The Spanish Influenza in Madrid:*Neighborhood variation in excess mortality across three waves*

Laura Cilek[1], Beatriz Echeverri Davila[2], Gerardo Chowell[3], Diego Ramiro Fariñas[1], Yolanda Casado Ruiz[1]

Keywords: Influenza, Spanish Influenza, Gripe, Excess Mortality, Social Epidemiology

[1]) Centro de Ciencias Humanas y Sociales (CCHS)
Consejo Superior de Investigaciones Cientificas (CSIC)
laura.cilek@cchs.csic.es
[2]) Grupo de Estudios de Poblaciòn y Sociedad, Universidad Complutense de Madrid.
[3]) Georgia State University

1 Background

Variation in the manifestation of Spanish Influenza around the world 100 years ago continues to pique the interests of researchers today. Though studied extensively, questions remain about the ways in which spatial, temporal, and social differences affect influenza mortality, particularly during pandemics.

While climatic differences between tropical and northern countries play a role in seasonal influenza activity [1], the high world-wide mortality of the 1918 Spanish flu pandemic [2] indicates these climactic patterns did not play a major role in regional death outcome. Typically characterized into three waves (Spring, Fall, and Winter), the presence and severity of each wave differed by location, creating debate regarding transmission mechanisms and the role acquired immunity in consecutive breakouts may have played in the tempering of each successive wave. While a virus will spread more slowly in a population with some immunity (i.e. the reproduction number $R_{effective}$ in a partially-immune population will be lower than R_0), the mortality of those exposed to the flu may differ according to their exposure and environmental surroundings [3]. Furthermore, mutation and evolution of the virus may further play a role in the extent to which a person, once immune to an earlier strain, is again succeptable to a ciruclating virus [3, 4].

Prior research focusing on entire cities examines transmission mechanisms and the role acquired immunity in consecutive breakouts may have played in the tempering of each successive wave of Spanish Influenza [5, 6, 7], but without biological evidence, this link is difficult to prove.

Other research clearly points toward a social gradient in the transmission and strength of seasonal influenza outbreaks and preventative vaccination campaigns in the US and across the world, especially in elderly and minority populations[8, 9]. Yet the extent that these patterns remain the same or differ during pandemic events is debated. For example, many of the earlier analyses of the 1918 influenza pandemic led researchers to postulate that the viral strain present in the 1918-920 pandemic events was so virulent that aside from affecting all age groups, the airborne nature of the disease outweighed the potential of any other social variables to create class mortality differentials [10, 11, 12]. Several examples of these studies, such as in a Great Britain Ministry of Health survey about fatality and social status in 1920, can be found in a 2006 review [13, 11].

Other recent research explores the correlation between district level demographic and social levels and flu mortality during the 1918 pandemic, yet the papers focus on single waves; for example, a census tract analysis found that during the strong fall wave in Chicago (the end of September to November 1918), influenza and related mortality was higher in places with greater illiteracy [14]. Mamelund used individual and household level data to find that both neighborhood of residence and apartment size, as a proxy for household wealth, had effects on influenza survival in Kristiana during 1918[13]. Madrid, in the middle of a large period of growth during a stage of urbanization in Spain, also had a large degree of social and economic district variation, which may have played a role in the spread of influenza and spatial mortality differences [15, 13].

These small-area studies within cities often focus on one or two waves of pandemic outbreak, but this can ignore the changing transmission mechanisms and role acquired immunity plays in consecutive breakouts [5, 6, 7]. Few papers combine the strength and timing of each wave with the impact of the socio-economic status, as this requires information about the effected's income and/or occupation across more than one wave in a location. The recorded presence of herald waves and their measured strength is limited to some large cities, and social data can be difficult to discern in an effected place. A study of influenza prevalence in Bergen, Norway suggests a change in the effect of social status (measured as apartment size), where those with smaller apartments experienced higher morbidity in a herald wave and less in the fall wave[16]. The role compositional

factors played in mortality during each of the three influenza waves in Madrid may have changed according to changing immunity in the population and virus makeup.

1.1 Influenza in Madrid

The epidemic influenza outbreaks hit Madrid uniquely compared to generally observed patterns in much of the world. The first wave arrived in Madrid in May 1918, producing an overall mortality peak higher than observed in other locations experiencing a herald wave [17, 18, 4]. The strength of this wave is postulated to have minimized the mortality of the succeeding fall wave, which, though protracted with three distinct peaks in October, December, and February, was relatively mild compared to the outbreaks during this time in Europe and North America [18, 17]. The less studied fourth wave of the pandemic (sometimes referred to an an "echo" wave) produced the largest mortality peak of the period in Madrid from December 1919 to January 1920.

During these outbreaks, age-specific excess mortality also follows a different pattern; with the exception of the fall wave (which experienced a small mortality bump), young adult absolute excess remains low. Nonetheless, due to low baseline rates among these adults, the highest *relative* excess rates are observed in those between 5 and 70.

Given the role of environment, exposure, and social factors associated with influenza mortality, as well as the wave and age specific patterns of influenza in Madrid, this paper seeks to understand and disentangle the the relationship between the two at the neighborhood level. At the time of the outbreaks, Madrid was in a period of great growth and consisted of 100 neighborhoods in various stages of development and with ranging types of population. Here, we look at mortality in these neighborhoods to determine how the type of buildings, apartments, and population present changed the mortality levels, and if these factors varied. by wave.

2 Data

2.1 Death Records

To determine the excess mortality by neighborhood, we use death records from the Civil Register records on deaths in Madrid from 1917-1922. During this time, the register holds approximately 103,500 records, an average of about 17,500 deaths per year; in addition to death date, each record contains demographic and socio-economic information and place of death. This unique, individual-level data allows for analyses to demonstrate the flow and characteristics of each wave across districts, as well as the changes in death by age group and specific cause within and between each location.

During the influenza epidemic, it is probable that many people died at another location than where they were living. As noted, many death certificates are coded with the address of death rather than the deceased's address. Thus, is it likely that due to the epidemic, many neighborhoods with large hospitals or first aid centers experienced inflated excess mortality, due to higher numbers of deaths from people living outside the neighborhood. In order to eliminate this confounding effect in our results, we conducted our analysis with two sets of data; one included all deaths, as coded on the death certificate and the other included only deaths that we believe to have occurred outside of a medical institution. We first collected information on all institutions in Madrid, not limited to hospitals, first aid clinics, other medical centers. From this list of 324 locations, we collected more information and decided on 140 institutions that may have experienced a disproportionate increase

in mortality during the epidemic waves, leading to falsely inflated excess mortality in the neighborhood, as we are concerned primarily with the relationship between neighborhood mortality and the structure and population that lived there. We exclude all deaths that happened at one of these 140 institutions.

Initially, we begin with 103,323 typed death records between 1917-1922, but when ascribing a neighborhood to each record, we face several problems, both due to missing values on the records and current data limitations. Some death records do not have an address written from which to determine the barrio, and a few have no death date written. We use the 1915 road map of Madrid to code addresses into barrios, which creates additional problems due to the large growth of the city between 1915 and our period of observation. A number of death records have a written address not listed in our 1915 database. This also hinders our ability to include several deaths on the outskirts in all years of our analysis.

In total, due to the limitations of the data (roughly equally distributed), about 2500 deaths are not coded to a neighborhood each year. Among the data, we find 91 of the 100 neighborhoods in Madrid at at the time have reasonably accurate data when coded to the geographic level. Thus, we use only these neighborhoods in the analysis.for which we feel the deaths are accurately accounted for.

As mentioned, in order to verify the role hospital deaths may have inflated excess mortality in some barrios, we limit our analysis to only death for which the address is not at one of 140 identified medical centers. From these records, we calculate neighborhood absolute and relative excess mortality rates for all epidemic periods combined, as well as individual waves from May to August 1918, September 1918 to April 1919, and November 1919 to February 1920. We determine the level of excess as the difference between observed deaths during the periods of each wave in 1917 and the number of deaths during each waves. These calculations are based on 86,704 deaths. As visible in the figure 1, some neighborhoods with hospitals continue to experience extreme overall mortality even in the absence of these deaths, such as the neighborhood of Doctor Forquet, slightly south of the city center. This particular neighborhood is also home to an asylum and several orphanages, which had underlying rates of high mortality.

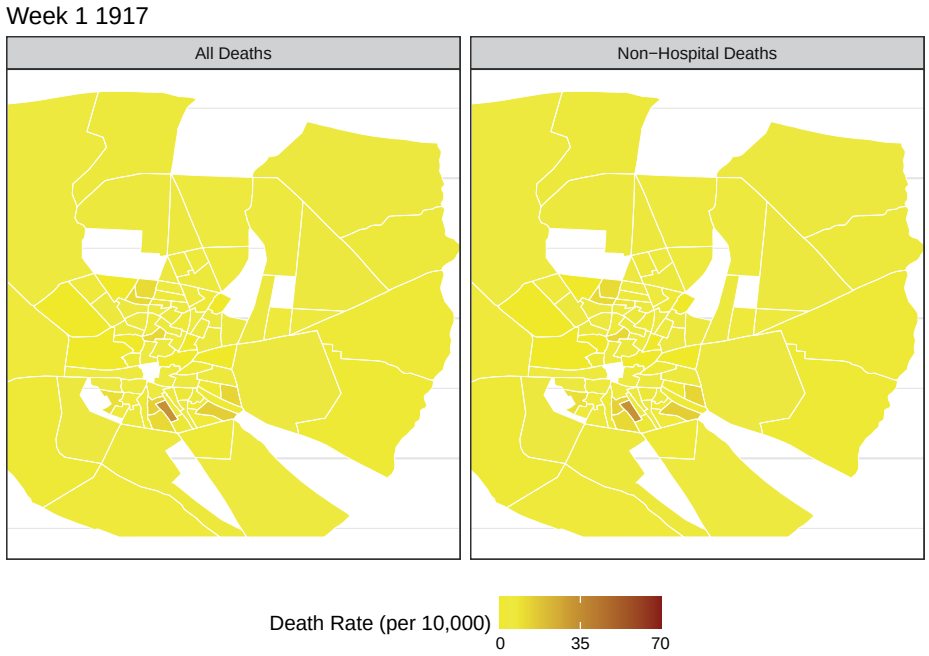
2.2 Padrón and other data

We also take advantage of the 1915 padrón of Madrid, which provides information about the constructed composition of the city, such as the number and type of buildings, rental prices, and measures of density. The padrón also provides limited data about the social composition of the districts, including the proportion of literate population and tabulations of occupations. While the level of detail in the padrón provides a fascinating glimpse into the city composition during this time, certain elements must be combined and altered in order to create variables that bring meaning to the analysis. Thus, several changes were implemented.

The padrón provides numbers of units for several different types of housing and apartments in the city; we aggregated housing generally occupied by those with the lowest income and social position as a percentage of total housing units in the neighborhood. In conjunction with this supplied housing data, we calculated a weighted average of the neighborhood housing prices from a 14-category table with counts of apartment rental prices in each neighborhood. An additional "commercial" variable contains, as a percentage of total available units in the neighborhood, the total amount of private businesses in the neighborhood.

We also manipulated the raw padrón data to create variables describing the social interaction and stratification in each neighborhood with a HISCAM score [19]. We linked a listing of occupations in the neighborhoods to HISCO codes which were then linked to HISCAM codes using both the male-only universal scale (U2), using

Figure 1: Weekly mortality rates with (right) and without (left) hospital and medical center deaths



only the listed male occupations, and the later period scale of both men and women (L), using all occupations in the neighborhood. After creating an average HISCAM score for each neighborhood through a weighted average of the count of each code in the neighborhood, we averaged the aggregate HISCAM scores. We used two HISCAM coding schemes for several reasons, not limited to: a) the HISCAM numbers were used from data not found in southern European countries, b) while most married women did not work, many single females were employed, c) while the results showed that the two schemes were highly correlated, there were some differences in neighborhoods between the two schemes' scores.

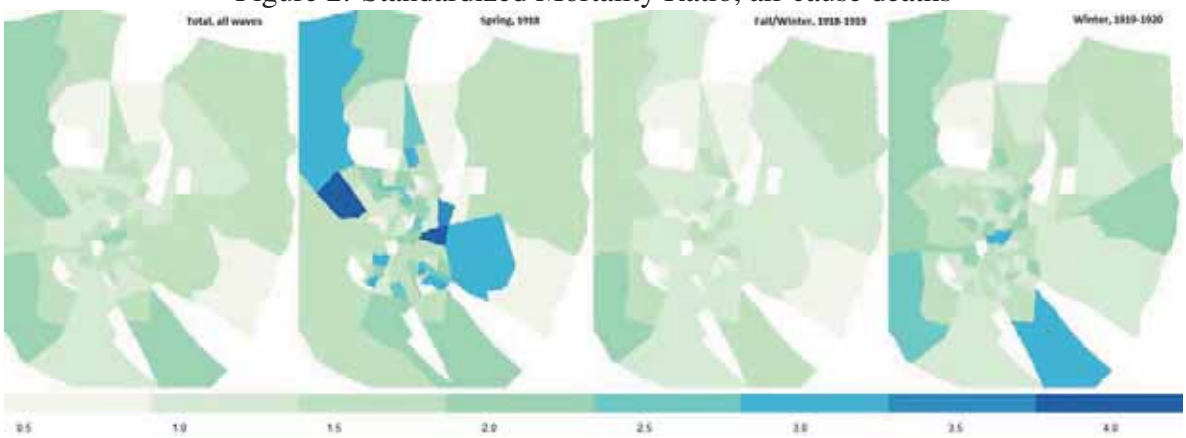
Because age-specific mortality follows a specific pattern in the city of Madrid [17], we also calculate the proportion of the population in each neighborhood that falls within the age range of 5 to 75. During each of the waves, this group had the lowest rates of absolute and highest rates of relative excess mortality.

Figure 2 shows the total and wave-specific standardized mortality ratio for each of the 91 included neighborhoods in our analysis, while summary statistics of other included variables are found in table 2.2. Figure 1 shows weekly all cause mortality rates by district from 1917 until the end of the flu. Further descriptive visuals highlight the relationship between the explanatory variables and excess mortality (absolute and relative) in the Appendix.

2.2.1 Checking for Spatial Autocorrelation

The administrative neighborhoods of the city are clearly defined, but the extent to which each neighborhood can be considered an independent observation must be determined. That is to say, the proximity of neighborhoods to each other may play a role in the variables observed in each barrio, leading to spatial autocorrelation in the data and model issues [20, 21]. We also note that many of the explanatory variables used in the the neighborhood

Figure 2: Standardized Mortality Ratio, all-cause deaths



	Mean	Std. Dev.	Min	Max
Total_Excess	0.30	0.26	-0.20	1.12
Spring_Excess	0.73	0.67	-0.40	3.00
Fall_Excess	0.14	0.24	-0.34	0.93
Winter_Excess	0.52	0.43	-0.16	2.37
Literacy	0.82	0.07	0.65	0.95
HISCAM	51.68	3.67	44.68	58.32
Poor_Housing	0.18	0.14	0.06	0.70
People_Building	49.84	17.26	11.36	103.64
Average_Rent	52.98	37.97	9.68	206.13
Commercial	0.10	0.04	0.02	0.21
Population_5_75	0.91	0.02	0.86	0.95

analysis *are* highly spatially correlated. Thus, after computing the standard mortality ratio for total and each wave in all neighborhoods, we computed Moran's I for the absolute and relative mortality in each neighborhood, finding no evidence of spatial autocorrelation. Further analysis of the residuals and other measures found also found a lack of spatial dependence.

3 Methods

While the strength of an outbreak my be quantified in terms of relative and absolute excess mortality rates, modeling the impact of factors on disease-related mortality requires a slightly modified technique. We examine the probability of experiencing the observed number of cases in a neighborhood based on the expected (baseline) amount of cases. More than absolute and relative excess mortality values, this takes into account the underlying mortality distribution of disease and population size [22, 23]. For example, variations in population size across neighborhoods mean that a single death in a lower populated neighborhood raises the mortality rate more than a single extra death in a larger neighborhood. In our case, the number of observed cases (dependent variable) is offset by the log of the total number of expected deaths in a count-data regression model [24]. The coefficients

can be roughly interpreted as the factor of change to the incidence rate of the disease.

An assumption of Poisson models is that the mean and variance within the distribution are equal to each other. Following tests, this condition of the Poisson distribution is clearly violated in our data, as it is very overdispersed. Ignoring this violation will result in small standard errors and overestimated significance of the model. A negative binomial regression model can account for this violation, producing more precise results due to the inclusion of a theta term θ that accounts for the unobserved heterogeneity in the data and over dispersion [25]. More specifically,the negative binomial and poisson regressions differ in that the mean and variance estimation in a Poisson regression are:

$$E[Y_i|\mathbf{x}_i] = \lambda_i \text{ and } Var[Y_i|\mathbf{x}_i] = \lambda_i$$

However, the negative binomial model allows for over dispersion in the data by:

$$E[Y_i|\mathbf{x}_i, \varepsilon_i] = e^{(\alpha+\beta\mathbf{x}_i'+\varepsilon_i)} = h_i\lambda_i$$

The h_i parameter has an assumed gamma distribution with a mean equal to 1 and a variance equal to $\frac{1}{\theta}$ [26]. Estimation of the parameters is completed using the MASS package of R, which fits the additional h_i parameter through maximum likelihood estimation and provides the θ value of h_i's distribution [27]. As they are nested models, the comparison of the Poisson and negative binomial models can and is done through a likelihood ratio test. Poisson and negative binomial regressions were compared for all models.

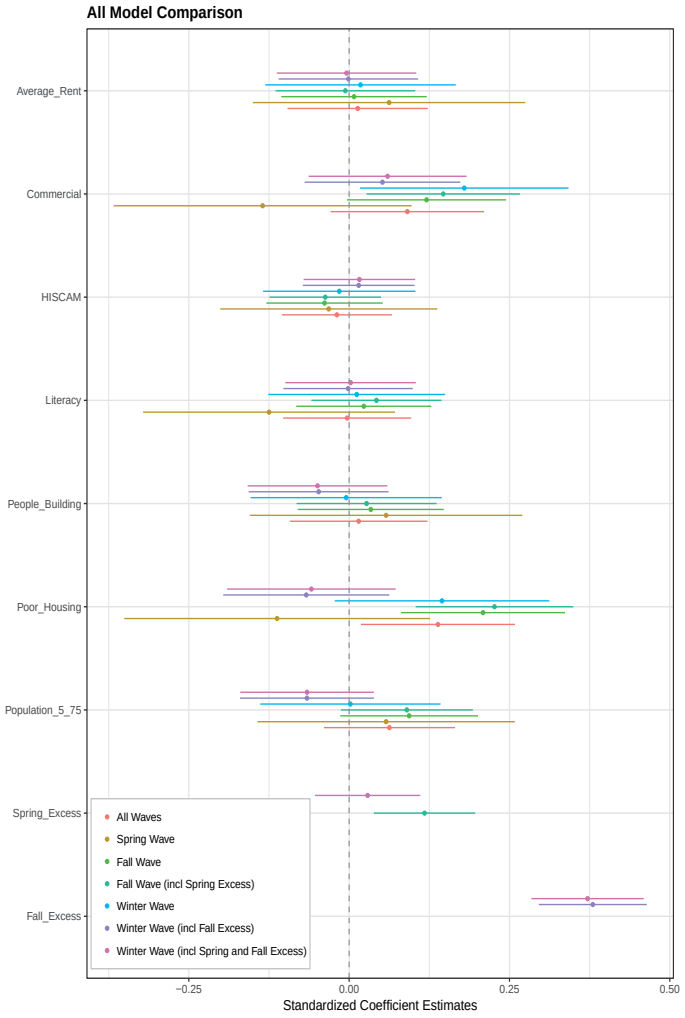
4 Results

Table 1 presents results of negative-binomial regressions for total and wave-specific mortality with all variables. The coefficients are presented as exponents for ease in interpretation; see figure 4 for a plot of the standardized coefficients of the presented models. In all models, the data fit the negative-binomial distribution with parameter θ significantly better than the poisson distribution. The models presented in table 1 each contain all of the descriptive variables outlined above; addition and deletion of variables in other models does not change the significance of any of the others.

Considering total mortality across all epidemic periods, a single variable can be said to have a strong relationship with excess mortality: the amount of living quarters that are generally poorly equipped. As the percentage of small basement and rooftop apartments in a neighborhood increases, the rate of excess mortality across all waves also increases. This relationship holds *and* strengthens in the fall wave, where the magnitude of this variable doubles from overall mortality. Not including mortality from previous waves, this housing variable has a weak relationship in the winter wave.

In each model for the fall and winter wave, the amount of excess mortality in the immediately preceding wave is highly correlated with the level of excess experienced in the same neighborhood during the studied wave. In the fall, the coefficient (exponentiated, almost 1) is nearly proportional to the amount of spring mortality, meaning the level of excess is similar from one way to the next. However, in the winter wave, the expected excess mortality is more than doubled from the fall, likely demonstrative of the magnitude of the winter wave. This indicates that despite the other variables in the model, neighborhoods with high mortality in one wave generally also had high levels of mortality (relative to other neighborhoods) in the other waves.

Figure 3: Plot of standardized regression coefficients for presented models



While the model focusing only on the herald wave mortality finds no relationship between any of the variables and excess mortality, some significant relationships do exist in the subsequent waves, most notably in the fall. In addition to the quality of housing, the proportional amount of those older than 5 and less than 75, as well as the amount of commercial property in the neighborhood have a slight positive effect on the overall wave mortality. This relationship strengthens with the addition of the spring excess mortality. Conversely, the significance of the proportion poor housing and literacy during the initial winter wave model is tempered by the addition of both spring and fall excess. In the winter model which only includes excess mortality for the immediately preceding fall wave, the positive relationship between the percentage businesses and excess increases in strength (but with smaller magnitude) from the initial model.

However, there are some additional inferences that can be made from the results of model fit to encourage continued work on the subject. For example, while no model appears to aptly describe the relationship between our descriptive variables and the excess mortality, we can hypothesize that characteristics of the neighborhood population played different roles in the individual waves. Both the first and last model, analyzing the SMR of the total and last winter 1919-1920 wave, fits best when looking contextually at the built environment and

Table 1: Models with all variables (exponentiated coefficients, original standard errors)

	Dependent variable:						
	Total Observed Deaths	Spring Observed Deaths	Fall Observed Deaths		Winter Observed Deaths		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Poor_Housing	1.658*	0.664	2.143**	2.288***	1.697′	0.783	0.806
	(0.224)	(0.444)	(0.238)	(0.229)	(0.312)	(0.241)	(0.245)
Literacy	0.977	0.396	1.184	1.372	1.091**	0.988	1.017
	(0.378)	(0.744)	(0.399)	(0.384)	(0.522)	(0.382)	(0.385)
HISCAM	0.997	0.996	0.995	0.995	0.998	1.002	1.002
	(0.006)	(0.012)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)
People_Building	1.000	1.002	1.001	1.001	1.000	0.999	0.999
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average_Rent	1.000	1.001	1.000	1.000	1.000	1.000	1.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Commercial	3.473	0.157	5.243′	7.489*	11.746*	2.041**	2.277
	(0.839)	(1.627)	(0.869)	(0.838)	(1.138)	(0.850)	(0.862)
Population_5_75	5.684	4.931	13.324′	12.073′	1.055	0.162	0.162
	(1.443)	(2.835)	(1.517)	(1.453)	(1.986)	(1.474)	(1.471)
Spring_Excess				1.091**			1.022
				(0.030)			(0.031)
Fall_Excess						2.174***	2.138***
						(0.088)	(0.091)
Constant	0.249	1.244	0.086	0.075	1.044	6.584	6.189
	(1.212)	(2.375)	(1.270)	(1.216)	(1.663)	(1.240)	(1.241)
Observations	91	91	91	91	91	91	91
Log Likelihood	−463.460	−553.871	−783.259	−779.254	−702.659	−673.766	−673.521
θ	34.886 (5.962)	7.909 (1.191)	26.916 (3.982)	29.376 (4.351)	15.757 (2.331)	29.659 (4.441)	29.819 (4.465)
Akaike Inf. Crit.	942.919	1,123.742	1,582.518	1,576.509	1,421.319	1,365.531	1,367.042

*p<0.1; **p<0.05; ***p<0.01

total population. Both find that the lower density of built area per person in neighborhoods plays a large role in higher influenza-related relative mortality. While this variable remains slightly significant in models for the protracted fall-winter 1918-1919 waves (as well as the total population size), the model fits more poorly. In both this and the spring wave, we find demographic indicators of population size and age composition to better fit a model describing the relative excess. Despite this interesting finding regarding model fit, we must remember that despite this, there is no real relationship between any of our variables and the mortality in the spring herald wave.

Despite other recent findings in Chicago [14], Literacy is not present in any best-fit model as an indicator (significantly or not) of mortality differences between baseline and epidemic mortality.

5 Discussion

The relationship between the dependent variables and regression model differ in each wave. While certainly, much variation in excess mortality is not explained via these models, the results may be indicative of varying levels of immunity throughout the population acquired from previous waves, especially provided that the level of mortality in each preceding wave has the strongest relationship to the observed excess mortality in the wave of interest. However, the strength of the preceding wave as an indication of current wave excess mortality may also simply indicate only that those neighborhoods with high excess mortality during the outbreaks continued to have high excess mortality when the epidemic returned. To this effect, the relationship between excess mortality during the spring wave and the protracted fall wave is strong, but nearly proportional. Also, this relationship is always positive, meaning that the the higher the excess mortality in the preceding wave, the studied waves mortality is predicted to also be higher.

In the spring, we find no significant relationship between our variables and the amount of neighborhood excess mortality. While other cities have reported strong, but not particularly lethal herald waves, the initial outbreak in May 1918 produced high levels of excess mortality, the waves peak being much higher than any observed week in the subsequent wave from fall 1918 through winter 1919. Perhaps the virulence and transmission speed of this particular wave was such that it did affect all parts of the city with equal ferocity.

In terms of mortality impact, the protracted fall wave, compared to other parts of Spain and the world, was muted in Madrid, continuing with three small peaks dispersed between September 1918 and April 1919. It is also the wave were a slight absolute excess mortality bump does occur in young adults, as has been found in many populations throughout the world. Here, we find the strongest relationships between the composition of a neighborhood and its mortality experience. As has been indicated in Chicago and Kristiana, areas and individuals with lower social status may have experienced a heightened mortality risk [13, 14]. Here, we find what may be an indication of this effect through the amount and type of housing found in a neighborhood, but we do not observe this in other variables that may have described the social makeup of the neighborhood, such as the mean HISCAM score, the average rental price, and the percentage of population that could both read and write.

The last wave of pandemic influenza struck with a vengeance between December 1919 and January 1920. In this wave, we also find little social impact on mortality, beyond that of the increased mortality in neighborhoods with higher numbers of stores and businesses as a percentage of total building units. The interpretation of this may be vague, but it could be a result of transmission dynamics in neighborhoods where people traveled to for work or to run errands.

While the results do not directly contradict recent findings, they also are not able to provide strong further evidence of a social dimension to influenza mortality during the Spanish flu outbreaks. Certainly, the historical nature of and available data for the analysis limit the extent to which we can examine the relationship. Yet as the overall mortality experience of the flu differed throughout the world, the extent to which social status within a single urban environment influenced mortality also likely varied. Our analysis provides another example from which researchers can advance research to disentangle this relationship between environment, social and material resources, and health. The continued effort to understand this relationship can lead to contemporary solutions in future influenza epidemics.

References

[1] J. D. Tamerius, J. Shaman, W. J. Alonso, K. Bloom-Feshbach, C. K. Uejio, A. Comrie, and C. Viboud, "Environmental predictors of seasonal influenza epidemics across temperate and tropical climates," *PLoS Pathog*, vol. 9, no. 3, p. e1003194, 2013.

[2] N. P. Johnson and J. Mueller, "Updating the accounts: global mortality of the 1918-1920" spanish" influenza pandemic," *Bulletin of the History of Medicine*, vol. 76, no. 1, pp. 105–115, 2002.

[3] J. D. Mathews, J. M. Chesson, J. M. McCaw, and J. McVernon, "Understanding influenza transmission, immunity and pandemic threats," *Influenza and other respiratory viruses*, vol. 3, no. 4, pp. 143–149, 2009.

[4] L. Simonsen, G. Chowell, V. Andreasen, R. Gaffey, J. Barry, D. Olson, and C. Viboud, "A review of the 1918 herald pandemic wave: importance for contemporary pandemic response strategies," *Annals of Epidemiology*, 2018/03/12.

[5] D. M. Morens and A. S. Fauci, "The 1918 influenza pandemic: insights for the 21st century," *Journal of Infectious Diseases*, vol. 195, no. 7, pp. 1018–1028, 2007.

[6] D. R. Olson, L. Simonsen, P. J. Edelson, and S. S. Morse, "Epidemiological evidence of an early wave of the 1918 influenza pandemic in new york city," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 31, pp. 11059–11063, 2005.

[7] E. C. Holmes, E. Ghedin, N. Miller, J. Taylor, Y. Bao, K. St George, B. T. Grenfell, S. L. Salzberg, C. M. Fraser, D. J. Lipman, *et al.*, "Whole-genome analysis of human influenza a virus reveals multiple persistent lineages and reassortment among recent h3n2 viruses," *PLoS biology*, vol. 3, no. 9, p. e300, 2005.

[8] E. Cordoba and A. E. Aiello, "Social determinants of influenza illness and outbreaks in the united states," *North Carolina Medical Journal*, vol. 77, no. 5, pp. 341–345, 2016.

[9] J. M. Nagata, I. Hernández-Ramos, A. S. Kurup, D. Albrecht, C. Vivas-Torrealba, and C. Franco-Paredes, "Social determinants of health and seasonal influenza vaccination in adults 65 years: a systematic review of qualitative and quantitative data," *BMC Public Health*, vol. 13, no. 1, p. 388, 2013.

[10] F. R. Van Hartesveldt, *The 1918-1919 pandemic of influenza: the urban impact in the Western World*. Edwin Mellen Press, 1992.

[11] S. M. Tomkins, "The failure of expertise: Public health policy in britain during the 1918—19 influenza epidemic," *Social History of Medicine*, vol. 5, no. 3, pp. 435–454, 1992.

[12] A. W. Crosby, *America's forgotten pandemic: the influenza of 1918*. Cambridge University Press, 2003.

[13] S.-E. Mamelund, "A socially neutral disease? individual social class, household wealth and mortality from spanish influenza in two socially contrasting parishes in kristiania 1918–19," *Social Science & Medicine*, vol. 62, no. 4, pp. 923–940, 2006.

[14] K. H. Grantz, M. S. Rane, H. Salje, G. E. Glass, S. E. Schachterle, and D. A. Cummings, "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, pp. 13839–13844, 2016.

[15] C. J. Murray, A. D. Lopez, B. Chin, D. Feehan, and K. H. Hill, "Estimation of potential global pandemic influenza mortality on the basis of vital registry data from the 1918–20 pandemic: a quantitative analysis," *The Lancet*, vol. 368, no. 9554, pp. 2211–2218, 2007.

[16] S.-E. Mamelund, "1918 pandemic morbidity: The first wave hits the poor, the second wave hits the rich," *Influenza and other respiratory viruses*, vol. 12, no. 3, pp. 307–313, 2018.

[17] L. Cilek, G. Chowell, and D. R. Farias, "Age-specific excess mortality patterns during the 1918-1920 influenza pandemic in madrid, spain," *American Journal of Epidemiology*, p. kwy171, 2018.

[18] B. Echeverri, "Spanish influenza seen from spain," *The'Spanish'Flu pandemic of*, 1918.

[19] P. S. Lambert, R. L. Zijdeman, M. H. D. V. Leeuwen, I. Maas, and K. Prandy, "The construction of hiscam: A stratification scale based on social interactions for historical comparative research," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 46, no. 2, pp. 77–89, 2013.

[20] L. A. Waller and C. A. Gotway, *Applied spatial statistics for public health data*, vol. 368. John Wiley & Sons, 2004.

[21] P. Elliott and D. Wartenberg, "Spatial epidemiology: current approaches and future challenges," *Environmental health perspectives*, vol. 112, no. 9, p. 998, 2004.

[22] M. Choynowski, "Maps based on probabilities," *Journal of the American Statistical Association*, vol. 54, no. 286, pp. 385–388, 1959.

[23] V. Gómez-Rubio, J. Ferrándiz-Ferragud, and A. López-Quílez, "Detecting clusters of disease with r," *Journal of Geographical Systems*, vol. 7, no. 2, pp. 189–206, 2005.

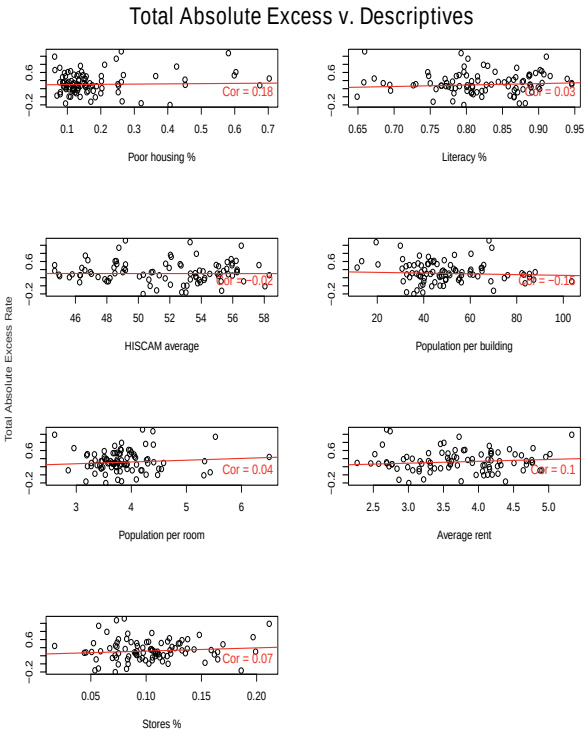
[24] V. Gómez-Rubio, P. Moraga, and J. Molitor, "Fast bayesian classification for disease mapping and the detection of disease clusters," tech. rep., Technical report, Universidad de Castilla-La Mancha, Spain, 2009.

[25] M. Burger, F. Van Oort, and G.-J. Linders, "On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation," *Spatial Economic Analysis*, vol. 4, no. 2, pp. 167–190, 2009.

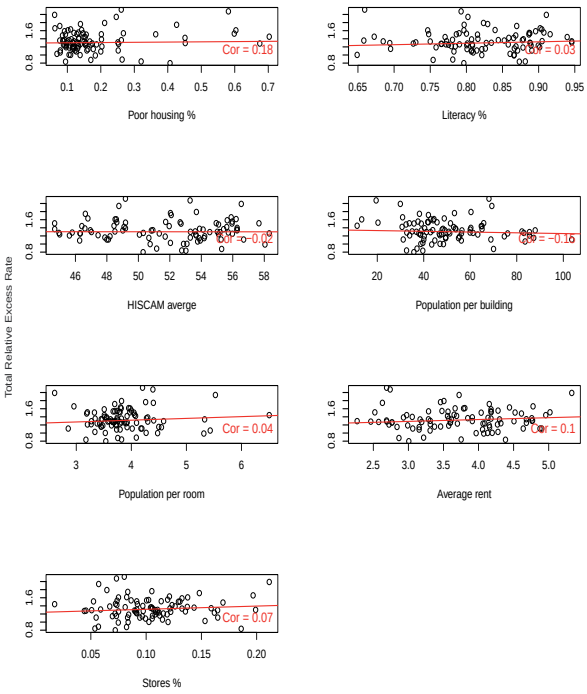
[26] W. Greene, "Functional forms for the negative binomial model for count data," *Economics Letters*, vol. 99, no. 3, pp. 585–590, 2008.

[27] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*. New York: Springer, fourth ed., 2002. ISBN 0-387-95457-0.

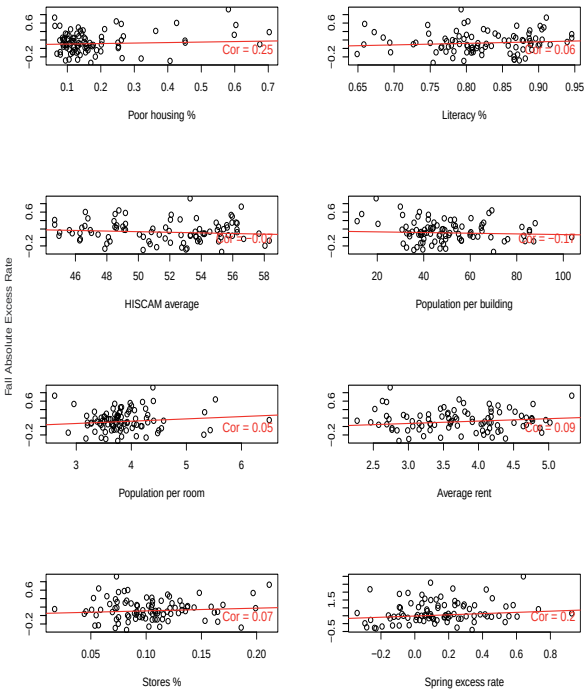
6 Appendix


Total Absolute Excess v. Descriptives
Total Absolute Excess Rate
Cor = 0.18
Poor housing %
Cor = 0.03
Literacy %
Cor = −0.02
HISCAM average
Cor = −0.16
Population per building
Cor = 0.04
Population per room
Cor = 0.1
Average rent
Cor = 0.07
Stores %

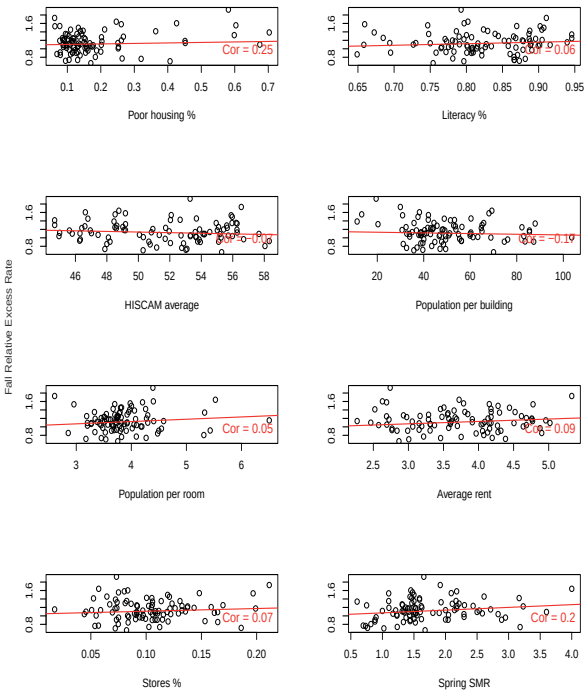
Total SMR v. Descriptives



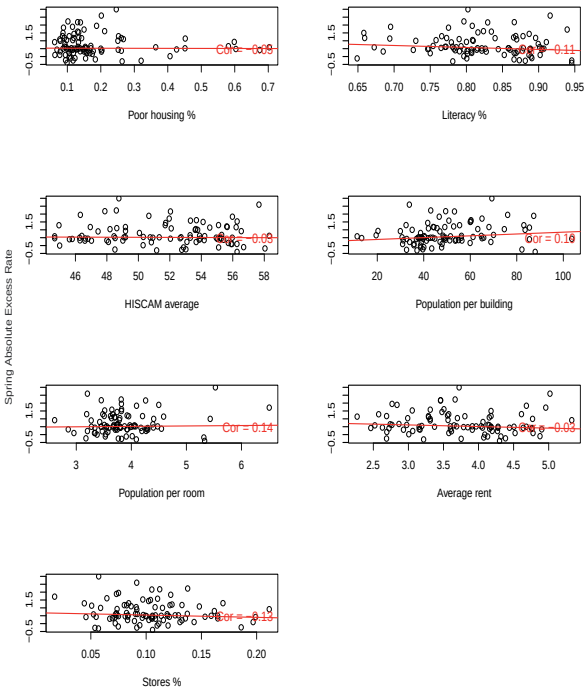
Fall Absolute Excess v. Descriptives



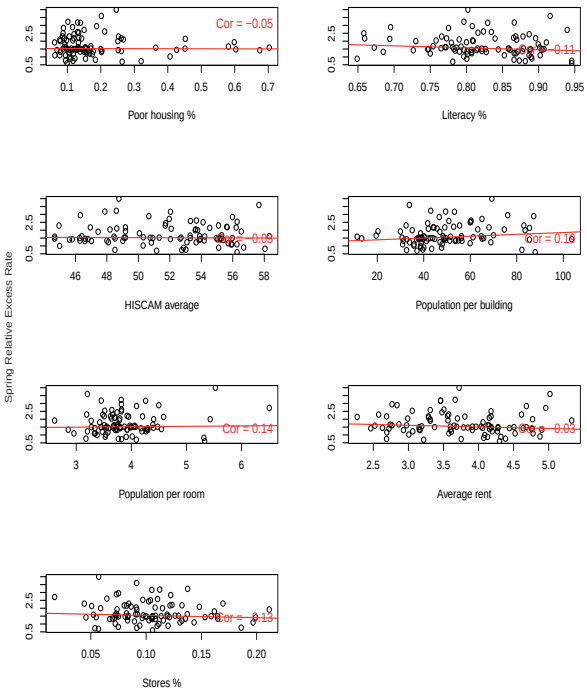
Fall SMR v. Descriptives



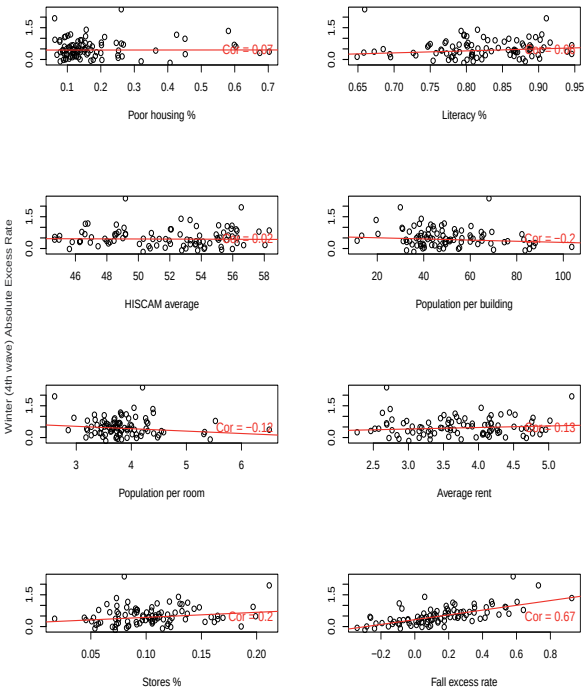
Spring Absolute Excess v. Descriptives

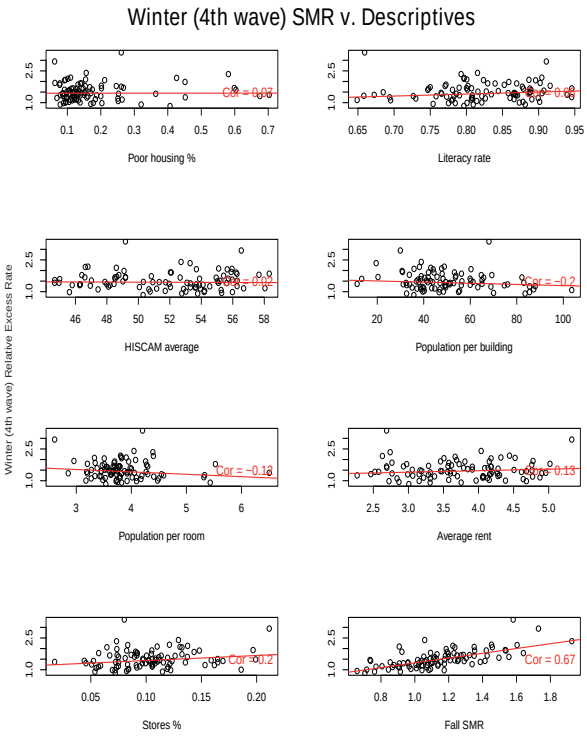


Spring SMR v. Descriptives



Winter (4th wave) Absolute Excess v. Descriptives




Winter (4th wave) SMR v. Descriptives
Winter (4th wave) Relative Excess Rate
Poor housing %
Cor = 0.05
Literacy rate
Cor = 0.18
HISCAM average
Cor = 0.02
Population per building
Cor = −0.2
Population per room
Cor = −0.18
Average rent
Cor = 0.13
Stores %
Cor = 0.2
Fall SMR
Cor = 0.67