Estimating Reproduction Numbers for the 1889-90 and 1918-20 Influenza Pandemics in the city of Madrid.

Jim Oeppen[*] Diego Ramiro Fariñas[†] Sara García Ferrero[†]

SALUD Y CIUDADES ÉN ESPAÑA, 1880-1940

SEMINARIO – BARCELONA 8-9 JULIO 2010

Keywords: Influenza; Pandemic; Reproduction Number; Madrid; Spain; 1889; 1918.

Abstract

The Reproduction Number of an epidemic is the number of secondary cases produced by each primary case in a totally susceptible population. Its magnitude, together with the Serial Interval between primary and secondary cases, is an important factor in the transmissibility of the disease and the possibility for intervention. For the epidemic to die out, the Reproduction Number must fall to less than unity. Large values indicate epidemics that may not be susceptible to interventions designed to reduce transmissibility below the critical threshold. It is surprising, given its importance for epidemic preparedness, that there are relatively few estimates of the Reproduction Number for 1918. Analyses of this pandemic in the United States and North-West Europe produce estimates of around 1.5 to 5.

The Madrid data comprise individual death records which allow us to examine a number of hypotheses with regard to covariates affecting the Reproduction Number, such as age, sex, and marital status. We also contrast the 1889-90 pandemic with successive waves of 1918-20. The city of Madrid represents an important example as there is evidence that the rise in relative mortality during 1918 was greater in Spain and Italy than in North-West Europe, North America and Australasia. So far, the explanations have been speculative rather than quantitative. This paper offers a clear example of the way in which historic data can contribute to the understanding of contemporary problems.

[*]Max Planck Institute for Demographic Research, Rostock, Germany.
[†]Institute of Economics, Geography and Demography, Spanish Council for Scientific Research, Madrid, Spain.

1 Introduction

The Reproduction Number of an epidemic is the number of secondary cases produced by each primary case in a totally susceptible population. Its magnitude, together with the Serial Interval between primary and secondary cases, is an important factor in the transmissibility of the disease and the possibility for intervention. For the epidemic to die out, the Reproduction Number must fall to less than unity. Large values indicate epidemics that may not be susceptible to interventions designed to reduce transmissibility below the critical threshold. Demographers are already familiar with these concepts from population dynamics as the Reproduction Number is equivalent to the Net Reproduction Rate and the Serial Interval is equivalent to generation length.

It is surprising, given its importance for epidemic preparedness, that there are relatively few estimates of the Reproduction Number for 1918. Analyses of this pandemic in the United States and North-West Europe produce estimates of around 1.2 to 3.75. These values are surprisingly low when compared with estimates as high as 20 for other influenza epidemics, and when compared with other infectious diseases. It suggests that the 1918 disaster was caused by very high case fatality rates, rather than by extreme transmissibility, and that control might have been feasible with aggressive intervention.

The Madrid data comprise individual death records which allow us to explore new methodologies for estimating Reproduction Numbers and to examine a number of hypotheses with regard to covariates such as age, sex, marital status and geographical propinquity. Aggregate studies show that the Relative Mortality Risk for infants was close to unity in 1918, but rose sharply with age so that children over 4 were at high risk. We will also contrast the 1889-90 pandemic with the successive waves of 1918-20. It is known that the age-specific response in these two pandemics was different and that the former occurred in an era when influenza had almost disappeared as a cause of morbidity and mortality, so that the proportion of the population with a naïve immune response to influenza was probably much higher than in 1918.

The city of Madrid represents an important example as there is evidence that the rise in relative mortality during 1918 was greater in Spain and Italy than in North-West Europe, North America and Australasia. So far, the explanations have been speculative rather than quantitative.

2 Historical Background to Influenza

Figure 1 shows the historical development of the influenza mortality rate in England and Wales on a log scale. While we must accept that cause of death data can be inaccurate, there are a number of strong features that are supported by other sources. The 1918 pandemic stands out as an event that occurs only once in the 150 year record. In some ways the 1889/90 pandemic is even more remarkable. The log scale makes it clear that the relative rise is greater than in 1918 or 1847. While reporting influenza as a cause of death may have become 'unfashionable' and this may have reinforced the decline in the late nineteenth century, the same fall to negligible mortality levels is apparent in other regions reporting influenza mortality: Scotland, New Zealand, urban Denmark, and Berlin. Denmark is particularly interesting because influenza case incidence, which was reported in Copenhagen, was at a very low level before the pandemic. This suggests that incidence and lethality were linked. In comparing periods, we must remember that the data are crude rates and the population is much older, on average, today. Nevertheless, the conventional, modern, medical arguments for the decline of influenza mortality cannot have applied in the late nineteenth century. The virus was undiscovered, and there were no antibiotics to control opportunistic co-infections. From the modeling point of view, it is likely that the 1889 population was immunologically 'naiive' compared with the situation three decades later.

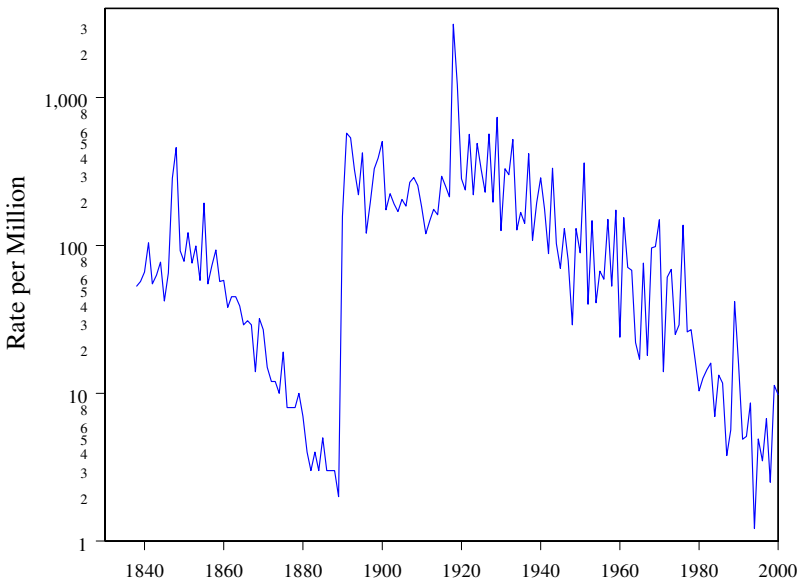


Figure 1: Influenza Mortality Rate: England and Wales.

2.1 The Timing of the 1889-1890 Influenza Pandemic in Madrid.

Influenza appeared in Madrid in the second week of December 1889 and by around the 19th to the 25th of December the pandemic had reached a crisis that lasted for four weeks. Valencia shared the experience of Madrid. Other Spanish populations with high incidence were: Coruña, starting from 22nd of December; from late December Saragossa, Irun, Segovia and Seville; from early January Valladolid, Tarragona, Santiago de Compostela; from mid January Figueras, Lerida and Balearic Islands; and the north of Catalonia in late January. During January most of Spain experienced influenza.[1] Apart from the first wave, there were another two in 1891 and 1892 and morbidity was very high, from 40% to 70% of the population.

In just two months, December 1889 and January 1890, there were 6315 deaths in Madrid. Of those deaths, 54% were due to Diseases of the respiratory system, and Acute Bronchitis and Pneumonia accounted for 43% of all deaths in the same period. The size of this epidemic can be compared to the second wave of the 1918-1919 Pandemic.[2] In that wave, influenza took three months (October to December 1918) to accumulate 5325 deaths. In the 1918-1919 wave, 553 influenza deaths were registered, in comparison with 1889-1890 when this diagnosis was underregistered.[3]

2.2 The Timing of the 1918-1920 Influenza Pandemic in Madrid.

The course of the 1918 pandemic in Spain has been extensively reviewed by Echeverri Dávila (1993), and in English by Echeverri (2003); Trilla et al. (2008a,b) and Erkoreka (2009).

The 1918-1920 influenza pandemic in Spain occurred in three waves. The first evidence of influenza was registered in Central Spain, in Madrid, Cuenca, Toledo and Salamanca, which showed increased mortality from influenza during May 1918. The first newspaper reports of influenza appeared in Madrid on May 22nd 1918, but it was thought that this mild illness had begun at the start of the month. The disease spread to all the country from these central areas, and disappeared after two months. Influenza mortality was not high and reached 0.04 to 0.65 per thousand during this first wave (Echeverri Dávila, 1993, 86-87). The most affected areas were Extremadura, Andalucía and great areas of the Castilla-La Mancha region.

The second wave began in August, although September and October represented the zenith of the epidemic, and ended late in December (Echeverri Dávila, 1993, 89). The second wave was the most important of the three. Mortality during October was 300% higher than the average for that month during previous years. In Spain 79484 deaths were officially registered as influenza: 52% of all deaths attributed to the pandemic during the period from May 1918 to June 1919 (Echeverri Dávila, 1993, 92). Mortality rates from influenza ranged from 0.5 to 14.0 per thousand: the higher figures in the Northern and Mediterranean regions of Spain.

The third wave, which took place from January to June 1919, accounted for 21094 deaths across Spain. It was the least explosive of the three and affected the areas that suffered most in the first wave. Mortality rates from influenza ranged by region from 0.07 to 1.4 per thousand (Echeverri Dávila, 1993, 93-94, 192). We can even talk about a fourth wave which took place during 1920, accounting for 17841 deaths from influenza for all Spain, particularly affecting children under 1 year old, the population without immunity to the influenza virus (Echeverri Dávila, 1993, 94).

[1] J.Teixidor y Suñol (1899) *La gripe. Naturaleza, formas y tratamiento.* Barcelona. San Martín de Provensals. p. 19.

[2] Porras Gallo, M. I. (1994) Una Ciudad en Crisis; La epidemia de Gripe de 1918-19 en Madrid. Tesis Doctoral. Facultad de Medicina. Universidad Complutense de Madrid.

[3] See Hauser, Ph. (1979) *Madrid bajo un punto de vista médico-social* (edición a cargo de Carmen del Moral, Madrid, Editora nacional), vol. 2, p. 101, for estimation of the underregistration of influenza deaths in 1889-1890.

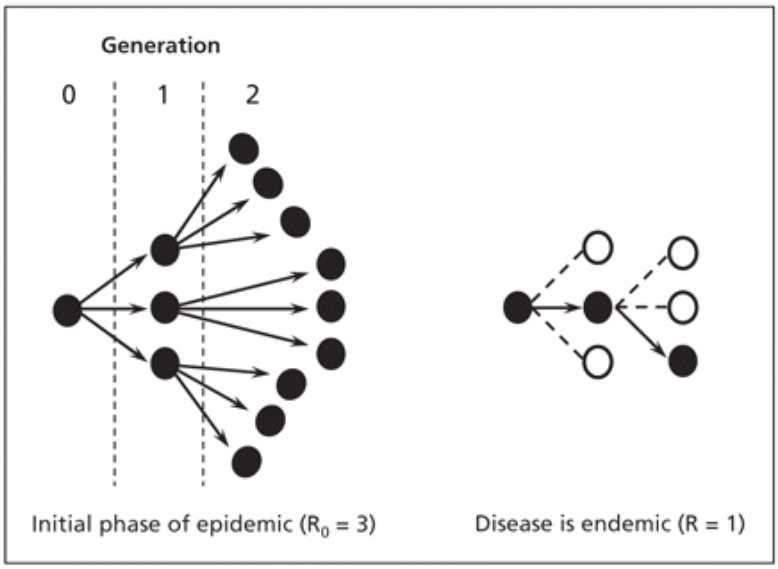
3 Reproduction Numbers

Human demographers are familiar with the concept of the Net Reproduction Rate (NRR), also represented by R_0. It is defined as the average number of daughters that a female birth cohort will produce during its reproductive span, while subject to a fixed age-specific pattern of maternity and mortality rates. Essentially, it is an inter-generational replacement rate and this is emphasised in the form

$$R_0 = NRR = e^{rT} \tag{1}$$

where r is the intrinsic growth rate of the population and T is the mean generation length (Preston et al., 2001, p.152). If $r = 0$, then $NRR = 1$ and the population reproduces itself exactly in each successive generation. Values of NRR that are greater or less than unity imply growth and decline, respectively.

Dietz (1993) notes that although this concept had been used in demography since 1886, it was introduced into epidemiology in 1952, but with different terminology. The generation length, T, corresponds to the average Serial Interval between linked cases, and is the sum of the average durations of the latent and infectious periods. R_0 is referred to as the basic Reproduction Number, or Ratio, of the disease. It is interpreted as the average number of secondary infections in a totally susceptible population resulting from one infectious individual over their complete infectious period. As with human demography, a value of unity represents a crucial threshold. If $R_0 > 1$ then the infection will spread; if less than unity it will die out. Figure 2 shows the initial epidemic phase in a population with total susceptibility, exponential growth, and $R_0 = 3$; and an endemic phase, where a proportion of the population is immune and $R_0 = 1$.



Source: Pan-InfORM, *Can. Med. Ass. J.*, 2009, 181 (3-4), 171-3.

Figure 2: Epidemic and Endemic phases of an Infectious Disease.

As with human fertility, it is useful to consider the proximate determinants. R_0 depends on three parameters: the duration of the infectious period; the contact rate; and the probability that an infection results from a contact. Dietz (Dietz, 1993) sub-divides the last factor into infectivity and susceptibility.

Over the past 800 years of recorded human history, there may have been as few as 4 or 5 cycles in NRR, but the basic Reproduction Number for influenza exhibits an annual cycle. During the northern-hemisphere Summer, R_0 is usually less than one, but it rises above unity during the Winter.[4] In some years, the value is much greater than unity and an epidemic or pandemic emerges. Having an estimated or expected value for R_0 during an epidemic may be of crucial importance for public health interventions and planning. Values close to unity may be reduced below the threshold for spread by simple measures designed to avoid inter-personal contact, such as closing schools or the isolation of infected individuals. If the expected value for R_0 is large, then the best strategy may be to reduce the susceptible population through vaccination.

The average generation time, or serial interval, can be estimated as the sum of the latent period, when a person is infected but not infectious, and half the length of the infectious period. The ability of a disease to spread rapidly is an exponential function of the Reproduction Number and the Serial Interval, as shown in (4). Influenza has a short generation time and the initial stages of infectiousness are asymptomatic. Both features contribute to rapid spread and the problems of planning successful intervention.

3.1 Estimation Methods for Reproduction Numbers

Although we have defined the Reproduction Number from a demographic viewpoint, there are a number of basic ways to estimate it, and many variants: see Dietz (1993); Heffernan et al. (2005). These range from the very simple - as used in this paper - to the complex (Chowell, 2009). Sophisticated deterministic and stochastic models with explicit transmission can be useful for simulations, intervention studies, and scenario building, but they are ruled out in this study by data limitations. We have mortality incidence data for pandemics, and no direct information on recovery, contact or transmission. Additionally, as we are concerned with pandemics we cannot assume that the disease is in an endemic equilibrium.

In this study we follow the demographic tradition in (4). An alternative with our data would be to calculate the Effective Reproduction Number, R_n, from

$$R_n = \Lambda^2(L \times D) + \Lambda(L + D) + 1 \tag{2}$$

where Λ is the growth rate of the cumulative number of deaths during the rising phase of the pandemic, and L and D are the average durations of the latent and infectious periods (Vynnycky et al., 2007; Jackson et al., 2010). R_0 and R_n are linked by

$$R_0 = \frac{R_n}{s} \tag{3}$$

where s is the susceptible proportion at the start of the epidemic. Thus R_0 can be interpreted as an upper bound on R_n, and equivalent when there is no immunity in the population. Noting that the proportion immune is $p_c = 1 - s$, we can calculate the immune proportion of the population that would prevent an epidemic starting by $p_c = 1 - 1/R_0$. This is termed 'herd immunity' - the extent to which the overall population may be protected by a proprtion of immunes.

Both methods share two advantages. First, by estimating the slope during the rising segment of the epidemic, one can ignore the decline in the number of susceptibles caused by death and

[4]The mechanism is not fully understood.

immunity. Second, the slope is independent of the level of under-registration, but only if the level remains constant.

At this first stage in our research, we have reservations about estimating R_n - a method that requires fitting a model to the cumulation of 'noisy' data when the breakpoints may be difficult to define - and choose to estimate R_0.

3.2 Influenza-Specific Reproduction Numbers

Table 1 suggests that, even in a pandemic, influenza has a relatively low Reproduction Number compared to other diseases.

Table 1: **Reproduction Numbers for Infectious Diseases.**

Disease	Transmission	R_0
Measles	Airborne	12–18
Pertussis	Airborne droplet	12–17
Diphtheria	Saliva	6–7
Smallpox	Social contact	5–7
Polio	Fecal-oral route	5–7
Rubella	Airborne droplet	5–7
Mumps	Airborne droplet	4–7
HIV/AIDS	Sexual contact	2–5
SARS	Airborne droplet	2-5
Influenza (1918 pandemic strain)	Airborne droplet	1.5-5

We show a range of values for the 1918 pandemic, but the true extent of variation in the Reproduction Number of influenza is difficult to assess. In addition to the variety of estimation methods described in the previous section, the types of data, the situations in which they occur, the degree of aggregation, and their quality, are immensely variable. Studies have used mortality and morbidity - either total or cause-specific; open and closed communities (cities versus ships, prisons, etc.); hospital entries; adult male labour-force records; etc.. Time units may be days or weeks.

Chowell (2009, 22) identifies 11 papers, in addition to his chapter, that estimate R_0 for 1918, with values ranging from 1.4 to 5.4. He notes that the estimates vary according to

> the specific location and pandemic wave considered, type of data, estimation method, and level of spatial aggregation, which has ranged from small towns to entire nations with several million inhabitants. The variability of R_0 estimates suggests that local factors, including geographic and demographic conditions, could play an important role in disease spread.

Vynnycky et al. (2007) restrict their estimates to morbidity data and calculate the Effective Reproduction Number, which is less than or equal to R_0. For open communities, Their estimates, adjusted to a totally susceptible population, range from $2.4 - 4.3$ for open communities, and from $2.6 - 10.6$ for closed ones.

As far as we know, there are no estimates for the 1889-90 pandemic.

4 A Simple Model of an Epidemic

Our conceptualisation of a typical influenza epidemic is extremely simple. We expect a background level of deaths (or cases) interrupted by a short period of rapid increase and decline, and that these four segments can be modeled by log-linear components. We also assume that the three breakpoints between the log-linear segments can be identified. Figure 3 illustrates the principle.

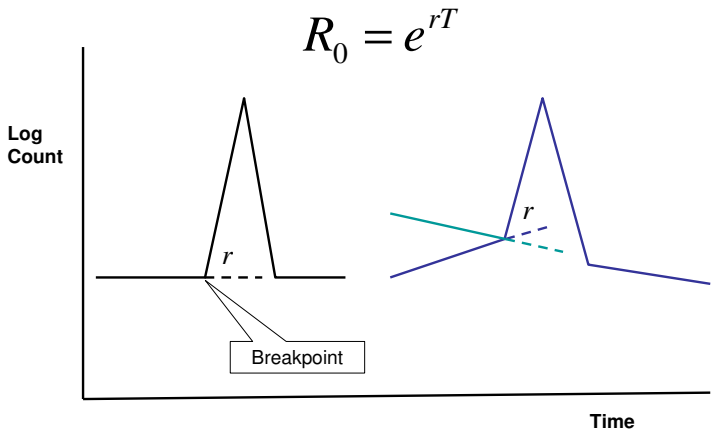


Figure 3: Schematic Representation of an Influenza Epidemic.

If the assumptions for the definition of R_0 are met, r in equation (4) can be estimated from the incremental slope of the log count during the rising segment of the epidemic.[5] We emphasise the importance of the incremental estimate in the second panel. If the background count is rising (the left blue segment), then an estimate of r based on the rising slope of the epidemic would be an over-estimate of the true Reproduction Number since a fraction of the rising count is 'normal'. Conversely, r would be underestimated if the background count is falling (the green line). In the literature, researchers sometimes subtract an estimate of the pattern of the 'normal' count and estimate the slope from the 'excess'. If cause of death data are available, one can extract the 'non-epidemic' component. In many cases this is not possible because the 'normal' pattern is not known; or the issue is ignored. In these latter situations, one would expect the Reproduction Number to be under-estimated for a Spring epidemic (green line), and over-estimated for Autumn and Winter (blue line).

Conventional methods for estimating r from the rising slope of the epidemic count assume that the relevant section of the time-series, and thus the two breakpoints, can be identified exogenously. This may be problematic if the population is small, the data are noisy, and the data have been aggregated into weekly counts. Our aim is to simultaneously estimate the breakpoints and the incremental slope.

[5] In human demography, the equivalent estimate of r can be obtained from the slope of a simple regression model on time, using the population, births or deaths in a stable population as the dependent count.

5 Segmented Regression

To estimate the slopes and breakpoints of the epidemic, we follow the model proposed by Muggeo (2003, 2008). Initially we only consider the first three elements of our epidemic model: the baseline, the first breakpoint, and the rising epidemic segment.

Muggeo suggested that a segmented linear relationship between a response variable and a covariate Z for observation $i = 1, 2, \ldots, n$ can be represented by

$$\beta_1 z_i + \beta_2 (z_i - \psi)_+ \tag{4}$$

where $(z_i - \psi)_+ = (z_i - \psi) \times I(z_i > \psi)$ and $I(\cdot)$ is the indicator function equal to unity when the logical statement is true.

In this model, β_1 is the left slope, β_2 is the difference-in-slope corresponding to the epidemic, and ψ is the breakpoint. The indicator function ensures that the change in slope, determined by β_2, is only applied after the breakpoint and is zero before. This formulation can be extended to multiple breakpoints, each with its own estimated β and ψ. The number of breakpoints, but not their position, is exogenously determined after inspection of the (possibly smoothed) time-series.

The variable of interest in this study, the number of deaths per time unit, is a count rather than a continuous variable, so it is appropriate to assume the log-linear Poisson form of the General Linear Model, with Z representing time. The model can include additional, non-segmented covariates, x_i, with linear parameters δ.

$$\begin{aligned} Y &\sim Pois(exp(\eta)) \\ \eta &= \alpha + x'_i\delta + \beta_1 z_i + \beta_2 (z_i - \psi)_+ \end{aligned} \tag{5}$$

Muggeo (2003) shows that the non-linear component in (4) has an approximate linear form which can be estimated iteratively via

$$\beta_1 z_i + \beta_2 (z_i - \psi)_+ + \gamma I(z_i > \psi)^- \tag{6}$$

where $I(\cdot)^- = -I(\cdot)$ and γ is a parameter that represents the gap between the two straight lines at the estimated breakpoint.

Given a guess at the initial breakpoint $\tilde{\psi}$, model (4) is estimated by iteratively fitting (6). At each iteration, the estimated breakpoint is updated by $\hat{\psi} = \tilde{\psi} + \hat{\gamma}/\hat{\beta}_2$. At the current estimate of the breakpoint, $\hat{\psi}$, the gap between the two straight lines is represented by $\hat{\gamma}$. At convergence this gap should be close to zero, which can be tested. Convergence is guaranteed, if the breakpoint exists and the variance is low.

A graphical representation of the model is shown in Figure 4. By our definition, the epidemic component $\beta_2(z_i - \tilde{\psi})_+$ must be zero at the breakpoint, when $z = \psi$. If the initial estimate of the breakpoint is poor, the epidemic 'wedge' cannot match the trapezoidal area under the curve and the model compensates via the step-function $\hat{\gamma}$. Then the update formula for ψ is applied and, in this case, moves the estimated breakpoint to an earlier date $\hat{\psi}$. As the pair of equations is iterated to convergence, the nuisance parameter $\hat{\gamma}$ shrinks towards zero.

In the full representation of the model in Figure 3 we simultaneously estimate multiple slopes and breakpoints. If the slope of the left segment is statistically equal to zero, we refit the model with $\beta_1 = 0$. This results in narrower confidence bands for the location of the breakpoint.

Muggeo (Muggeo, 2008) considers the advantages of estimating the locations of the breakpoints. He suggests that assuming fixed breakpoints underestimates the standard errors of the other parameters because the uncertainty about the breakpoints is not taken into account.

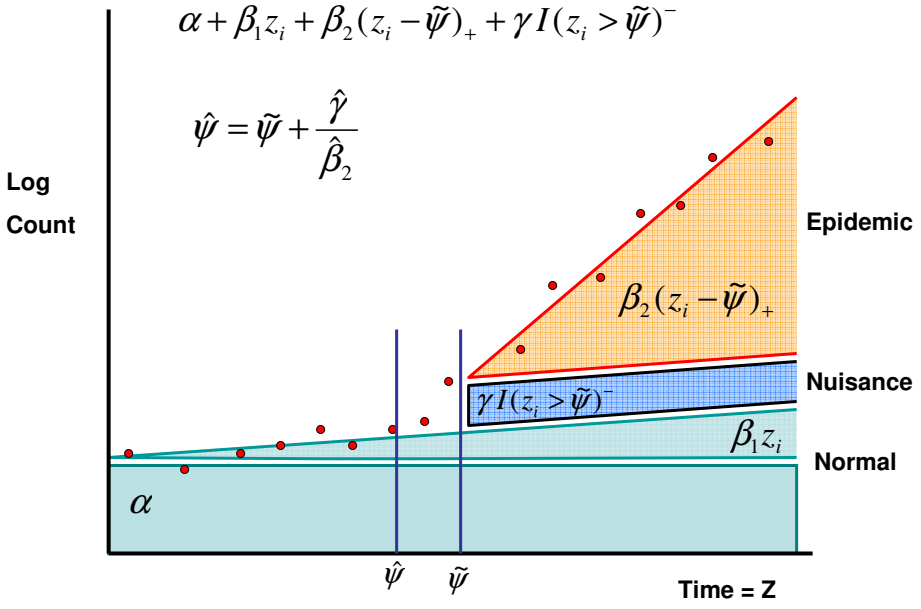


Figure 4: Schematic Representation of the Segmented Model.

The models in this paper were fitted using the software package 'segmented' (Muggeo, 2009) within the R statistical environment (R Development Core Team, 2009).

6 Data and Sources

This study makes use of two main data sources, Civil Register records on Deaths, and Listings of Burials for the City of Madrid published in the Official Bulletin of the State, with the title *Gazeta de Madrid* in Spanish. Madrid had c.470,000 inhabitants in 1887, and c.750,000 in 1920. Data were collected for the years corresponding to two major influenza pandemics, 1889-90 (data collected for 1889-1892, with c.20,000 burials per year) and 1918-19 (data collected for 1918-1921, with c.18,000 deaths per year). Detailed information can be found in Table 2.

Table 2: **Total burials and deaths for the City of Madrid in 1889-1892 and 1918-1921.**

Gazeta de Madrid			Civil Register		
Year	Burials	Collected	Year	Deaths	Collected
1889	20205	20205	1918	18974	8834
1890	21645	21645	1919	18330	15059
1891	17322	17322	1920	18055	14482
1892	16526	16526	1921	16215	4979
Total	75698	75698		71574	43354

Source: collected from *Gazeta de Madrid* and Civil Register. It was not possible to collect all deaths from the Civil Register due to the closure for reforms of the Municipal Archives.

The Civil Registration System in Spain was finally established by law in 1870, after several attempts during the XIX$^{\text{th}}$ Century.[6] The 1869 Constitution established the secularization of the Civil Register. The law established the need to record every individual so that they could claim their rights. Therefore the Civil Registration System was the guarantee that these rights would be preserved under the control of the State. From the beginning, it was compulsory for each individual to declare and register their births, marriages and deaths before a municipal Judge. For each individual the following information was recorded: date of inscription in the register, names and surnames, sex, age, birthplace, civil status, occupation, date of death, complete address where the death took place, cause of death and the Cemetery where the body was buried, adding some additional information regarding violent deaths.

With regard to the *Gazeta de Madrid*, it was the antecedent of the actual Official Bulletin of the State and during the XVII$^{\text{th}}$ to the early XX$^{\text{th}}$ centuries was named as *Gazeta de Madrid*, in which all the laws approved by the state were published and where most official announcements appeared. As in other European nations, these official bulletins started to be published from mid XVII$^{\text{th}}$ century onwards. From 1837 onwards the *Gazeta* was published daily. From 1888 until 1901 the information concerning all the burials taking place in the City of Madrid was published on a daily basis. The Section of Health and Bureau of Statistics of the General Directorate of Health and Beneficence, which was responsible to the Ministry of State, was in charge of publishing the records in the *Gazeta*. For each individual the following information was recorded: names and one surname, sex, age, civil status, date of death, cause of death and complete address where the death took place.[7]

[6]For more information on the formation of the civil registration system in Spain, see Ramiro (1998), 'La evolución de la mortalidad en la infancia' and Reher, D., Valero Lobo, A. *Fuentes de información demográfica en España.* Cuadernos Metodológicos Núm 13. CIS.

[7]For information of the formation and characteristics of the *Gazeta de Madrid* see Núñez del Prado, 'De la Gaceta de Madrid al Boletín Oficial del Estado', *Historia de la Comunicación Social* Vol.7 (2002) 147-160.

These two sources differ mainly in one important aspect, the *Gazeta* includes fetal deaths and deaths during the first 24 hours: deaths that do not appear at the Civil Register books. The problem of under-registration of both births and deaths in Spanish data comes from the legal concept of a *live birth*.[8] Registers in some other European countries, such as Belgium and France are similarly defective.[9] In Spain before 1975, children who died in the first 24 hours of life were not counted as live births. For the purpose of registration in the civil registers, as well as in the official series of Spanish vital statistics (*Movimiento Natural de la Población*), children dying during the first day of life were included until 1975 in a category for stillbirths and children dying in the first day of life.[10] Therefore, stillbirths and first day deaths were specified and recorded in separate books. The origin of this problem can be attributed to ambiguities in the regulations in force from 1889 which resulted in many of these deaths being recorded separately from the general records of the civil registers as *Legajos de abortos or Cuadernos de fetos* (Records of miscarriages or stillbirths).[11] The majority of these documents have been destroyed over time, although the figures had been included in the official published series of vital statistics. However, it is not possible to use this information today to reconstruct local demographic series. There are no significant differences for the rest of the deaths, just small differences in the registration of the day in which the death was recorded in the *Gazeta* in comparison with the Civil Register.

7 Results

7.1 The 1889/1890 Pandemic

In fitting the models, we arbitrarily choose a 'window' of data. Obviously this must include the epidemic, but we also want the estimated breakpoints to be as far as possible from the temporal limits of the data, without adding unnecessary segments and their breakpoints. The objective of maximising the data available before or after a breakpoint is to make its estimation as accurate as possible.

Figure 5 shows the segments and breakpoints fitted to total deaths in the 1889/90 epidemic. The red lines are from a model that estimates a slope for the first segment. The blue lines are estimated under the assumption that the initial slope is zero. The short horizontal lines at the bottom of the plot show the estimated locations of the breakpoints for each model with two standard-error limits. It is clear that setting the initial slope to zero is a worthwhile simplification.

As an experiment, a fourth breakpoint was specified with May 1st, 1890 as the starting position. The software reports an error stating that the search has moved outside the time-span of the data.

The Figure shows that the simple segmented model provides a good fit to the log counts of deaths. However, it is also clear that there is a change in the mortality regime before the supposed start of the epidemic. Table 3 shows the input breakpoint dates and the final estimates, together with the standard errors and the length of each segment in days. The estimated breakpoint on 11$^{\text{th}}$ November predates the explosion of the pandemic by about five weeks and we will return to this issue below. The standard errors for the breakpoints are remarkably small, but we must remember that the model has access to all the data, and thus the benefit of hindsight. This is important in appreciating the short period of the pandemic segment. The model suggests that a planner would only have had 15 days to determine that a pandemic was occurring and to mount an effective

[8] Marcelino Pascua, *La mortalidad infantil en España*, 13-18; Antonio Arbelo, 'Necesidad demográfico sanitaria', 393-405, and Antonio Arbelo, *La mortalidad de la infancia en España* 152-156.

[9] E.A. Wrigley, 'Explaining the rise in marital fertility in England in the 'long' eighteenth century', *The Economic History Review*, LI, (3), (1998), 435-464, esp. 445; and C. Gourbin and G. Masuy-Stroobant, 'Registration of vital data: are live births and stillbirths comparable all over Europe?' *Bulletin of the World Health Organization* Vol. 73 (4) (1995), 449-60.

[10] The first to notice this problem was Antonio Arbelo, 'Necesidad demográfico sanitaria', 393-405; and *La mortalidad de la infancia en España*, 154.

[11] Ramiro (1998), 'La evolución de la mortalidad en la infancia'

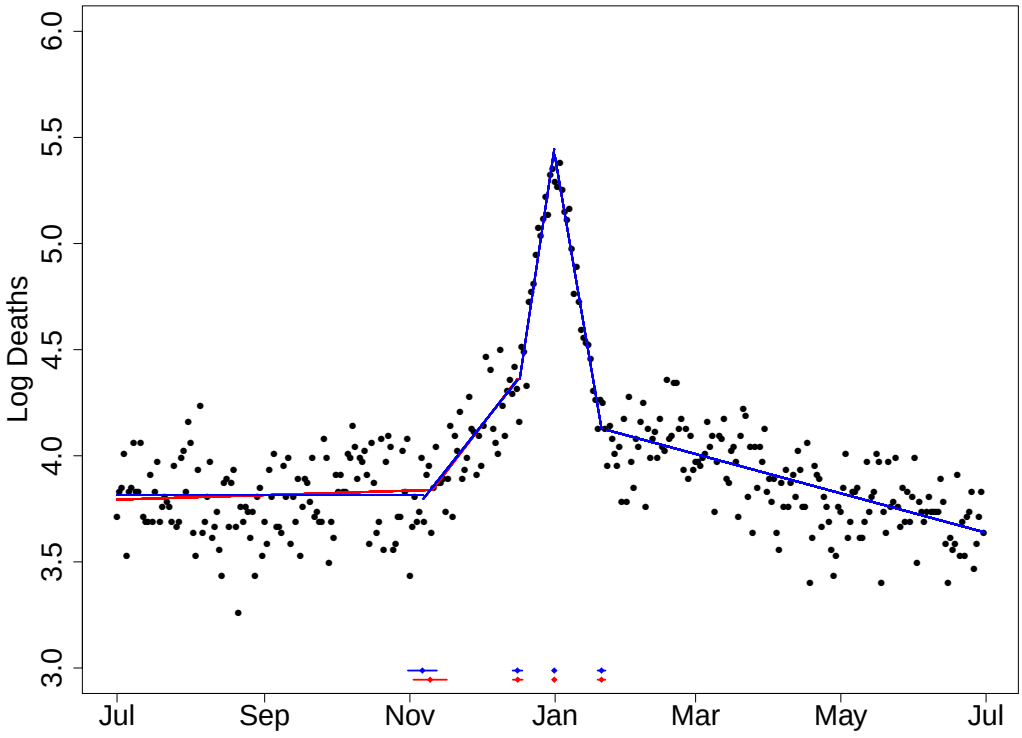


Figure 5: The 1889/90 Influenza Pandemic in Madrid.

policy.

Table 3: **Breakpoints: Total burials and deaths, Madrid 1889-1892.**

Breakpoints			Segment
Initial	Est.	St.Err	Length
$1889-07-01$			132
$1889-11-01$	$1889-11-11$	3.8	36
$1889-12-01$	$1889-12-17$	1.1	15
$1890-01-01$	$1890-01-02$	0.4	19
$1890-02-01$	$1890-01-21$	0.8	161
$1890-06-30$			

The estimates of the Reproduction Numbers are shown in Table 4. For comparative purposes we have adopted the range of Serial Intervals suggested by Vynnycky et al. (2007), with 4.0 days as the preferred estimate. The R_0 estimates are lower than the figures reported for the 1918 pandemic and indicate that an effective vaccination rate of about 25% could have stopped the pandemic. The columns R_0' and p_c' are based on the conventional slopes rather than the incremental slopes we propose. Separating the data by sex indicated that females had slightly higher values than males, but we did not control for differences by sex in age and marital status, although this is feasible with these data and the model.

Table 4: **Reproduction Numbers: Total burials and deaths, Madrid 1889-1892.**

Interval	R_0	R_0'	p_c	p_c'
2.5	1.15	1.19	0.13	0.16
4.0	1.25	1.32	0.20	0.24
6.0	1.40	1.52	0.28	0.34

As part of the investigation into the pre-pandemic segment during November and early December, we divided the data by age-groups. The Reproduction Numbers by age, with two standard error limits, are shown in Figure 6. The age-groups are 0-5, 5-20, and then ten-year groups to 80+. The vertical lines indicate $R_0 = 1$ and the estimate for all ages combined. It seems that there are two regimes. Young people under 20 are less affected by the pandemic, although the small number of deaths between 10 and 19 makes it difficult to detect the separation between the two regimes. This means that the overall estimate is misleading. Two features are surprising. First, the values for young people indicate relative immunity, but the immune systems of their elders should have acquired experience of a wider spectrum of influenza viruses. Second, there is no indication that the mortality risk rose among the elderly as the error limits overlap, although there is a some evidence that the disease spread more rapidly among those aged 40-49. It is well known that there were marked age-effects in the 1918 pandemic, but this is not always the case. In their study of morbidity during the 1968 (Hong Kong) influenza pandemic, Jackson et al. (2010) found that age-effects were small.

Figures 7 and 8 show the segments and breakpoints for young people and adults, respectively. Although the timing of the declining segment is common to both groups, the origin is not. For young people, the pandemic seems to be the continuation of a longer-running process, whereas the adult pattern matches our conceptualisation of a classic pandemic. The adult model could be simplified by dropping the first breakpoint.

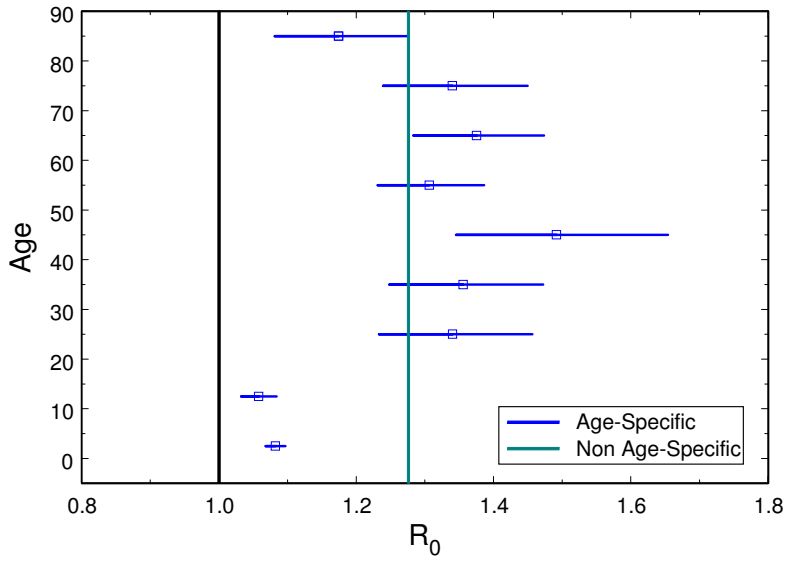


Figure 6: The 1889/90 Influenza Pandemic in Madrid: Reproduction Numbers by Age.

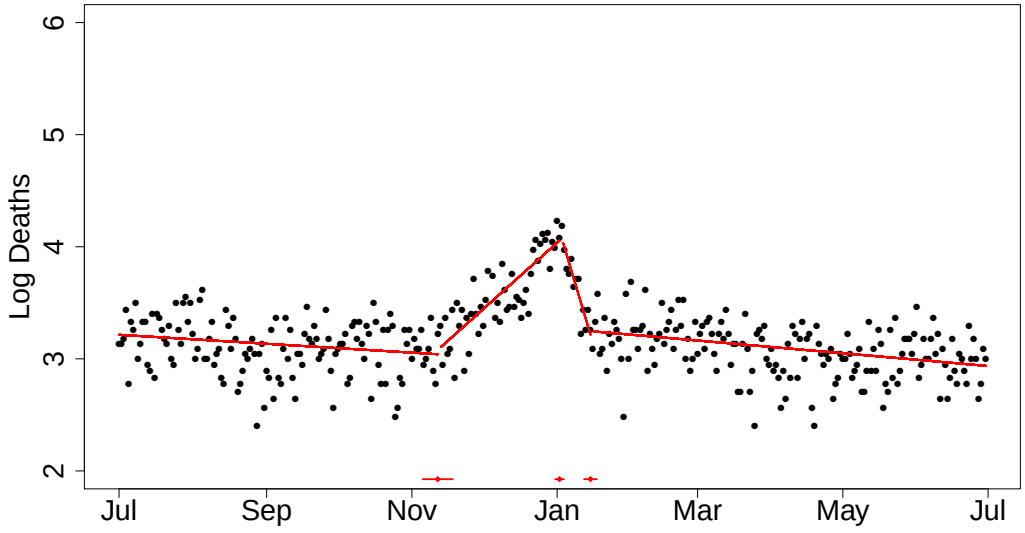


Figure 7: The 1889/90 Influenza Pandemic in Madrid: Young people < 20.

7.2 Preliminary Results for the 1918/1920 Pandemic Waves

Data entry for the period 1918-1920 is not complete, so the preliminary results presented below should be treated with caution. First, we do not yet have full geographical coverage of the city and there could be a selection effect as areas that include hospitals may be over- or under-represented. Second, the available counts do not include those who were recorded as 'fetos', although the data are available and will be entered later.

Only four districts are entered for the whole period from 1918 to 1922. These are Buenavista, Centro, Chamberi, and Hospicio. The time series is shown in Figure 9.

The graph confirms that Madrid experienced the three waves observed in other countries: Spring 1918, Autumn 1918, and Winter 1919/20. The Autumn wave is not extreme, which has been 3experienced in some other locations, but may reflect the limited area covered by the districts.

For the preliminary estimation, we restrict our attention to the third wave. We have eight of the ten districts for the period covering the third wave in 1919/20 so the results are probably representative. They are Buenavista, Centro, Chamberi, Congreso, Hospicio, Hospital, Inclusa and Latina. The estimates are shown in Figure 10. For comparative purposes we have re-estimated the model for 1889/90, excluding the 'fetos' deaths, and plotted the results offset by exactly 30 years e.g. 01-01-1890 is plotted at 01-01-1920.

If these preliminary results are confirmed when all the data are entered, they show a remarkable result. The third wave of the 1919/20 pandemic in Madrid shares the timing and shape of the 1889/90 pandemic. Why this should be true is not clear, but the richness of the individual data for Madrid means that we will be able to check if the two pandemics share the same patterns with respect to age, sex, cause of death, and geography.

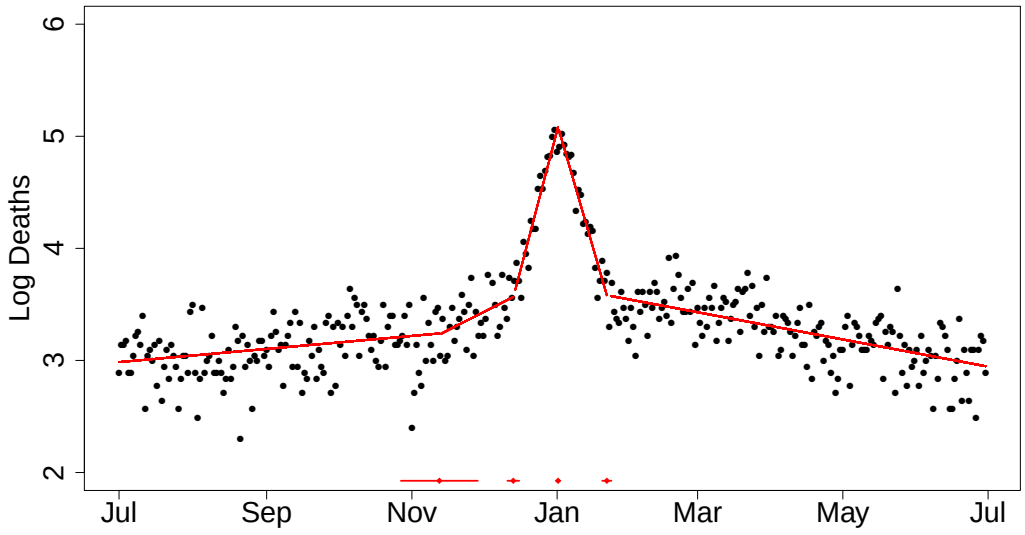


Figure 8: The 1889/90 Influenza Pandemic in Madrid: Adults $\geq$ 20.

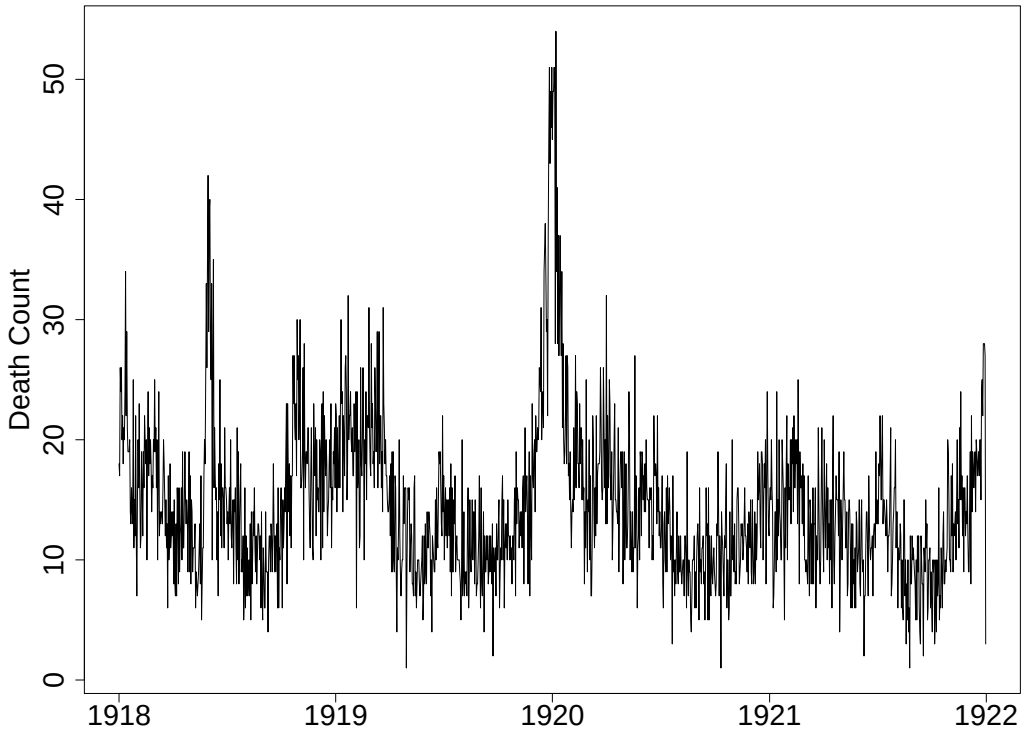


Figure 9: Deaths in 4 districts of Madrid (not including 'fetos').

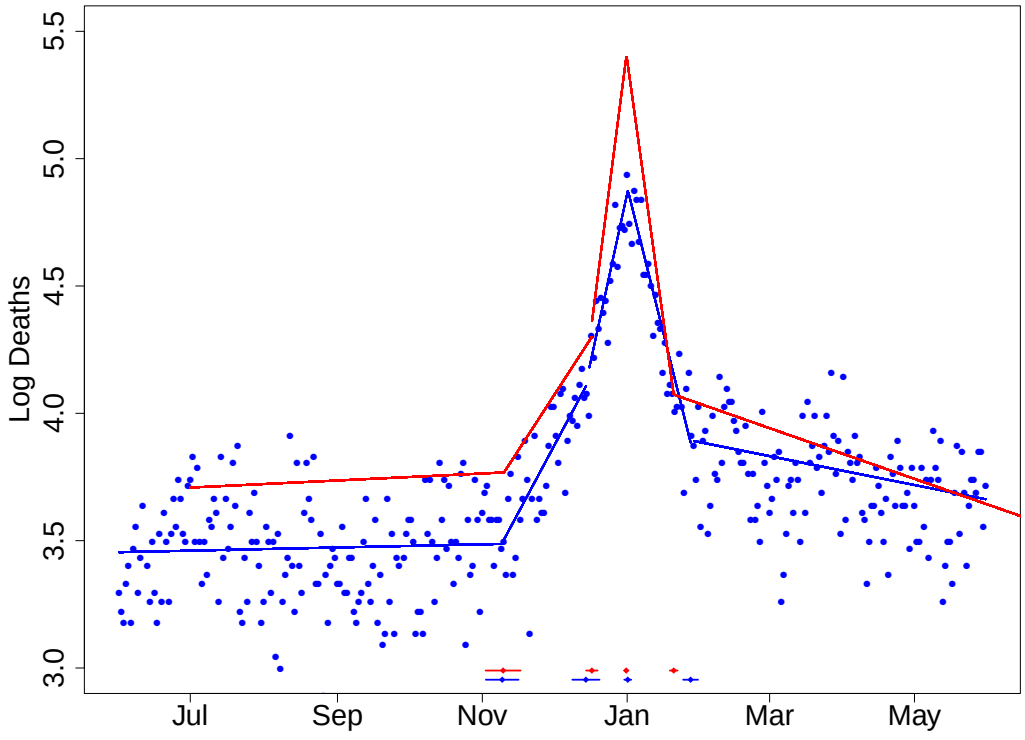


Figure 10: Influenza Pandemics in Madrid: 1889/90 red, 1919/20 blue.

8 Discussion

The Madrid data are probably unique in historical research into influenza pandemics. We know of no individual mortality database of this size, either for 1889/90 or 1918/20. Having individual data, recorded on a daily basis, for a population that falls between the size of Chicago and New York offers a rich opportunity. This paper no more than scratches the surface of what might be done. Although our model formulation is capable of including covariates, we have not yet used those that are available: sex, marital status, cause of death, and geographic area. We expect them to give us indirect insights into varying contact rates and susceptibility to influenza.

References

Chowell, G. (2009). The basic reproduction number of infectious diseases. In *Mathematical and Statistical Estimation Approaches in Epidemiology*, Chapter 1, pp. 1–30. Springer Netherlands.

Dietz, K. (1993). The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research 2*(1), 23–41.

Echeverri, B. (2003). Spanish influenza seen from Spain. In H. Phillips and D. Killingray (Eds.), *The Spanish influenza pandemic of 1918-19*, Chapter 11. Routledge.

Echeverri Dávila, B. (1993). *La gripe española. La pandemia de 1918-1919.* Number 132 in Colección Monografías. Centro de Investigaciones Sociológicas.

Erkoreka, A. (2009). Origins of the Spanish influenza pandemic (1918-1920) and its relation to the First World War. *J. Molecular and Genetic Medicine 3 (2)*, 190–194.

Heffernan, J., S. RJ, and W. LM (2005). Perspectives on the basic reproductive ratio. *Journal of the Royal Society, Interface / the Royal Society 2*(4), 281–93.

Jackson, C., V. E, and M. P (2010). Estimates of the transmissibility of the 1968 (Hong Kong) influenza pandemic: evidence of increased transmissibility between successive waves. *American Journal of Epidemiology 171*(4), 465–78.

Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine 22*, 3055–3071.

Muggeo, V. M. R. (2008). segmented: an R package to fit regression models with broken-line relationships. *R News 8/1*, 20–25.

Muggeo, V. M. R. (2009). *segmented: Segmented relationships in regression models.* R package version 0.2-5.

Preston, S. H., P. Heuveline, and M. Guillot (2001). *Demography: measuring and modeling population processes.* Malden, MA: Blackwell Publishers.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Trilla, A., G. Trilla, and C. Daer (2008a). The 1918 "Spanish Flu" in Spain. *Clinical Infectious Diseases 47*, 668–673.

Trilla, A., G. Trilla, and C. Daer (2008b). Correction to: The 1918 "Spanish Flu" in Spain. *Clinical Infectious Diseases 47*, 987.

Vynnycky, E., A. Trindall, and P. Mangtani (2007). Estimates of the reproduction numbers of Spanish influenza using morbidity data. *International Journal of Epidemiology 36*(4), 881–9.