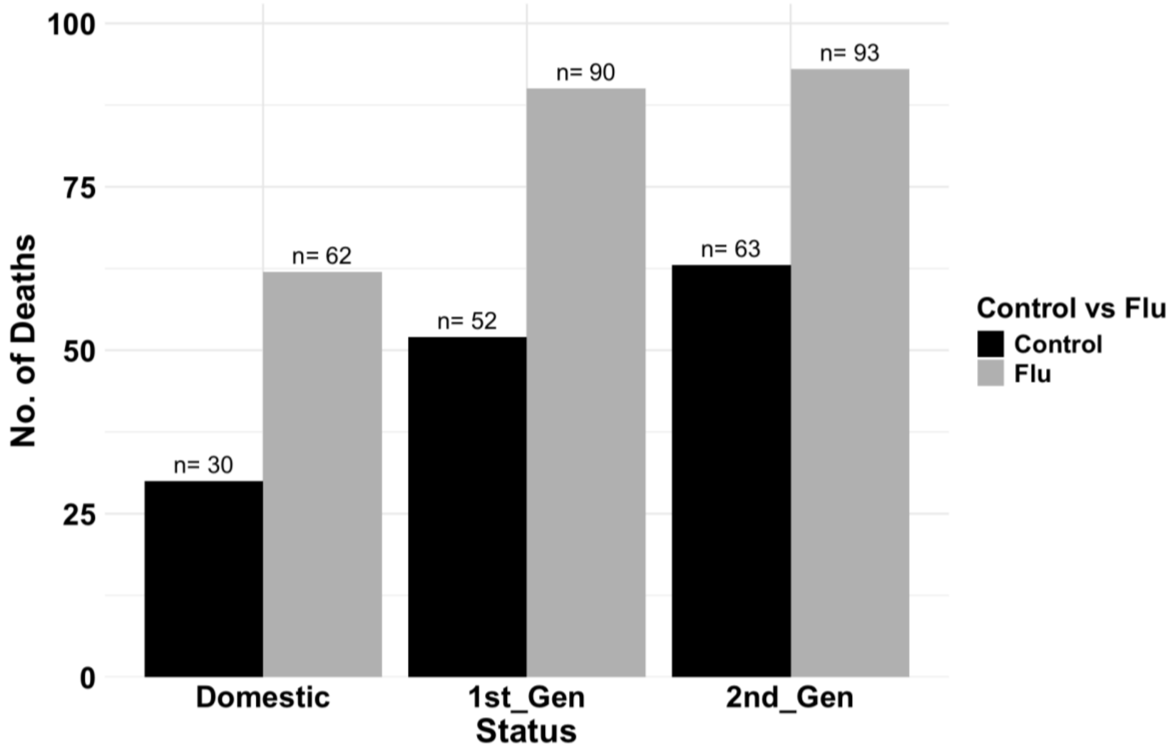
	0-19	20-39	40-59	60+
Control	10	5	10	7
Flu	37	60	19	9

Call:
chisq.test(data2)

Pearson's Chi-square test:

χ^2	Degrees of Freedom	p-value
15.41	3	1.498×10^{-3}

H) Bar graph showing the total number of death certificates (n=390) collected, for all causes of death, between the Control (black) and Flu (grey), by immigration status (i.e., domestic-status, first-generation, and second-generation).



I) R output comparing the number of deaths between Control and Flu timelines, across status groups for all causes of death.

Data: data4

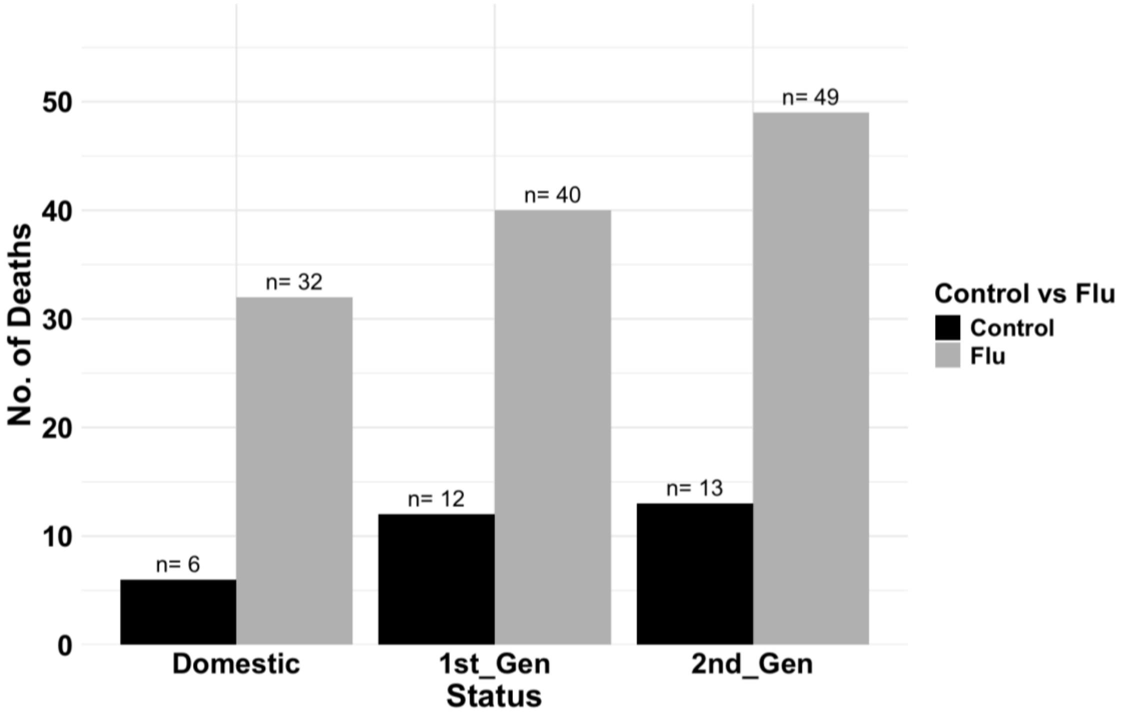
	Domestic	First-Generation	Second-Generation
Control	30	52	63
Flu	62	90	93

Call:
chisq.test(data4)

Pearson's Chi-square test:

χ^2	Degrees of Freedom	p-value
1.5281	2	0.4658

J) Bar graph showing the total number of death certificates (n=152) collected for pneumonia/influenza (PI)-related deaths, between the Control (black) and Flu (grey) timelines by immigration status (i.e., domestic-status, first-generation, and second-generation).



K) R output comparing the number of deaths between the Control and Flu timelines across status groups for only pneumonia/influenza (PI)-related causes of death.

Data: data5

	Domestic	First-Generation	Second-Generation
Control	6	12	13
Flu	32	40	49

Call:
chisq.test(data5)

Pearson's Chi-square test:

χ^2	Degrees of Freedom	p-value
0.73936	2	0.691

L) R output comparing the number of deaths between the Control and Flu timelines, across status and age groups for all causes of death.

Model 1: Count ~ Control_Flu * Status * Age
Model 2: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Control_Flu:Age + Status:Age

Call: anova(Model _1, Model_2, test = "Chi")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	0	0.0000	–	–	–
2	1	3.2843	-6	-3.2843	0.7724

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

M) R output comparing the number of deaths between the Control and Flu timelines across status and age groups, for all causes of death, removing the interaction between Control and Flu timelines and age.

Model 3: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Status:Age
Model 4: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Control_Flu:Age + Status:Age

Call: anova(Model _3, Model_4, test = "Chi")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3	9	28.8169	–	–	–
4	6	3.2843	3	25.533	1.195×10^{-5} ***

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

N) R output of significant p-values (p<0.05) from pairwise post-hoc analysis, comparing the marginal means of the number of deaths between the Control and Flu timelines, across status and age group combinations for all causes of death.

Model 1: Count ~ Control_Flu * Status * Age
emm1<- emmeans(Model_1, ~Control_Flu * Status * Age)

Call:
contrast(emm1, method = “pairwise”, by= c(“Status”, “Age”), adjust= “bonferroni”)

Status = First_Generation, Age= 60+:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-0.606	0.293	Inf	-2.069	0.0386

Status = First_Generation, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.386	0.354	Inf	-3.921	0.0001

Status = Second_Generation, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.386	0.423	Inf	-3.281	0.0010

Status = Domestic, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.735	0.626	Inf	-2.770	0.0056

O) R output of pairwise post-hoc analysis comparing the marginal means of the number of deaths between the Control and Flu timelines, across statuses for the age group 20-39, for all causes of death.

Model 1: Count ~ Control_Flu * Status * Age
emm2<- emmeans(Model_1, ~Control_Flu * Status | Age, at = list(Age = "20-39"))

Call:
contrast(emm2, method = "pairwise", by= c("Control_Flu"), adjust= "bonferroni")

Control_Flu = Control:

Contrast	Estimate	SE	df	z.ratio	p.value
(Domestic 20-39) – First_Generation 20-39)	-1.204	0.658	Inf	-1.829	0.2022
(Domestic 20-39) – Second_Generation 20-39)	-0.847	0.690	Inf	-1.228	0.6585
(First_Generation 20-39) – (Second_Generation 20-39)	0.357	0.493	Inf	0.724	1.0000

Control_Flu = Flu:

Contrast	Estimate	SE	df	z.ratio	p.value
(Domestic 20-39) – First_Generation 20-39)	-0.856	0.290	Inf	-2.955	0.0094
(Domestic 20-39) – Second_Generation 20-39)	-0.499	0.307	Inf	-1.623	0.3138
(First_Generation 20-39) – (Second_Generation 20-39)	0.357	0.246	Inf	1.448	0.4432

Results are given on the log (not the response) scale.
P value adjustment: bonferroni method for 3 tests.

P) R output comparing the number of deaths between the Control and Flu timelines, across status and age groups, for only pneumonia/influenza (PI)-related causes of death.

Model 1: Count ~ Control_Flu * Status * Age

Model 2: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Control_Flu:Age + Status:Age

Call: anova(Model _1, Model_2, test = "Chi")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
1	0	0.0000	–	–	–
2	6	9.9034	-6	-9.9034	0.1288

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Q) R output comparing the number of deaths between the Control and Flu timelines, across status and age groups, for only pneumonia/influenza (PI)-related causes of death, removing the interaction between the Control and Flu timelines and age.

Model 3: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Status:Age

Model 4: Count ~ Control_Flu + Status + Age + Control_Flu:Status + Control_Flu:Age + Status:Age

Call: anova(Model _3, Model_4, test = "Chi")

	Resid. Df	Resid. Dev	Df	Deviance	Pr(>Chi)
3	9	27.2109	–	–	–
4	6	9.9034	3	17.307	6.11×10^{-4} ***

Signif. Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R) R output of significant p-values (p<0.05) from pairwise post-hoc analysis, comparing the number of pneumonia/influenza (PI)-related deaths between the Control and Flu timelines, across status and age groups.

Model 1: Count ~ Control_Flu * Status * Age
emm1<- emmeans(Model_1, ~Control_Flu * Status * Age)

Call:
contrast(emm1, method = "pairwise", by= c("Status", "Age"), adjust= "bonferroni")

Status = First_Generation, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-2.773	0.73	Inf	-3.804	0.0001

Status = Second_Generation, Age= 0-19:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.099	0.44	Inf	-2.517	0.0118

Status = Second_Generation, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-2.773	1.03	Inf	-2.690	0.0071

Status = Second_Generation, Age= 40-59:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.609	0.77	Inf	-2.078	0.0377

Status = Domestic, Age= 0-19:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.540	0.64	Inf	-2.421	0.0155

Status = Domestic, Age= 20-39:

Contrast	Estimate	SE	df	z.ratio	p.value
Control – Flu	-1.792	0.76	Inf	-2.346	0.0190

S) R output of pairwise post-hoc analysis, comparing the number of pneumonia/influenza (PI)-related deaths between the Control and Flu timelines across statuses for the age group 20-39.

Model 1: Count ~ Control_Flu * Status * Age
emm2<- emmeans(Model_1, ~Control_Flu * Status | Age, at = list(Age = "20-39"))

Call:
contrast(emm2, method = "pairwise", by= c("Control_Flu"), adjust= "bonferroni")

Control_Flu = Control:

Contrast	Estimate	SE	df	z.ratio	p.value
(Domestic 20-39) – First_Generation 20-39)	0.000	1.000	Inf	0.000	1.0000
(Domestic 20-39) – Second_Generation 20-39)	0.693	1.220	Inf	0.566	1.0000
(First_Generation 20-39) – (Second_Generation 20-39)	0.693	1.220	Inf	0.566	1.0000

Control_Flu = Flu:

Contrast	Estimate	SE	df	z.ratio	p.value
(Domestic 20-39) – First_Generation 20-39)	-0.981	0.339	Inf	-2.898	0.0113
(Domestic 20-39) – Second_Generation 20-39)	-0.288	0.382	Inf	-0.753	1.0000
(First_Generation 20-39) – (Second_Generation 20-39)	0.693	0.306	Inf	2.264	0.0708

Results are given on the log (not the response) scale.
P value adjustment: bonferroni method for 3 tests.

T) Population (# of individuals) of Cuyahoga in 1920, with the percent that are illiterate (%), for native-white individuals (encompassing domestic-status and second-generation immigration status) and foreign-born individuals (encompassing first-generation immigrant status).

Data collected from (US Census Bureau, 1920)

Age	All Classes		Native-White		Foreign-Born White	
	pop. (#)	% illiterate	pop. (#)	% illiterate	pop. (#)	% illiterate
≥10	746,955	4.8	453,403	0.2	262,369	12.7
≥21	583,071	6.1	314,920	0.1	242,316	13.7

U) Percent of illiterate individuals residing in Ohio who are 21 years of age or older. These are divided by domestic status (native-white), first-generation, and second-generation (foreign-born white) immigrant status individuals.

Data collected from (US Census Bureau, 1910, 1920)

Year of Census	All Statuses (%)	Domestic (%)	First-Generation (%)	Second-Generation (%)
1910	4.0	2.2	11.8	1.1
1920	3.6	1.2	13.4	0.7