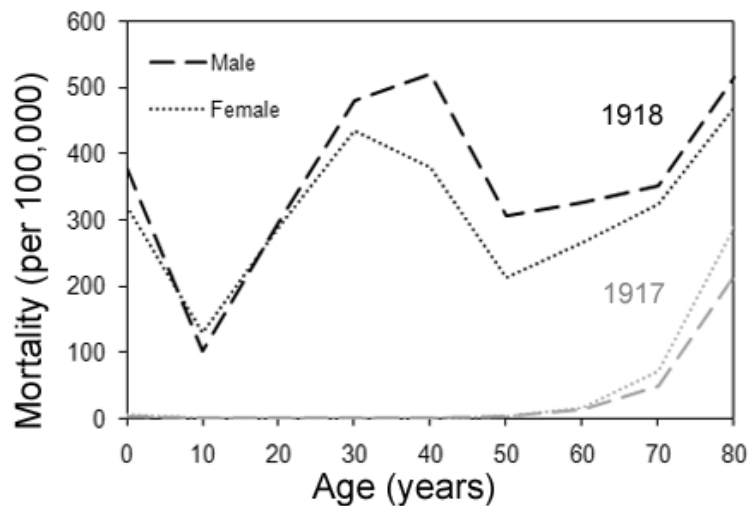# The Role of Tuberculosis
in Characterizing the Age-specific Risk of
Severe Influenza from 1918-19

Thesis

August 2010

Student:

Welling Oei

Tutor : Prof. J.A.P. Heesterbeek, PhD
Reviewer : Prof. M.J.M. Bonten, MD PhD
Supervisor : H. Nishiura, MD PhD

Age-specific mortality of influenza and pneumonia (per 100,000 population)
in the Netherlands during the epidemic in 1917 (grey) and 1918 (black)

Abstract

The W-shaped age-specific mortality curve from the latest influenza pandemic in 2009, caused by the new influenza A (H1N1), has only been previously seen in one other flu epidemic, the highly fatal influenza pandemic of 1918, referred to as Spanish flu. An understanding of this atypical W-shaped feature of the Spanish flu, compared to the usual U-shape seen in other pandemics, is required to improve decision making. The aim of our study is to develop a formal hypothesis that tuberculosis (TB) was responsible for the W-shaped mortality in the Spanish flu pandemic.

We obtained data from published epidemic records of Spanish flu in the United States of America, Japan, Iceland and the Netherlands, an epidemiological investigation of Spanish flu in TB and non-TB individuals in a Swiss sanatorium and TB mortality data across age, time and space in the three countries during the period 1900-1940. Three distinct steps of analyses were performed: Descriptive analyses of influenza mortality, morbidity, and case fatality ratio; Causal inference 2x2 table of TB infection vs influenza deaths; And analyses of epidemiological trends of TB mortality that separate the effect of age, period, and cohort.

The study shows that the W-shaped pattern was seen not only in the mortality, but also in the case fatality, suggesting an underlying risk factor elevating the risk of influenza death. Causal inference between TB and influenza mortality given influenza infection showed that among those non-TB flu individuals, none died of influenza. The Age-Period-Cohort (APC) model was the best model to explain trends of TB mortality. The model showed that TB was 'washed out' during 1918, i.e. deaths resulting from the Spanish flu was in fact subtracted from future TB mortality, which was illustrated by a spike and followed by a constant decline thereafter.

We concluded our study demonstrated, through various theoretical exercises, that TB may have been responsible for the W-shaped mortality curve of the Spanish flu.

Key-words: Spanish flu, 'w-shape', tuberculosis, age-period-cohort model

Preface

This prayer has always been the source of my strength during the writing of my thesis. This thesis records the whole process I have gone through during my first research project. I must admit that there were times when I nearly got lost and had no clue how to proceed, given that I had no background in epidemiology of infectious disease at the start of the project. My passion to gain proper knowledge of infectious disease has led me to get reacquainted with mathematical approaches which I abandoned more than ten years ago. I am lucky that I was given a project tailored to my 'undeveloped capability' which encouraged me to learn tremendously from it and try to deliver the best I could. However, this project also taught me that there are things that are not attainable, given that I was trying to prove something that occurred nearly a century ago using the limited record system of that time, or things that are beyond my comprehension that sometimes frustrated me. In the end, I realize that I should accept the current limitations, of the project and of my own potential, and I can only hope that this thesis will be of any use to future research.

Welling Oei, Den Haag, August 2010

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Influenza pandemics since the 20th century

An influenza pandemic is an epidemic of an influenza virus that spreads on a worldwide scale and infects a large proportion of the human population. There have been three influenza pandemics in the 20th century, in 1918, 1957, and 1968, which are known as the Spanish (H1N1), Asian (H2N2), and Hong Kong (H3N2) influenza respectively [17]. The new influenza A (H1N1) virus which caused the latest influenza pandemic in 2009 was less lethal than previous pandemics but concern was raised that it might become more lethal as it spread [22]. Given the possibility of this virus' mutation or re-assortment with current circulating influenza virus strains such as the highly pathogenic variation H5N1, understanding the previous pandemics and their common epidemiological features is crucial to help design the most effective countermeasures. As of 22 July 2010, the cumulative number of confirmed human H5N1 cases worldwide reported to the World Health Organization (WHO) reached a total of 297 deaths among 501 cases (`http://www.who.int/csr/disease/avian_influenza/country/cases_table_2010_07_22/en/index.html`).

1.2 The atypical mortality curve of Spanish flu

The most devastating recorded human influenza pandemic is the Spanish flu. It was estimated that one-third of the world's population contracted the flu and more than 2.5% of those infected died [32]. The most peculiar feature of Spanish flu was the
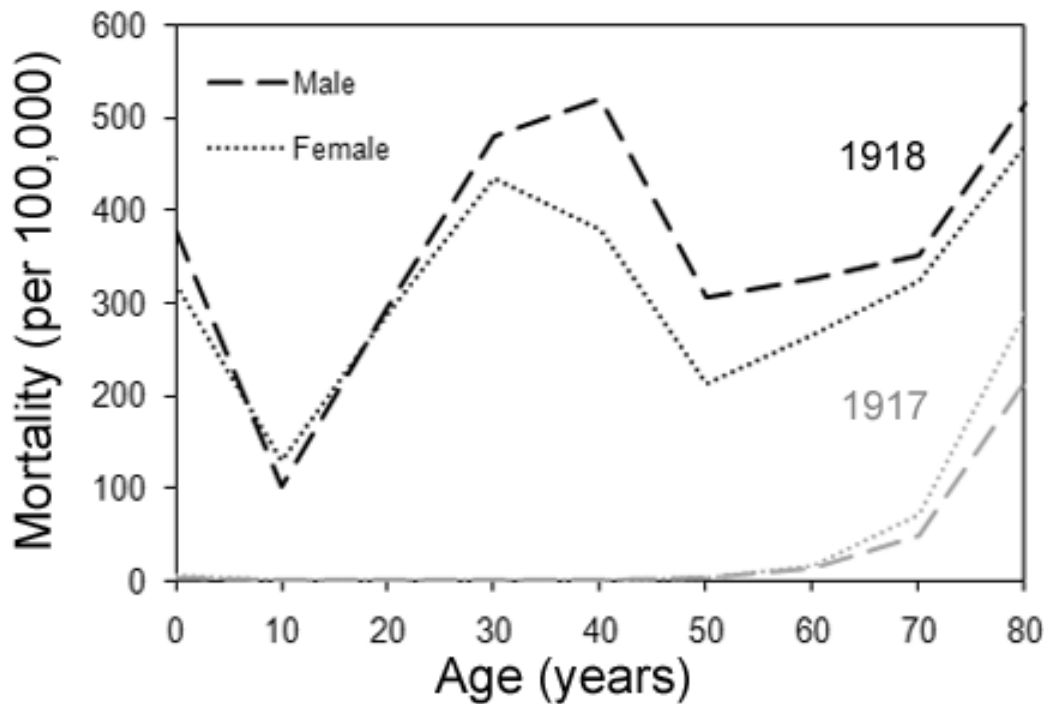
Figure 1.1: Age-specific mortality of influenza and pneumonia (per 100,000 population) in the Netherlands during the epidemic in 1917 (grey lines) and 1918 (black lines). The dashed- and dotted lines represent the male and female population for both years, respectively. The influenza death per population during seasonal influenza epidemic in 1917 shows the typical 'U-shape pattern' while the pandemic in 1918 results in the 'W-shaped' pattern, i.e. additional peak of death among young adults.

atypical W-shape of the age-specific mortality curve [29, 33], which was not observed in other pandemics apart from the 2009 pandemic [27]. Other influenza pandemics, as well as seasonal influenza epidemics, disproportionately killed the very young and the elderly, resulting in a 'U-shaped' age-specific mortality curve (fig. 1.1). The 1918-19 Spanish influenza pandemic also caused high mortality among those aged 20-40 (and especially those aged 25-35), resulting in the 'W-shaped' curve [9, 32]. That is, one of the distinguishing features of the 1918-19 pandemic was the additional peak in mortality among young adults.

1.3   Hypotheses for the 'W-shaped' curve

No satisfactory explanation has been put forward for the observed W-shaped mortality pattern. Nevertheless, historians and demographers suggested several hypotheses:

1. The virus responsible for the 1918-19 influenza pandemic was closely related to a virus of the same subtype H1N1 that might have circulated prior to 1918. This offered acquired immunity to middle-aged and elderly persons [20, 21].

2. World War I occurred in 1914-1918, coinciding with the Spanish flu pandemic. It is not difficult to imagine that poor nutrition and low socioeconomic status due to the war could have increased both TB and influenza death among young adults [9]

3. Cytokine theory, i.e. a hyper reaction of the immune system that potentially causes severe damage and death, might explain the more severe outcomes observed among young adults (who have stronger immunity) and the fewer deaths observed in the very young and the elderly (who have weaker immune systems) [1, 12].

4. Another potential co-morbidity: Scarlet fever [7, 28], i.e. a clinical syndrome characterized by high fever and rash, resulting from infection and toxin of a strain of Streptococcus pyogenes which was first identified in the early 20$^{\text{th}}$ century

5. In 1918-19, there was an underlying co-morbidity that caused an elevated risk of death among young adults compared with other influenza epidemics. It has been suggested that those with tuberculosis (TB) in 1918 were more likely to die of influenza compared with those individuals without TB [25].

Among these hypotheses, the proposed causal link between TB and influenza among young adults has been studied twice before. The first study was conducted by Noymer and Garenne in 2000 by exploring the time-trends of tuberculosis mortality in the United States [25]. The second was conducted by Noymer in 2009 by analysing individual data of the Union Army in the USA during the late 19$^{\text{th}}$ century [24]. Much stronger evidence is called for and a multi dimensional analysis could prove to be useful in supporting this hypothesis. It should be noted that other studies from the early 20$^{\text{th}}$ century have claimed that TB was a protective factor against an acute infection such as influenza [31, 23]. The present study examined if TB was a potential causal risk factor for an elevated risk of influenza death. Given the previous note on the protective role of TB in reducing the risk of influenza death, we need to test our hypothesis by examining aggregated historical datasets and investigating this issue using different theoretical methods. Our analyses are
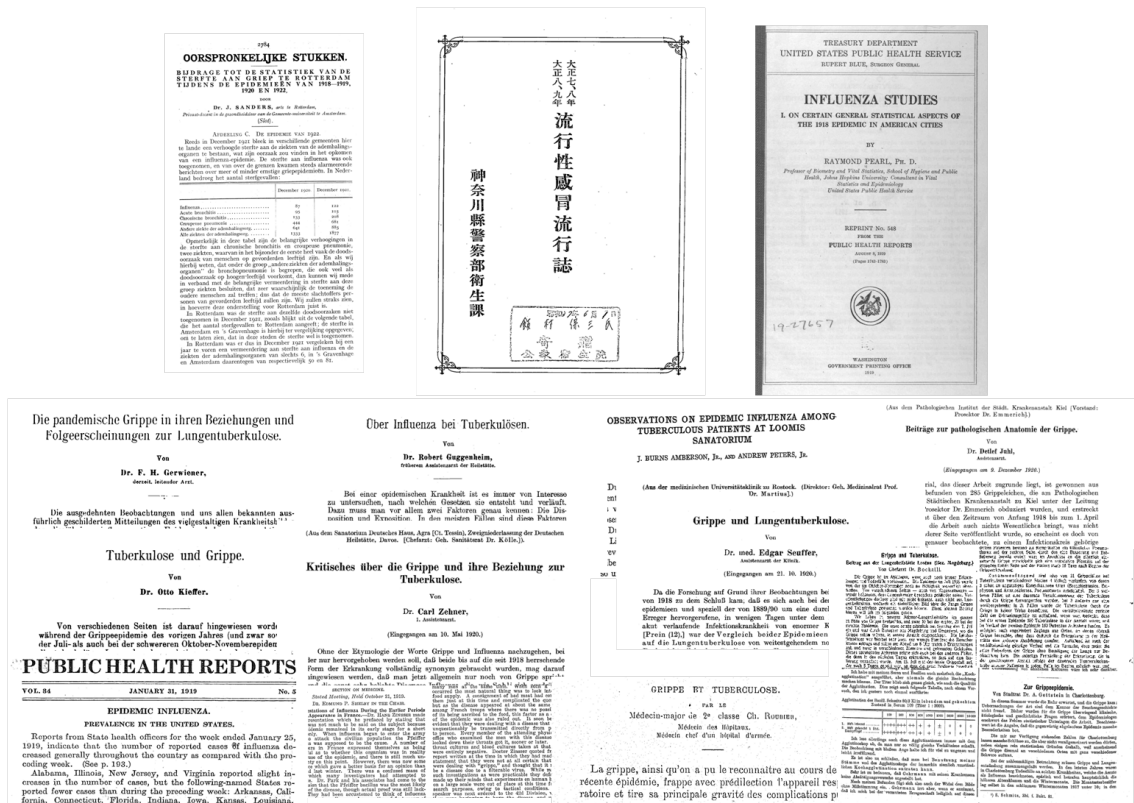
Figure 1.2: Various sources of published historical datasets of influenza and tuberculosis mortality included in the study: the United States of America, Japan, Iceland, Switzerland, and the Netherlands.

conducted on historical datasets of influenza and TB in the United States of America, Japan, Iceland, Switzerland, and the Netherlands.

1.4   Research objective and research questions

The present study aims to clarify if TB was responsible for elevating the risk of influenza death among young adults. We defined three research questions to achieve this in three separate steps:

1. Is W-shape also seen in the age-specific case fatality of the 1918-19 influenza pandemic?

2. Did tuberculosis increase the risk of influenza death?

3. Was the prevalence of tuberculosis 'washed out' following the 1918-19 pandemic?

Testing this hypothesis is of practical importance for today's population and is useful for future pandemic planning, because the prevalence of TB remains high in some developing countries and is highly variable worldwide [6, 25].

1.5   Approaches

To clarify the causal link between TB and influenza during the 1918-19 influenza pandemic both directly and indirectly, we tested our hypothesis using the following three major steps:

1. We estimated age-specific mortality, morbidity, and the case fatality ratio in the 1918-19 influenza pandemic. The presence of the W-shaped age-specific case-fatality ratio was examined to see if there was an underlying causal risk factor (or factors) for influenza death among young adults, i.e. age-groups 21-30 and 31-40 years. Identification of the W-shaped age-specific case fatality ratio firmly implies the presence of an underlying risk of death. Through analyzing multiple datasets of this kind, we offer the most appropriate interpretation of this issue, clearly distinguishing the risk of infection from the risk of death.

2. We tested the association between TB and 1918-19 influenza mortality. Using observational epidemic data in specific health care settings, the association was measured by odds ratio, chi square and Fisher's exact test under the null hypothesis that there is no relation between TB infection and death from influenza. Due to limited historical data, we performed univariate analysis alone.

3. We explicitly analyzed the epidemiological time course of TB using the age-, period- and cohort-model, which can separately quantify these three effects (i.e. age-effect, period-effect, and the cohort-effect). We examined how the influenza pandemic would have influenced TB mortality by taking into account the effects of age, period and cohort. If our hypothesis holds, TB mortality should have been washed out by the Spanish flu. This would be characterized as a rise in TB mortality during the pandemic, immediately followed by a decline after 1918. This pattern implies that the slow rate of TB mortality

was subtracted from future mortality by the explosive influenza pandemic, resulting in a significant decline in further mortality following 1918 [24].

1.6   Report layout

Following the introduction, Chapter 2 will describe the methodology of the research. Chapter 3 and Chapter 4 summarize the findings and present the discussion respectively. Finally, Chapter 5 concludes this report and gives some recommendations for further research in the future.

Chapter 2

Materials and Methods

2.1 Study design

The design for this research is the retrospective analysis of published historical data. Age-related epidemiological features of influenza are descriptively studied, and subsequently, the causal relationship between TB and influenza is tested by analyzing the observational epidemic data and by performing univariate analysis. Since confounding factors for the causal relationship cannot be eliminated using the historical data alone, and because misclassifications of both TB and influenza will most likely influence our conclusion on the causal relationship, we further performed time-trend analysis of TB mortality stratified by age and gender.

The domain for this study is human and the study populations depend on three separate analyses: (A) the general population for the first analysis (i.e. descriptive study of the age-specific distribution of Spanish influenza) which experienced the 1918-19 influenza pandemic and lived in the United States, Iceland, Japan and the Netherlands; (B) for analysis of Spanish influenza outbreak, those TB-infected and non-TB-infected individuals in particular health care settings (a TB sanatorium in Switzerland, a TB hospital in Japan, the catchment area of public health service in Maryland) who were affected by the influenza outbreak; (C) those who died of TB at a certain time and age in the United States, Japan and the Netherlands.

Accordingly, the inclusion criteria for (A) are:

1. Those who remained healthy, who were infected with influenza (cases), and who died of influenza (deaths).

2. In various geographic locations (USA, Iceland, Japan, and the Netherlands).

The inclusion criteria for (B) are:

1. Healthy (non-TB) and TB-infected individuals.

2. In specific health care settings: a sanatorium, a hospital, and a public service's catchment area.

The inclusion criteria for (C) are:

1. Populations who died of TB and were reported in the records

2. In various countries (USA, Japan, and the Netherlands)

At the beginning of this research, we conducted a preliminary study consisting of two analyses to answer the research questions by conducting the aforementioned separate analyses using smaller datasets. For the first step, a number of 42,920 influenza cases and 730 influenza deaths from a total of 146,203 individuals in various localities in the United States were analysed descriptively [4]. For the second step, 102 TB patients and 33 non-TB employees in a TB sanatorium were investigated, following the diagnoses given in the historical publication [11]. Details of this preliminary study will be described in the following section, together with various additional datasets.

2.2   Data collection

Information on influenza cases and deaths was extracted from published epidemic records of Spanish flu in the United States of America (USA) [4, 8], Iceland [10], Japan [5] and the Netherlands [3]. Epidemiological investigations of influenza cases and deaths among TB-infected and non-TB individuals were extracted and analyzed from published medical records of a Switzerland sanatorium [11], a hospital in Japan

[13], and a public health survey in Maryland, USA [26], while TB mortality data were obtained from annual reports in the United States of America, Japan and the Netherlands [34, 16, 30, 3].

Influenza deaths were defined as all deaths listed as specifically caused by influenza and pneumonia, whereas TB deaths were defined as all deaths listed as specifically caused by any form of tuberculosis infection. For the USA, datasets were stratified by 10-year age-groups and annual TB mortality data were available from the year 1900 to 1940. For Japan and the Netherlands, datasets were stratified by gender and age (5-year age-groups for Japan and 10-year age-groups for the Netherlands) and annual TB mortality data were available from 1899-1943 for Japan and 1901-1940 for the Netherlands. Data were presented per 100,000 of population according to the population survey of the relevant year.

2.3   Data analysis

2.3.1   Descriptive analysis

We present data on the influenza pandemic in 1918-19 in terms of mortality, morbidity, and case fatality ratio for the USA [4, 8] and Japan [5], and mortality only for Iceland [10] and the Netherlands [3] because no data of cases per population could be obtained for the latter two countries.

Mortality is a measure of the number of deaths (in general or due to a specific cause) in a certain population, per unit of time. Morbidity is the number of infected individuals in the population during a given time period. The case fatality ratio is the ratio of deaths per cases, over a certain period of time. Figure 2.1 illustrates how mortality curves are decomposed into morbidity and the case fatality ratio to demonstrate the distinction between risk of dying and the risk of dying given infection. Consistent W-shaped patterns found not only in mortality, but also in the case fatality ratio, indicate that there is an underlying risk factor (co-morbidity) that elevates the risk of dying among young adults (given the presence of infection).
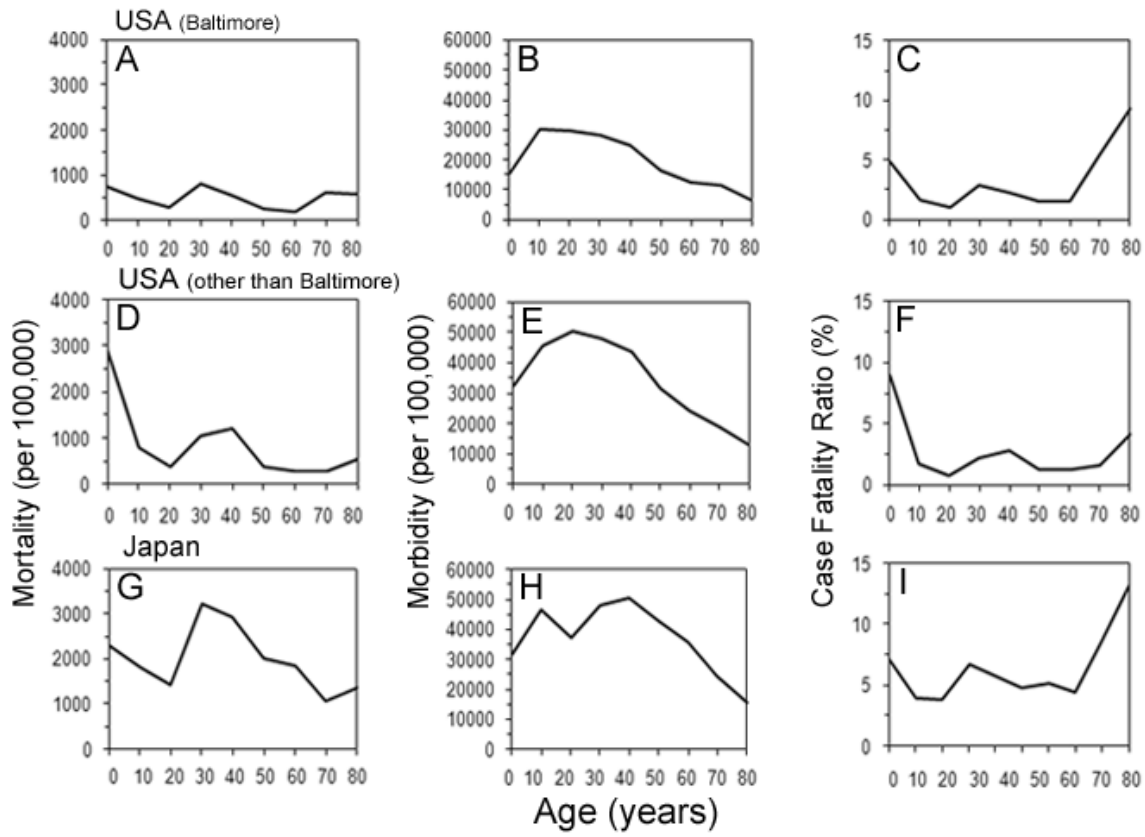
Figure 2.1: Age distribution of influenza pandemic in 1918 in the United States of America & Japan. Note: Panels A,B,C; D,E,F; and G,H,I show respectively mortality, morbidity, and the case fatality ratio in Baltimore, localities other than Baltimore, and Japan [8, 5]. Mortality and morbidity refer to as the total number of deaths per 100,000 of population and the total number of cases per 100,000 of population respectively. The case fatality ratio is the proportion of deaths among cases.

United States of America

Two sources of datasets from published historical reports for public health service in the USA are used in this study [4, 8]. The Frost datasets [8] show the number of cases and deaths from influenza and pneumonia (all forms) occurring among males and females of different ages in Baltimore and certain other localities in Maryland (Cumberland, Lonaconing, Frederick, Salisbury and three rural districts in Frederick, Washington and Wicomico Counties) during the influenza outbreak in 1918. Figure 2.2 presents this data in terms of mortality, morbidity and the case fatality ratio, stratified by gender and 10-year age-groups, distinguishing those of Baltimore and localities other than Baltimore. The figure shows the W-shaped pattern for mortality and the case fatality ratio. Panels D and F show some missing datapoints
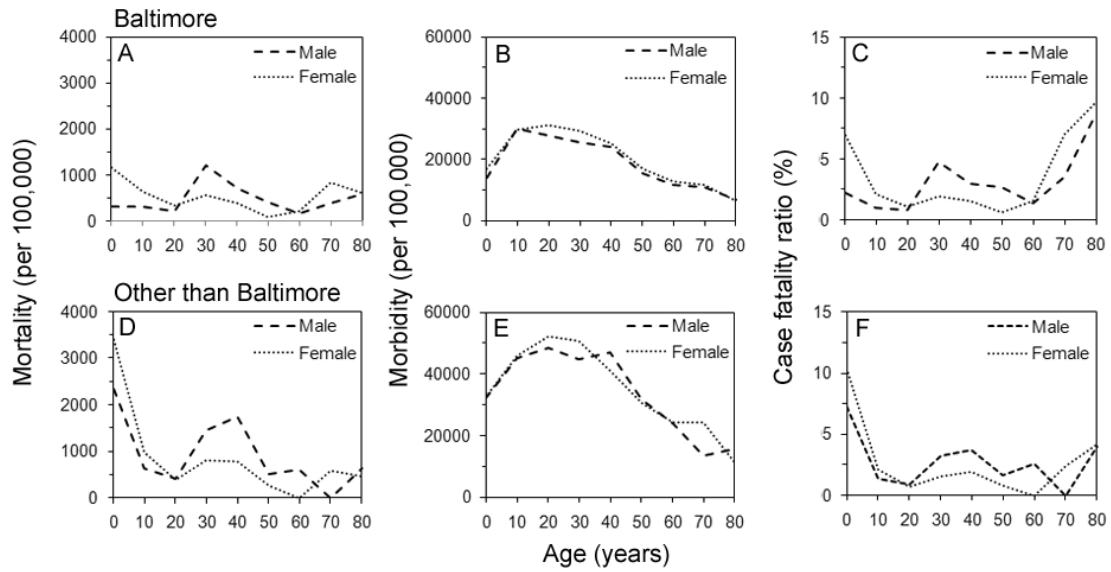
Figure 2.2: Age-specific distribution of influenza pandemic in 1918 in Baltimore and localities other than Baltimore USA [8]. Mortality (deaths per population) curves are decomposed into morbidity (cases per population) and the case fatality ratio (deaths per cases). Dashed-line and dotted-line represent the male and female population respectively

in the 60 and 70 years age-groups.

Public health reports published by Collins [4] many years later in 1931 report the number of cases and deaths of certain respiratory diseases occurring within four month period since 1 September 1918 in certain surveyed localities were reported. Those localities are New London, Conn, Baltimore and five minor Maryland towns,Spartanburg, Augusta, Macon, Angusta, Luisville, Des Moines, Iowa, Litte Rock, San Antonio Tex and San Antonio Calif. Figure 2.3 presents this data in terms of mortality, morbidity and the case fatality ratio, stratified by gender and age (5-year age-groups) and also shows the W-shaped pattern for mortality and the case fatality ratio.

Japan

The datasets are extracted from historical public health reports giving the number of cases and deaths from influenza in the whole of Japan [5], stratified per 10-year age-groups. The denominator is the total population of Japan as of the beginning of 1918. Here again, mortality (deaths per population) is decomposed into morbidity (cases per population) and the case fatality ratio (deaths per cases). Panels G,H,I
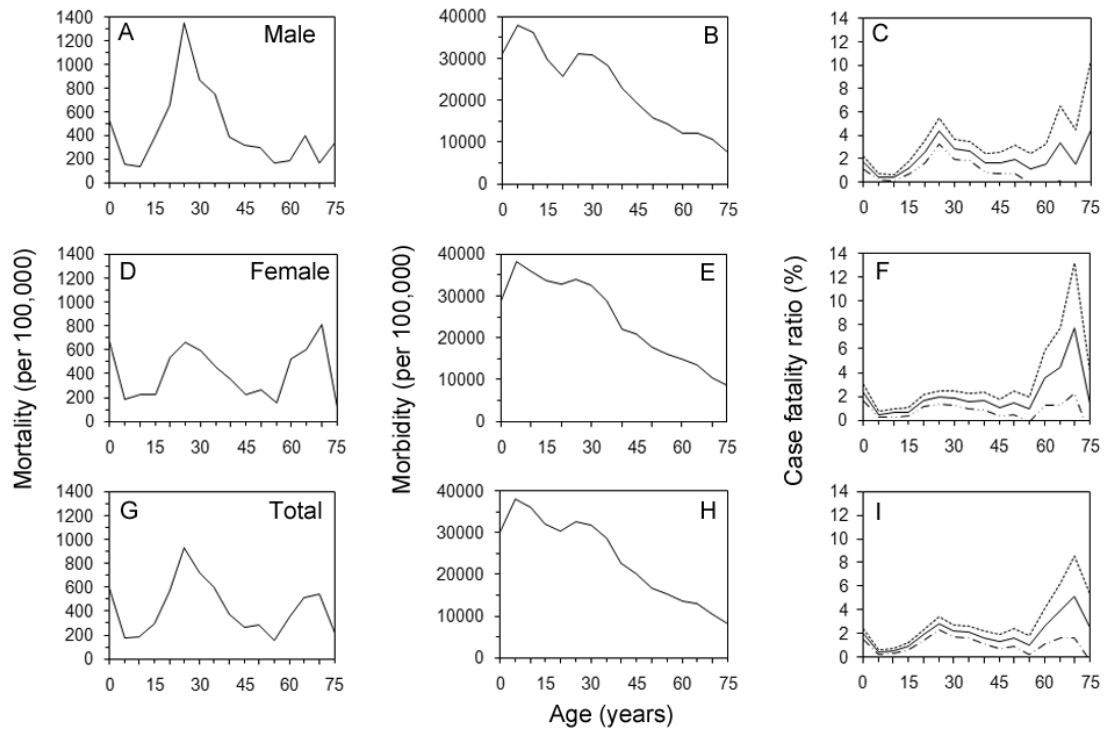
Figure 2.3: Age-specific epidemiological features of influenza pandemic in 1918 of 11 surveyed localities in the United States of America [4]. Mortality, morbidity and the case fatality ratio curves are stratified by age (5-year groups) for male (panel A, B and C), female (panel D, E and F), and the total population (panel G, H and I). Dotted-lines in the case fatality curves represent the 95% confidence band

in Figure 2.1 show the mortality, morbidity and the case fatality ratio of influenza during the 1918 pandemic in Japan. The additional peak among young adults (20-40 years age-group) is seen not only in the age-specific mortality, but also in the case fatality ratio, resulting in the W-shape.

Iceland

Number of cases and deaths from influenza observed by a practicing physician in Reykjavik during the Spanish flu epidemic were recorded and reported by Gottfredsson in 2008 [10]. We extracted the percentage of age-specific mortality from that published public health report and compared it with the population data retrieved from the Iceland statistics database (http://www.www.statice.is) and further analyzed this to produce the age-specific mortality curve per 100,000 population (fig 2.4). The age-specific mortality curve shows the W-shape pattern where the additional peak of infections and deaths from influenza occurs among the young adult
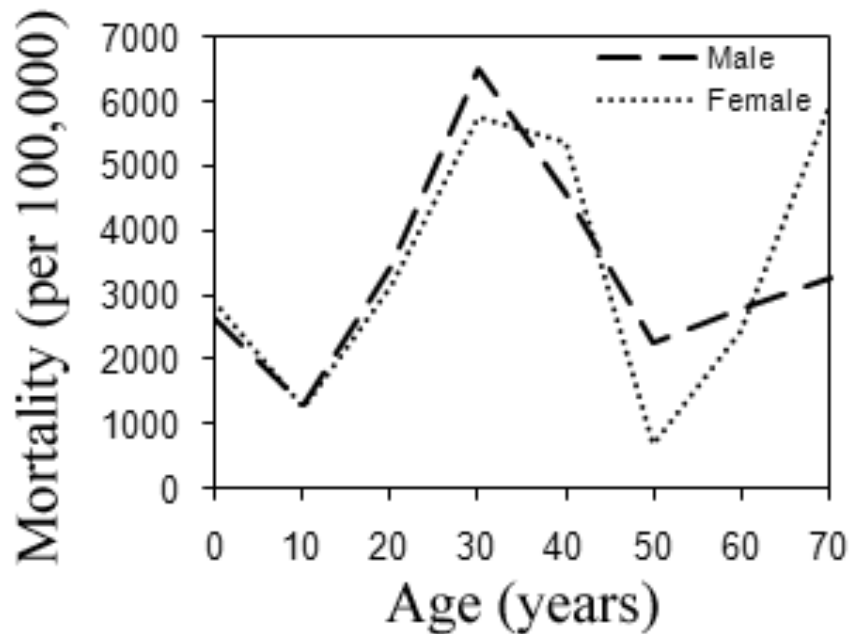
Figure 2.4: Age-specific mortality of influenza during the 1918 pandemic in Reykjavik, Iceland [10]. Dashed-line and dotted-line represent male and female population respectively. Percentage of mortality was extracted and compared with population data retrieved from the Iceland statistics database (http://www.www.statice.is). The mortality curves for both males and females show the 'W-shaped' pattern with an additional peak of deaths among the 20-40 years age-groups.

population (age 20-40 years old). However, the number of patients reported in the report is not representative of the whole Reykjavik population and the representative number of cases could not be retrieved from the Iceland statistics database, thus we can only present the data collected in terms of absolute number of deaths and cases. The case fatality ratio from these numbers, stratified by age and gender, consistently shows the W-shaped pattern (fig 2.5).

The Netherlands

Datasets for the Netherlands were collected from the published annual reports by the Netherlands' Central Bureau of Statistics [3]. The number of deaths per year specifically caused by influenza and pneumonia, stratified by age and gender, were collected from the year 1901 until 1940. The denominator is the total population of the Netherlands (stratified by gender and age) as recorded by the population survey every 10 years. Figure 2.6 depicts the mortality from the influenza pandemic in 1918/19 in the Netherlands in comparison with the USA, presenting the magnitude
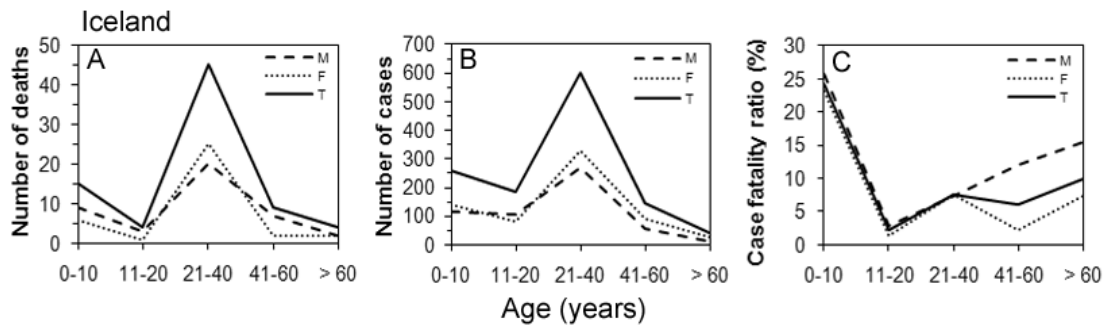
Figure 2.5: Age-distribution of deaths, cases and the case fatality ratio of influenza during the pandemic in 1918 in Iceland [10]. Dashed, dotted, and bold lines represent female, male and the total observed individuals. The 'W-shaped' age-profile, i.e. an additional peak of death among young adults aged 20-40 year-groups, is seen in the case fatality ratio for females and the total population.

of the influenza mortality compared with seasonal influenza outbreaks from other years. Figure 2.7 shows only the mortality (deaths per population), stratified by age and gender, because there were no records of influenza cases during that period from which to derive the morbidity and the case fatality ratio. In this figure, again, the mortality curve has the W-shaped pattern.
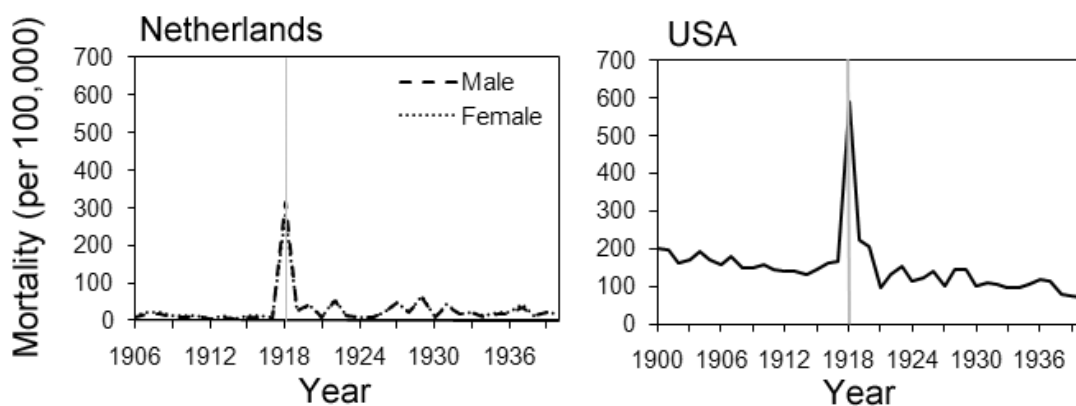
Figure 2.6: Influenza mortality (per 100,000 of population) during the early of 20th century in the Netherlands (for males and females) and the United States of America (for both genders). The vertical lines represent the year 1918 for both panels. The high increase of deaths of influenza and pneumonia during the influenza pandemic in 1918 shows the magnitude of the pandemic compared with other influenza epidemics.
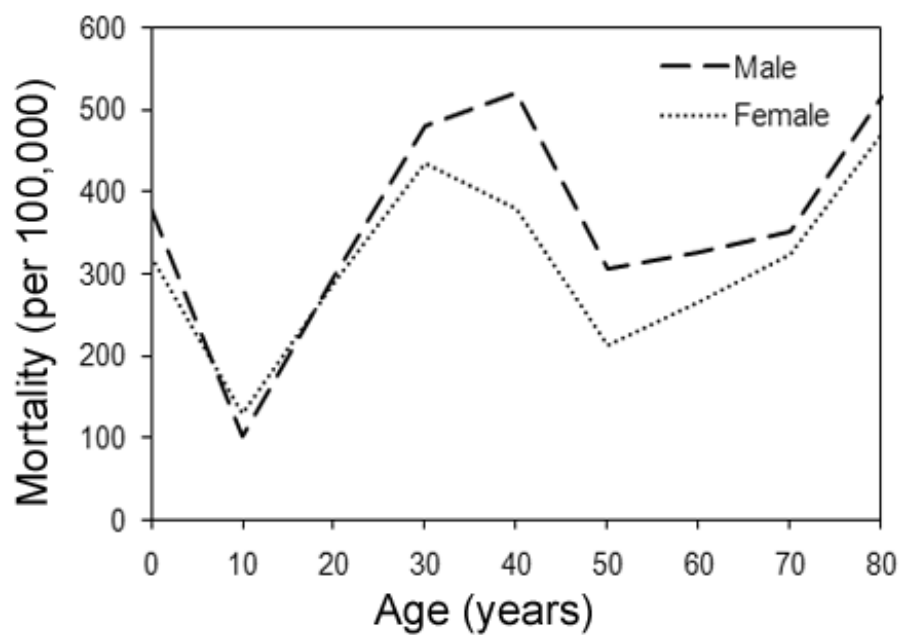
Figure 2.7: Age-specific mortality (per 100,000 population) during influenza pandemic in 1918 in the Netherlands. Dashed- and dotted-lines represent males and the females. The figure shows an additional peak of deaths among the young adults (aged 30-40 year-groups) for both male and female populations, resulting in the 'W-shaped' profile.

2.3.2   Causal inference: tuberculosis *vs* influenza

We wanted to investigate the influence of TB infection on increasing the risk of influenza death given influenza infection. We will analyze that tuberculosis increased not only the risk of contracting influenza infection, but also the risk of dying of influenza given infection. The increasing risk of influenza seen among the TB patients is analyzed from several datasets of public health report in Maryland USA [26], a TB sanatorium in Switzerland [11], and a TB hospital in Japan [13]. A study conducted elsewhere on a Union Army data in the late 19th century confirmed that TB increased the risk of influenza mortality [24]. Furthermore, we will use the data of the Swiss sanatorium [11] to establish the association that TB increases the risk of dying from influenza given infection.

To see if TB infection influenced the risk of contracting influenza during the flu pandemic in 1918, the TB patients of a hospital in Japan, TB-infected and non-TB infected individuals of a sanatorium in Switzerland, and some sections of the population in Maryland, USA, are analyzed. In the public health report from Maryland, 2375 TB infected individuals are compared with 8820 non-TB individuals where 2566 of the total number contracted the influenza infection (tab. 2.1). The odds ratio of 1.16 (1.08-1.30, p-value 0.006) means that the risk of contracting the influenza infection is 16% higher among the TB population compared with the non-TB population. The odds ratio and 95% confidence interval are calculated by formulas 2.1 and 2.2. We calculated the chi square test of this 2x2 table, given by formula 2.3, that gives a value of 7.75 and a p-value of 0.005. The Yates-corrected chi square (eq. 2.4) gives a value of 7.60 and a p-value of 0.006.

$$\text{Odds ratio} = \frac{(595 \times 6849)}{(1971 \times 1780)} \tag{2.1}$$

Table 2.1: Tuberculosis infection vs influenza infection in Maryland during the influenza pandemic in 1918 [26]. An odds ratio (OR) of 1.16 is calculated from the table by formula (2.1) means that the the risk of contracting the influenza infection is 16% higher among those TB population compared with those non-TB population. The 95% confidence interval (1.09-1.30) is calculated by formula (2.2), with the accompanying p-value of 0.006.

		Influenza infection	
		Yes	No
Tuberculosis infection	Yes	595	1780
	No	1971	6849

$$95\% \text{ Confidence Interval} = e^{\ln(\text{OR}) \pm 1.96 \times \sqrt{\frac{1}{1971} + \frac{1}{6849} + \frac{1}{595+1971} + \frac{1}{1780+6849}}} \qquad (2.2)$$

where OR is the odds ratio from equation 2.1.

In other reports analyzed, the TB infection seemed to give a protective effect from contracting influenza. The data from the Swiss sanatorium covers 102 patients with TB and 33 employees without the TB infection (tab. 2.2). In this sanatorium population, 64 TB patients and 24 non-TB employees were infected with influenza. An odds ratio of 0.63 (0.27-1.5, p-value 0.4) means that the risk of contracting the influenza infection among the TB patients is 37% smaller compared with those who did not have the TB infection. The chi square test, using the same method of calculation as given in equation 2.3, is 1.09 with p-value of 0.30.

$$\chi^2 = \frac{\left((595 \times 6849) - (1971 \times 1780)\right)^2 \times (595 + 1780 + 1971 + 6849)}{(595 + 1971) \times (1780 + 6849) \times (595 + 1780) \times (1971 + 6849)} \qquad (2.3)$$

$$\text{Corrected } \chi^2 =$$
$$\left[\frac{\left(|(595 \times 6849) - (1971 \times 1780)| - (0.5 \times (595 + 1780 + 1971 + 6849))\right)^2}{(595 + 1971) \times (1780 + 6849) \times (595 + 1780) \times (1971 + 6849)}\right]$$
$$\times (595 + 1780 + 1971 + 6849) \qquad (2.4)$$

The data from the TB hospital in Japan shows the different effects of several clinical courses of TB infection in contacting the influenza infection. A number of 57 TB patients are analyzed according to the changes in their clinical status (worsened or unchanged) and category (<3 or ≥ 3) of weight, temperature, coughing, sputum,

Table 2.2: 2x2 table of tuberculosis infection vs influenza infection in a TB sanatorium in Switzerland [11]. An odds ratio (OR) of 0.63 is calculated with the 95% confidence interval of 0.27-1.5 and an accompanying p-value of 0.4. The chi square calculated from the table is 1.09 with a p-value of 0.30.

		Influenza infection	
		Yes	No
Tuberculosis infection	Yes	64	38
	No	24	9

appetite, smear positive, and culture of the TB bacteria. Table 2.3 summarizes each
association and shows that among these patients characteristics, a worsened status in
TB-progressive patients increased the risk of contracting influenza by 2.86 times (OR
2.86, 95%CI:0.73-11.14, p-value 0.14). TB patients in the category <3 for appetite
and positive culture slightly increased risk of contracting the influenza infection
(OR 1.09, 95%CI:0.19-6.06, p-value 1.00; and OR 1.05, 95%CI:0.30-3.74, p-value
1.00 respectively). The other conditions appear to reduce the risk of contracting
influenza infection. Chi square tests were calculated with their accompanying p-
values, given in the table. These statistically non significant associations, which
may be due to the small sample size of the analysis, could have led scientists at that
time to think that having a chronic disease such as TB might protect people from
getting influenza. This inappropriate interpretation will be addressed more carefully
in the discussion.

A study on Union Army veteran data in the USA in the late 19th century was
conducted by Noymer in 2009 to see the association between having tuberculosis and
influenza mortality [24]. In that study, the number of soldiers who were registered
as having the TB infection during the influenza epidemic is analyzed. The odds
ratio in that paper shows that the risk of dying from influenza is 0.88 (0.52-1.52,
p-value 0.8) among the tuberculous compared with the non-tuberculous. This result
was elaborated further using Cox regression to show the effect of having TB on the
hazard of influenza death. This gave a 20% higher hazard ratio (analogous to the
odds ratio), although this is not statistically significant, given the sample size.

Compared to the study mentioned earlier, our study analyzed that TB is a risk
factor for influenza death given infection, supporting the previous findings of 'W-
shaped' in the age-specific case fatality ratio. Using the same data from the Swiss
TB sanatorium [11], we analyze the risk of influenza death, given infection, among
the TB patients compared with non-TB individuals. Table 2.4 shows the TB pa-
tients infected with influenza compared with employees infected only with influenza,
with death and survival as the outcomes. Among the 24 non-TB employees, none
died, whereas among the 64 patients infected with TB, 7 died of influenza. Due to
the smallest expected sample in one of table cells (less than 5), the risk of dying
from influenza given infection among the TB patients compared with non-TB em-
ployees was analyzed by Fisher Exact test (eq. 2.5), in preference to the χ^2 test with
continuity correction [2]. The probability of 0.09 was calculated conditional on the
observed marginal totals, under the null hypothesis that there was no association

between the tuberculosis and influenza deaths. This exact test has an undesirable feature that the average value of the significance level, when the null hypothesis is true, exceeds 0.5.

$$\text{Fisher's exact test} = \frac{(7+0)! \times (57+24)! \times (7+57)! \times (0+24)!}{7! \times 0! \times 57! \times 24! \times (7+0+57+24)!} \quad (2.5)$$

2.3.3   Separating the effect of age, period and cohort

We explicitly analyzed the epidemiological time course of TB mortality using a statistical model that separately quantifies the three temporal effects of age, period and cohort. Temporal patterns such as age, period and cohort are important aspects of disease that need to be taken into account separately in epidemiological analysis to get a true association with each component. Age, an important risk factor in most diseases, must be controlled in any analysis of temporal trends because older age groups might be at a relatively higher risk of certain diseases or conditions. Period, the effect associated with calendar time, might identify a change in the population exposure to an important risk factor that causes sudden increases or decreases in disease rates at a certain time. Period is the time when the disease occurred that is sometimes affected by artifacts that change the estimates of disease rates, but do not affect the true underlying disease burden. For example, an increase awareness of a certain disease can cause higher reporting of a disease's occurrence or increased knowledge of disease can change the case definition or diagnosis code. Such a change can influence the period change because all age groups may be similarly affected. Another analytic perspective arises from looking at the experience of the same generation of individuals (known as the cohort effect), often referred to by their year of birth, as they are less prone to the dynamic population change over time. However, an inherent redundancy develops from these three perspectives, i.e. where period = age + cohort, that needs to be taken into account in the analysis.

We used datasets of TB mortality from various sources to look into those age-, period-, and cohort-effects and put them into a table for further analysis. We stratified the annual TB mortality data by gender (except for the USA) and age (10-year age-groups for the USA and the Netherlands, 5-year age-groups for Japan). Figure 2.8 summarizes the TB mortality from the USA, Japan and the Netherlands across different age categories (children, young adults, and the elderly). The missing lines in the middle part of the mortality curve in panels C, F, H, and I were due
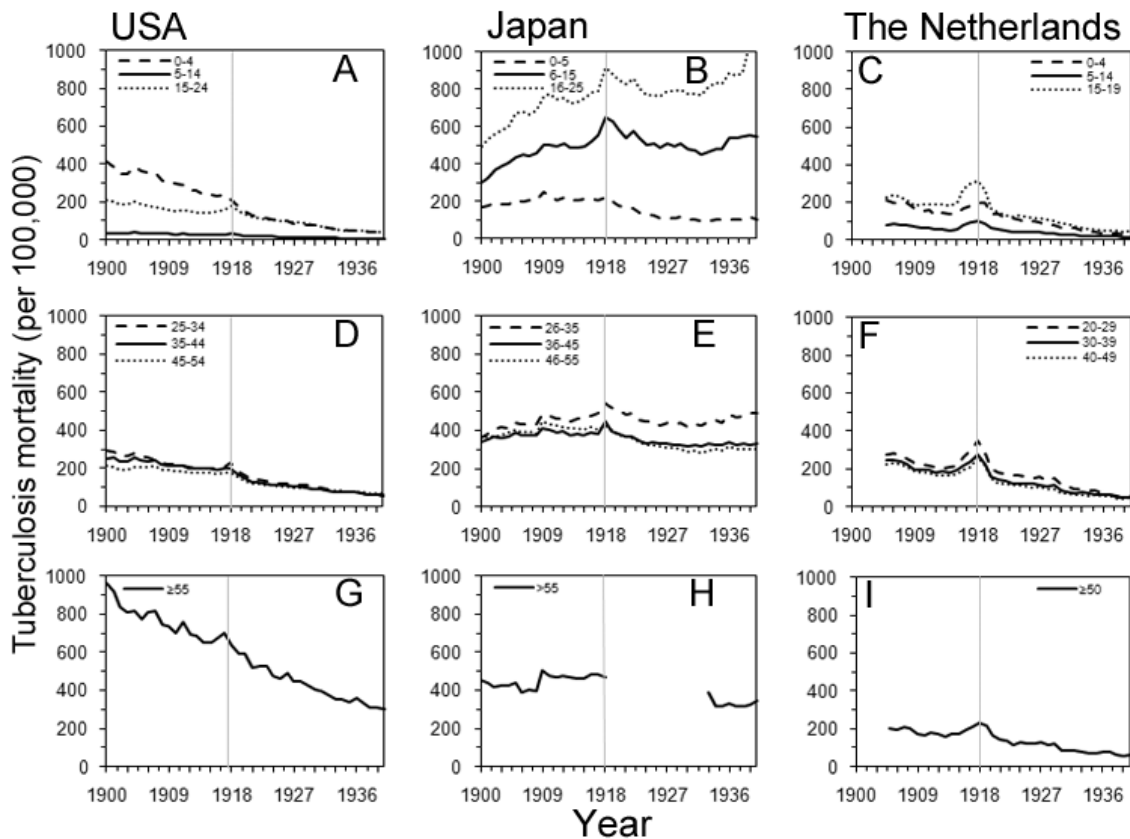
Figure 2.8: Tuberculosis mortality (per 100,000 of population) across different age groups in the United States of America (panels A,D,G), Japan (panels B,E,H), and the Netherlands (panels C,F,I) during the period 1900-1940. Panels A,B,C show children, panels D,E,F show young adults, and panels G,H,I show the elderly for the three countries. The vertical solid lines in each panel represent the year 1918.

to the limited published historical information of TB mortality in Japan and the Netherlands.

Table 2.5 shows an example of how to extract the three components of age, period and cohort from the mortality data (per 100,000 population) in Japan. The table depicts a part of the age-specific TB mortality in Japan from the period 1899 to 1904 with the age group stratified by 5 years (from ≤5 year to >60 year of age groups), with the inherent cohort groups. For example, the TB mortality of individuals aged > 60 years in the year 1899 is 151.1 per 100,000 of total Japan population, and these people are in the birth cohort group of 1834-1838.

We modeled the data from the table and we estimated the parameters by means of maximum likelihood using the JMP Statistics analytical package version 7.0.1. Relative risk is interpreted as the exponential form of the estimated parameters. We

performed the same process for all the TB age-specific mortality data from the USA, Japan, and the Netherlands, with some smaller datasets from elsewhere [18, 15] for the exercises.

During the exercises, we extracted and re-conducted the analysis of age-, period-, and cohort-effect described in the published papers analyzing the effects of age, period, and cohort [15, 18]. In those papers, two datasets comprising 4 age, 4 period, 7 cohort, and 7 age, 5 period, and 11 cohort categories respectively are analyzed using the Poisson regression model. Several models were built to see the effect of age alone (A) (eq. 2.6), period alone (P) (eq. 2.7), age-period (AP) (eq. 2.8), period-cohort (PC) (eq. 2.9) and the age-period-cohort (APC) model (eq. 2.10). We derived the additive model from the multiplicative model by taking the log-linear form (eq. 2.11), where μ, $\alpha_{(i)}, \beta_{(j)}, \gamma_{(k)}$ represent the natural log of K, $A_{(i)}, P_{(j)}$, and $C_{(k)}$, respectively.

$$\lambda_{(i)} = K \times A_{(i)} \tag{2.6}$$

$$\lambda_{(j)} = K \times P_{(i)} \tag{2.7}$$

$$\lambda_{(i,j)} = K \times A_{(i)} \times P_{(j)} \tag{2.8}$$

$$\lambda_{(j,k)} = K \times P_{(j)} \times C_{(k)} \tag{2.9}$$

$$\lambda_{(i,j,k)} = K \times A_{(i)} \times P_{(j)} \times C_{(k)} \tag{2.10}$$

$$\ln \lambda_{(i,j,k)} = \mu + \alpha_{(i)} + \beta_{(j)} + \gamma_{(k)} \tag{2.11}$$

Figure 2.9 shows a visual comparison fitting the AP and APC model to the mortality data in the 15-24 years of age-group, which shows that the APC is a better fitted model than AP. Figures 2.10 and 2.11 show the relative risks for each age, period and cohort category using the APC model for both datasets.

We fitted both multiplicative and additive models of each age model, period model, age-period model, and age-period-cohort model to the datasets and found that there is no difference in the results from the multiplicative or additive models. Hence, we
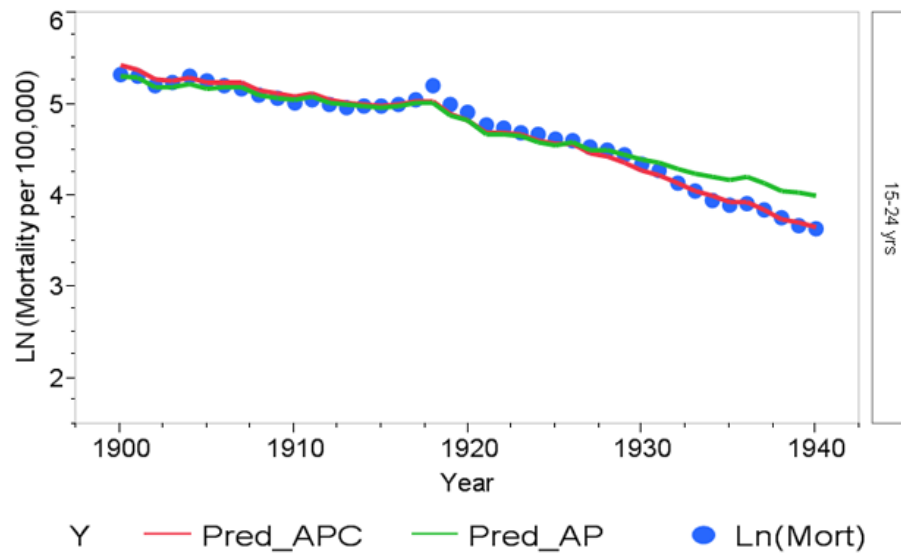
Figure 2.9: Visual comparison of AP and APC models in explaining the mortality data per 100,000 population (in logarithmic scale) in the 15-24 years of age-group in the United States of America during the period 1900-1940 [34].
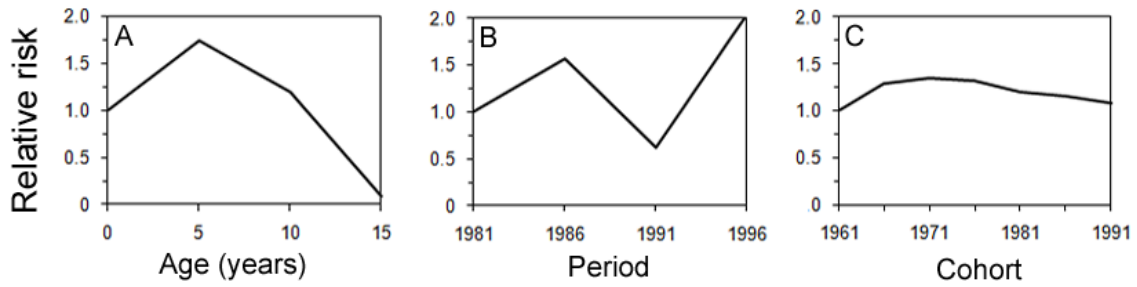
Figure 2.10: Relative risk of dengue disease in each age (panel A), period (panel B), and cohort groups (panel C) using the APC model [18]. The precisions are per 5 years for all groups.

used additive model only when analyzing TB mortality data from the USA, Japan and the Netherlands.

To analyze TB mortality data from the USA [34], Japan [16, 30], and the Netherlands [3], we built the same models that take into account the effect of age only (A), period only (P), age and period (AP), and age-period-cohort altogether (APC) (eq.2.11). $Ln\lambda_{(i,j,k)}$ represents the unknown true mortality rate, which was predicted by age, period, and cohort categories. μ is the natural log of the constant K. $\alpha_{(i)}$, $\beta_{(j)}$, and $\gamma_{(k)}$ are parameters representing the effect of age, period, and cohort. Relative risk is interpreted as the exponential form of these parameters.

The age groups of 5-14 years, 0-4 years, and younger than 20 years, are the reference age groups for the USA, Japan and the Netherlands datasets respectively. These different age-groupings are due to the limited consistency of the historical data across different countries. The period groups of 1900, 1899 and 1905, and the birth cohort
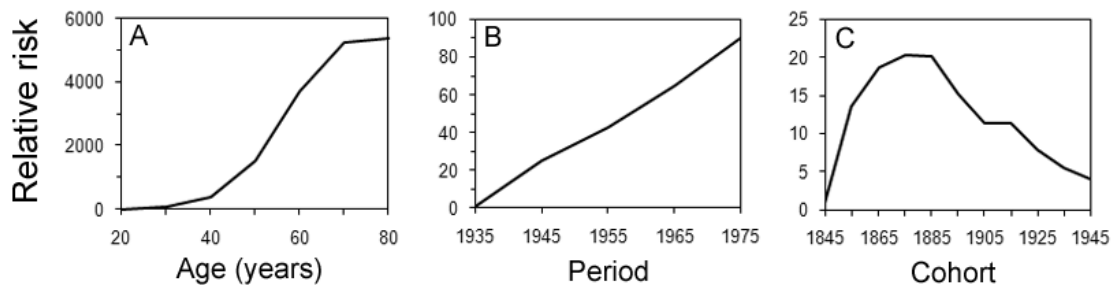
Figure 2.11: Relative risk of lung carcinoma mortality in each age, period, and cohort group using the APC model [15]. The precisions for age (panel A), period (panel B) and cohort (panel C) groups are per 5 years.

groups of 1801-1810, 1831-1835, and 1831-1840 are the reference groups for the three
countries respectively. We could not quantify the cohort effect straightforwardly
from the data since the intervals of age (per 5-year for Japan and 10-year for USA
& the Netherlands) differ from the period category (per 1-year). This discrepancy
was dealt with by creating parameters of cohort grouping based on 5-year (for Japan)
and 10-year groups (for the USA and the Netherlands). An example of the analysis
of TB mortality in the USA datasets is given below.

The USA TB mortality data has 9 age categories at 10-year intervals, running from
5-14 to > 85, 41 period categories (annually from the year 1900 to 1940), and 14
cohort groups at 10-year intervals, starting with 1801-1810 as cohort group 1, until
1931-1940 as cohort group 14. Applying the additive APC model (eq. 2.11) for the
data point of age category 1 (5-14 age-group), period category 1 (year 1900), and
cohort category 1 (1886-1895), gives the equation 2.12. In this equation, the effects
of age category 1 and period category 1 are already included in the constant because
they serve as the reference groups. Part of the cohort category 1 lies in the cohort
group 9 (1886 to 1890) and the other part in cohort group 10 (1891-1895). This
cohort grouping model calculation is applied to all TB mortality datasets.

$$\ln \lambda_{(1,1,1)} = \mu + \frac{1}{2} \times C_9 + \frac{1}{2} \times C_{10} \qquad (2.12)$$

where C_9 and C_{10} are cohort groups 9 and 10.

Parameters were estimated by maximizing the negative log-likelihood of the Poisson distribution (eq. 2.13) using the JMP statistical package (version 7.0.1 2007),
presented with their lower and upper confidence limits. Equation 2.14 shows the
formula of the negative log-likelihood. We assessed the goodness-of-fit of the fitted
models using Akaike's Information Criterion (AIC), given by equation 2.16. AIC
is not a measure of how good a model fits the data, rather, it is a measure of how
good the model fits to the data as compared with the other models that are fitted.
The best model is the one that gives the smallest AIC, unless the difference between
AIC of models is small (≤ 2) then the simplest model is preferred (Occam's Razor
principle).

$$f(x; \lambda) \quad = \quad \frac{\lambda^x \times e^{-\lambda}}{x!} \qquad (2.13)$$

$$\text{NLL} = \lambda - x \times \ln\lambda + \sum_{k=0}^{x-1} ln\,(x-k) \tag{2.14}$$

$$\text{AIC} = -\,(2 \times \text{log-likelihood}) + (2 \times \text{number of parameters}) \tag{2.15}$$

$$\text{AIC} = (2 \times \text{NLL}) + (2 \times \text{number of parameters}) \tag{2.16}$$

where λ and x are predicted and observed mortalities, NLL is the negative log-likelihood.

We also performed the RMSE and the chi square test for the model prediction with the accompanying p-value. RMSE estimates the standard deviation of the residual error, which is the square root of the MSE. MSE shows the mean squared error, which is the estimate of the variance of the residual error, i.e. the SSE (residual sum of squares error, equals to chi square test) divided by the DFE (degrees of freedom error). The Chi square calculation is given by equation 2.17 and its p-value is given by equation 2.18. Significant p-value (≤ 0.05 is interpreted as the model is significantly different to the empty model (contains no parameters but only the constant).

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{(\text{Expected})} \tag{2.17}$$

$$\text{p-value} = 1 - \left[\text{chi square distribution}\left(\chi^2, \text{d.f.}\right)\right] \tag{2.18}$$

where d.f. is the degree of freedom.

As stated previously, the best fitted model explaining the TB mortality data was the one with the smallest AIC. A comparison of the AICs from the various fitted models is found in table 2.6 and table 2.7, together with their number of parameters, degrees of freedom, RMSE, and p-values.

Figure 2.12 shows the relative risk of the estimates using the age-period-cohort (APC) model. Relative risk is calculated as the exponential of the estimate from each parameter, where the risk of the reference groups is set to 1. The missing lines in panel H are due to the fact that we only included the period 1905-1930 for TB mortality in the Netherlands in the model, given the inconsistency of existing historical data before and after that period. The highest risk is found in the 15-25 age
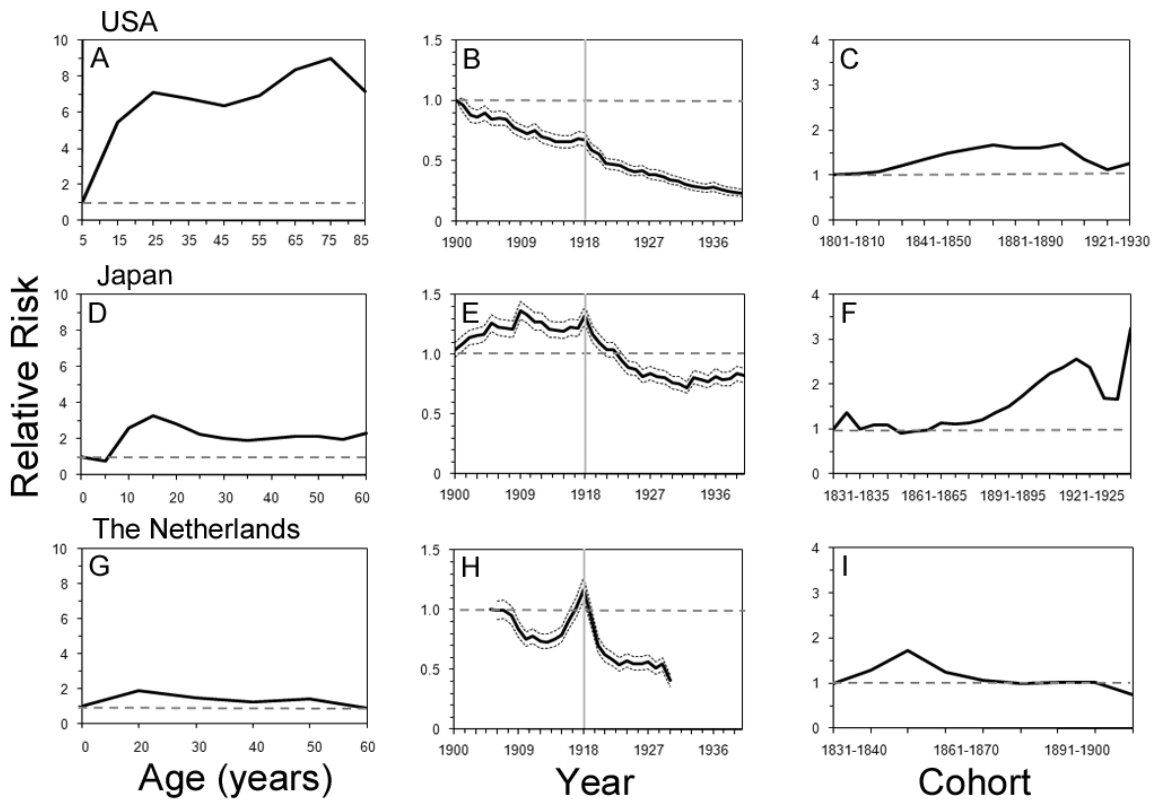
Figure 2.12: Tuberculosis mortality during period 1900-1940, analyzed using the Age-Period-Cohort (APC) model, in the United States of America, Japan, and the Netherlands. Panels A, D, G show the age effect in the United States of America, Japan, and the Netherlands respectively. Panels B, E, H show the period effect, and panels C, F, I show the cohort effect for the three countries. Age groupings were every 10 years for USA and the Netherlands, and every 5 years for Japan. The vertical solid lines and the dotted lines in panels B, E, and H represent the year 1918 and the confidence interval of period effect. The dashed horizontal lines in all panels mark out the relative risk of 1.

group for all three countries, although there is also an incline among the elderly in the USA dataset. Regarding the period effect, there is a consistent pattern of a spike in TB mortality for all three countries during 1918 that was followed immediately by a significantly steeper decline. This prominent feature of the period effect implies that TB mortality was "washed out" by the influenza pandemic.

Figure 2.13 shows the overall TB mortality in the three countries in comparison to the figure of predicted period effect on TB mortality from the APC model. This visual comparison shows that the model is reasonably good at predicting the observed TB mortality in the three countries. Observed TB mortality in this figure clearly shows that there was an increase in TB deaths in these three countries in the year 1918 and then a significantly steeper decline thereafter, suggesting that tuberculous
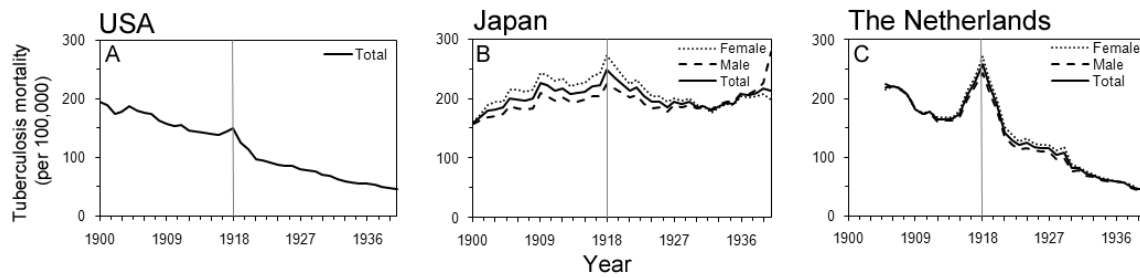
Figure 2.13: Tuberculosis mortality (per 100,000 population) of all ages during period 1900-1940 in the United States of America, Japan, and the Netherlands. The vertical solid lines and the dotted lines represent the year 1918 and the confidence interval of the period effect. It should be noted that there is a clear spike of tuberculosis mortality in 1918 which is followed by steeper decline thereafter, suggesting that tuberculosis was 'washed out' during the pandemic in 1918 for the three countries.

population was 'washed out' by the pandemic , i.e. the huge deaths resulting from the Spanish flu epidemic in 1918 was subtracted from future TB mortality.

Table 2.3: Compilation of 2x2 table of clinical courses of tuberculosis (TB) infection vs. influenza infection of a TB hospital in Japan during the influenza pandemic in 1918. The causal associations are analyzed by odds ratio and Chi square tests with their accompanying p-values. The odds ratio is interpreted as the risk of contracting the influenza among the TB patients with the presence of a certain category or clinical courses compared to those without, and is presented with the 95% confidence interval and the p-value. The tuberculosis patients with worsened status have 2.86 higher risk contracting influenza infection compared with those with an unchanged status.

Clinical course of tuberculosis		Influenza infected		OR^a (p-value)	95% CI^b (p-value)	χ^{2c} (p-value)	Corrected χ^{2d} (p-value)
		Yes	No				
Status	Worsened	5	9	2.86 (0.14)	0.73-11.14	2.40 (0.12)	1.37 (0.24)
	Unchanged	7	36				
Weight	< 3	1	9	0.36 (0.67)	0.04-3.20	0.89 (0.35)	0.27 (0.64)
	$\geq$ 3	11	36				
Temperature	< 3	1	8	0.42 (0.67)	0.05-3.74	0.64 (0.42)	0.12 (0.73)
	$\geq$ 3	11	37				
Cough	< 3	1	8	0.42 (0.67)	0.05-3.74	0.64 (0.42)	0.12 (0.73)
	$\geq$ 3	11	37				
Sputum	< 3	1	7	0.49 (1.00)	0.05-4.45	0.41 (0.52)	0.03 (0.86)
	$\geq$ 3	11	38				
Appetite	< 3	2	7	1.09 (1.00)	0.19-6.06	0.01 (0.93)	0.12 (1.00)
	$\geq$ 3	10	38				
Smear	< 3	6	25	0.80 (0.76)	0.22-2.86	0.12 (0.73)	0.00 (0.99)
	$\geq$ 3	6	20				
Culture	< 3	6	22	1.05 (1.00)	0.30-3.74	0.00 (0.95)	0.07 (1.00)
	$\geq$ 3	6	23				

[a] OR : Odds ratio
[b] 95% CI : 95% Confidence Interval
[c] χ^2: Chi square test
[d] Corrected χ^2 : Yates-corrected Chi square test

Table 2.4: 2x2 table of tuberculosis morbidity vs influenza mortality in a TB sanatorium in Switzerland. This table explains that tuberculosis increases the risk of dying of influenza, given infection, supporting the findings of the 'W-shaped' profile in the case fatality. Among 24 non-TB employees, none died whereas among 64 patients infected with TB, 7 died of influenza. Fisher's exact test, calculated by equation (2.5), results in a probability of 0.09.

	Dead	Survived
Tuberculosis and influenza	7	57
Influenza (without tuberculosis)	0	24

Fisher Exact Test; p-value 0.09

Table 2.5: Age-specific TB mortality in Japan (per 100,000 population). Data is stratified by 5-year age-groups starting from the period 1899 to 1904. The birth-cohort groups at the last columns and the last row represent the mortality situated next to it, for example, the TB mortality of individuals aged > 60 years in the year 1899 is 151.1 per 100,000 of total Japan population, and these people are the in the birth cohort group of 1834-1838.

Age (years)	Period						Cohort
	1899	1900	1901	1902	1903	1904	
≤ 5	151.5	164.1	174.9	188.4	184.0	179.9	1899-1903
6-10	76.4	82.5	92.6	102.9	107.7	109.9	1894-1898
11-15	201.5	217.0	233.6	268.4	281.9	291.9	1889-1893
16-20	251.9	265.8	285.7	304.7	317.1	326.3	1884-1888
21-25	218.8	226.6	242.2	258.1	262.7	273.1	1879-1883
26-30	190.7	190.9	201.8	213.9	217.7	216.1	1874-1878
31-35	167.8	170.5	177.6	189.6	197.0	194.8	1869-1873
36-40	161.9	168.5	179.1	184.8	176.2	178.6	1864-1868
41-45	168.6	166.4	174.9	182.3	186.9	187.7	1859-1863
46-50	174.6	183.2	184.1	187.2	189.5	189.6	1854-1858
51-55	168.7	178.0	182.6	182.1	185.6	198.8	1849-1853
56-60	157.4	167.6	158.9	162.3	168.4	167.1	1844-1848
> 60	273.4	287.9	277.1	258.9	254.0	254.8	1839-1843
Cohort	1834-1838	1835-1839	1836-1840	1837-1841	1838-1842	1839-1843	

Table 2.6: Comparison of different models of tuberculosis mortality in Japan and the United States of America, separating age, period and cohort effect. Columns represent the various models analyzing tuberculosis mortality data for two countries during the period 1900-1940, with their accompanying number of parameters and degrees of freedom. The goodness-of-fit test is assessed from the smallest AIC presented by the models.

Country	Gender	Model	Number of parameters	d.f.[a]	RMSE[b]	χ^2	p-value	AIC[c]
Japan[16, 30]	Female	A[d]	13	572	0.23	4,424	≤ 0.01	8,907
		P[e]	45	540	0.51	36,331	≤ 0.01	37,098
		AP[f]	57	528	0.19	2,767	≤ 0.01	7,270
		APC[g]	80	505	0.14	1,287	≤ 0.01	5,758
	Male	A	13	572	0.20	10,262	≤ 0.01	16,358
		P	45	540	0.49	27,927	≤ 0.01	33,924
		AP	57	528	0.19	8,233	≤ 0.01	14,471
		APC	80	505	0.12	3,955	≤ 0.01	9,845
	Female + Male	A	13	572	0.22	4,869	≤ 0.01	9,429
		P	45	540	0.44	26,495	≤ 0.01	28,800
		AP	57	528	0.20	3,624	≤ 0.01	8,207
		APC	80	505	0.14	1,572	≤ 0.01	6,081
USA[34]	Female + Male	A	9	360	0.44	6,942	≤ 0.01	9,681
		P	41	328	0.68	5,285	≤ 0.01	9,823
		AP	49	320	0.12	320	0.50	2,884
		APC	63	306	0.06	143	1.00	2,735

[a] d.f.: degrees of freedom
[b] RMSE: Square root of Mean Square Error (which is residual sum of squares error divided by degrees of freedom)
[c] AIC: Akaike's Information Criterion
[d] A: age-model
[e] P: period-model
[f] AP: age-period model
[g] APC: age-period-cohort model

Table 2.7: Comparison of different models for tuberculosis mortality in the Netherlands, separating age, period and cohort effect. Columns represent various models analyzing tuberculosis mortality data for the country during the period 1905-1930, with their accompanying number of parameters and degrees of freedom. The goodness-of-fit test is assessed from the smallest AIC presented by the models.

Country	Gender	Model	Number of parameters	d.f.[a]	RMSE[b]	χ^2	p-value	AIC[c]
The Netherlands[3]	Female	A[d]	6	150	0.31	2,182	≤ 0.01	3,345
		P[e]	26	130	0.31	1,716	≤ 0.01	2,984
		AP[f]	31	125	0.19	609	≤ 0.01	1,748
		APC[g]	40	116	0.13	302	≤ 0.01	1,467
	Male	A	6	150	0.36	2,860	≤ 0.01	4,095
		P	26	130	0.35	2,733	≤ 0.01	3,860
		AP	31	125	0.20	730	≤ 0.01	1,860
		APC	40	116	0.14	353	≤ 0.01	1,500
	Female + Male	A	6	150	0.33	2,492	≤ 0.01	3,685
		P	26	130	0.33	2,056	≤ 0.01	3,292
		AP	31	125	0.20	629	≤ 0.01	1,763
		APC	40	116	0.13	304	≤ 0.01	1,465

[a] d.f.: degrees of freedom
[b] RMSE: Square root of Mean Square Error (which is residual sum of squares error divided by degrees of freedom)
[c] AIC: Akaike's Information Criterion
[d] A: age-model
[e] P: period-model
[f] AP: age-period model
[g] APC: age-period-cohort model

Chapter 3

Results

This chapter summarizes the findings from the previous methodology section by presenting the results obtained from the three systematic steps of analysis.

3.1 The W-shaped age-distribution of the influenza pandemic

The W-shaped age-specific mortality curve is seen consistently in all the influenza datasets from the United States of America (fig. 2.2 and fig. 2.3), Japan (fig. 2.1), Iceland (fig. 2.4 and fig. 2.5), and the Netherlands (fig. 2.7) during the pandemic in 1918. This W-shape was also seen for the case fatality ratio when we decomposed the mortality curve into morbidity and the case fatality ratio curves. This atypical feature clearly distinguishes the Spanish flu pandemic in 1918 from other influenza pandemics, as well as seasonal epidemics, that disproportionately killed the very young and the elderly, resulting in a 'U-shaped' curve (fig. **??**). An additional peak of death given infection among the 20-40 years age group indicates that there was an underlying risk factor that elevated the risk of influenza death among young adults. TB infection could plausibly be the co-morbidity that explains this age-profile mortality peak since it was a common disease in young adults during the early part of the 20$^{\text{th}}$ century, and most commonly it infected the same site, i.e. the respiratory tract. Several other arguments supporting this hypothesis will be explained further in the discussion.

3.2   Tuberculosis *vs* influenza: An increased risk

The test for a causal association between the presence of TB infection and an increased risk of influenza death among the influenza-affected population that was performed in separate datasets, resulted in a positive relation. The risk of contracting the influenza infection is 16% higher among those TB population compared with those non-TB population in Maryland, USA (tab. 2.1), with 95% confidence interval 1.08-1.30 and p-value 0.006. The Chi square test and the corrected-one give the value of 7.75 and 7.60, and p-value of 0.005 and 0.006, implying that there is a significantly higher risk of contracting influenza among TB-infected compared with non-TB-infected individuals.

Whereas, similar causal inference from the Swiss datasets (tab. 2.2) results in an odds ratio of 0.63 (0.27-1.5, p-value 0.4). The chi square test, using the same method as given in equation 2.3, is 1.09 with p-value of 0.30. The clinical courses of TB patients (worsened status, category <3 for appetite and positive culture) in a hospital in Japan influenced the risk of contracting influenza (OR 2.86, 95%CI:0.73-11.14, p-value 0.14; OR 1.09, 95%CI:0.19-6.06, p-value 1.00; and OR 1.05, 95%CI:0.30-3.74, p-value 1.00, respectively) (tab. 2.3). Other clinical courses appear to have a protective effect (odds ratio ≤ 1) against contracting influenza infection. However, all of these clinical courses give non-significant p-values (≥ 0.05). Chi square tests were calculated with their accompanying non-significant p-values, presented in table 2.3. These results contradict our hypothesis. They may be due to the small sample size of the analysis, could have led scientists at that time to think that having a chronic disease such as TB might protect people from getting affected by the influenza pandemic. This inappropriate interpretation will be addressed more carefully in the discussion.

Another study conducted elsewhere on Union Army veteran data in the USA in the late 19th century to see the association between having tuberculosis and influenza mortality [24]. The odds ratio in that paper shows that the risk of dying from influenza is 0.88 (0.52-1.52, p-value 0.8) among the tuberculous compared with the non-tuberculous. This result was elaborated further using Cox regression to show the effect of having TB on the hazard of influenza death. This gave a 20% higher hazard ratio (analogous to the odds ratio), although this is not statistically significant, given the sample size.

Compared to the study mentioned earlier, our study analyzed that TB is a risk factor for influenza death given infection, supporting the previous findings of 'W-shaped' in the age-specific case fatality ratio. The risk of dying from influenza given infection among the TB-infected compared with the non-TB-infected resulted in a probability of 0.09. The Fisher Exact test was used to analyze this association due to the small sample size in one of table cells, under the null hypothesis that there was no association between TB infection and influenza death. Nevertheless, table 2.4 clearly shows that out of all the non-TB-infected individuals, none died from influenza, whereas out of 64 patients infected with TB, 7 died of influenza. This strongly suggests that TB was indeed the co-morbidity that increased the risk of dying among the influenza-affected population in the sanatorium.

3.3 Temporal trends: Age, period and cohort

The age-period-model (APC) is the best fitted model for explaining the TB mortality for all three countries (the USA, Japan ,and the Netherlands), compared with models of age alone, period alone, and age-period effect. The APC model give the smallest AIC values (2,735; 6,081; and 1,465 for the total populations of USA, Japan and the Netherlands, respectively) (tab. 2.7). This AIC is calculated by equation 2.16 and the negative log-likelihood for the APC additive model is given by equation 2.14. Previous APC model exercise on Dengue and lung cancer datasets showed that there was no difference in the results from applying an additive or multiplicative model and that the APC model is the best fitted model for explaining both datasets.

The parameters of the APC model are estimated by means of maximum likelihood (using JMP statistics analytical package version 7.0.1) and relative risk is interpreted as the exponential form of each estimated parameter. Figure 2.12 shows the relative risk from APC model for the various age, period and cohort groups, which is comparable to the observed TB mortality from the three countries in figure 2.13. The relative risk is the exponential of the estimates given by the additive APC prediction model (eq. 2.11), with the risk of the reference groups set to 1. The highest risk is found in the 15-25 age group for all three countries, although there is also an incline among the elderly in the USA dataset. Regarding the period effect, there is a consistent pattern of a spike in TB mortality for all three countries during 1918 that was followed immediately by significantly steeper decline. This prominent fea-

ture of period effect implies that TB mortality was "washed out" by the influenza pandemic.

Chapter 4

Discussion

The W-shaped feature is specific for the influenza pandemic in 1918 (it appears in mortality and the case fatality ratio, but only in mortality of the influenza pandemic 2009); Therefore, the underlying reason should also be specific for the Spanish flu.

4.1 Unique study: Did Spanish flu kill many TB patients?

A young adult disease

TB could have plausibly been the underlying cause for several reasons. First of all, the age-profile of the tuberculous population in the early 20th century (i.e. it was more commonly found in young adults) overlaps with the W-shaped feature of the Spanish flu. In other words, the huge loss of life among young adults resulting from the Spanish flu was in fact subtracted from future TB mortality. This is often referred to as the selection effect, or the 'cohort inversion' effect, i.e. a cohort experiencing an adverse event in its lifetime may be more robust after the event than before [14]. Thus, an adverse event in early life, in this case TB infection, can enhance the mortality risk from another disease, in this case influenza.

Affecting the same site as influenza

Tuberculosis is a chronic infectious disease that predominantly infects the same site

as influenza, i.e. the respiratory tract. Some studies have revealed that having an influenza infection aggravates the pulmonary condition of TB patients, so that a closed case may become open, an arrested lesion active, or an active case progressive [31]. The diagnosis of TB was presumably established in the early of 20th century, but based on the apparent clinical signs and symptoms of the disease because diagnostic aids were not yet commonly applied [19]. Post mortem autopsies, which were not generally used at that time, revealed that the exact cause of death often was not identified in cases of TB. This was mostly in the elderly, for whom death would be perceived to be more natural and acceptable. Even if the procedure had revealed TB was the cause of death, the families of the deceased may have requested the doctors not to declare it as such, due to the social stigma and consequences for health insurance. On the contrary, a diagnosis of TB frequently follows in the wake of influenza, where physicians do err at times on the side of caution in declaring patients as infected with TB especially when they have flu complications such as pneumonia [31]. This uncertainty might have caused some of the TB patients to be misclassified as non TB and vice versa, which could have lead to an underestimation or overestimation of TB mortality in the population.

Significant decrease after 1918-1919

The analysis of TB mortality data from different countries (the USA, Japan and the Netherlands) in the previous chapter of this report shows that there is a consistent pattern of a spike in TB mortality for all three countries during 1918 that was followed immediately by significantly steeper decline. This prominent feature of period effect implies that TB mortality was 'washed out' by the influenza pandemic. Potential deaths from TB infection in the near future after 1918 were subtracted all at once when the influenza pandemic occurred in 1918, referred to as 'selection effect' mentioned above. It should be noted that selection effect theory does not posit that the virulence of the Spanish flu was due to the confluence of the TB and influenza, but rather, the changes in TB epidemiology were a consequence of the influenza pandemic as suggested elsewhere by Noymer [25, 24]. Noymer distinguished this theory into active and passive selection (individual- *vs* cohort-level phenomenon). The disproportionate level of tuberculous death among young adults during the influenza pandemic in 1918 (resulting in the W-shape) describes perfectly passive selection. However, one might argue that every increase will be followed by a decrease and visual assessment on the TB mortality graphs is not an objective method nor scientifically valid.

Causal association: Tuberculosis *vs* influenza

Causal link that TB is a risk factor for influenza death (referred to active selection as mentioned above) has been studied twice previously and presenting more causal association analysis would support this hypothesis. TB not only increases the risk of influenza mortality, as shown by other study [24], but also is an underlying factor that increased the risk of influenza death, given infection (the influenza case fatality ratio), as shown from our study. We analyzed some individual-level TB *vs* influenza data obtained from published historical data and we found evidence suggesting that TB was indeed the co-morbidity that increased the risk of dying within the influenza-infected population. The sample population for making the causal inference between TB infection and influenza death was the influenza infected population. It was noted that TB patients were more likely to contract influenza compared with healthy sanatorium employees. This raised the question of whether or not the increasing risk of flu infection and death from flu was influenced by the presence of TB infection or by the accompanying TB pre-conditions (i.e. poorer nutrition status with lower immunity among the patients) which resulted in a higher likelihood of death among the TB patients with influenza, compared with the non-influenza-affected population. If the latter was proven to be the case, selection of the sample population would not be representative of the general population. Thus, there is a risk of incomparability because of the different risks of dying among the influenza cases and the non influenza cases, which potentially could bias the study results.

4.2 Several other hypotheses

The influence of the Great World War

Another hypothesis states that the atypical W-shaped feature in the case fatality ratio (not only mortality) could have been influenced by World War I which occurred during 1914-1918. Although we do not have explicit quantitative evidence, it is not difficult to imagine that poor nutrition and low socioeconomic status due to the war could have influenced both TB and influenza death among young adults. However, it was stated that mortality was higher among men who remained at home than those who were deployed for war [9]. We addressed this issue by examining historical

datasets in multiple countries including a "neutral" country which was not involved in the war (the Netherlands). However, the Netherlands, to some extent, was also affected by a fall in nutrition levels and socioeconomic status; This is shown in the figure 2.13 in which TB mortality increased starting in the year 1914 and declined after peaking in the year 1918. This raises the question of whether or not the peak was caused by the flu pandemic or was due to the war ending. If it was due to the war, it is unlikely that the increase in deaths driven by malnourishment would have resolved within a couple of years of the war ending. Alternatively, nutrition status could have affected those who had the disease differently from those who did not have it. There is no direct method available for measuring the effect of the war because complete historical data on the individual level (i.e. with nutrition and socio-economic status) are not available.

Another plausible co-morbidity

Another potential underlying disease that could explain the peculiar findings of W-shaped age-specific mortality during the time of Spanish flu is Scarlet fever [28, 7]. It is a clinical syndrome characterized by high fever and rash, resulting from infection and toxin of a strain of Streptococcus pyogenes which was first identified in the early 20th century. The disease can be fatal if it is complicated with sepsis, i.e. spreading of infection throughout the body, which might have been a common course of the disease during the period when antibiotic use was not yet common. However, the disease occurs infrequently before the age of 3 years and after the age of 15 years; thus, it is unlikely that Scarlet fever could have explained the higher risk of influenza death among young adults (compared with other ages). The true role of this co-morbidity needs further critical analysis, given such individual level data on influenza deaths in 1918-19 among the Scarlet fever infected and non Scarlet fever infected exist.

Acquired partial immunity

Acquired immunity among middle-aged adults has also been suggested as the mechanism underlying the W-shaped age-specific mortality. That is, H1N1 in 1918 is believed to have an intrinsically high virulence only among those who had been born before 1889, and so this potentially offered partial immunoprotection against the 1918 virus among those aged 35 or more [20, 21]. Nevertheless, it should be noted that this theory would present an additional paradox: an obscure precursor virus that left no detectable trace today would have had to have appeared and

disappeared before 1889 and then reappeared more than 3 decades later [32].

The cytokine storm

Some scientists have proposed that cytokine theory (i.e. a hyper reaction of the immune system that potentially cause severe damage and death) as an explanation for the more severe outcome observed among young adults (who have stronger immunity), in comparison with the fewer deaths seen among the very young and the elderly with weaker immune system [12, 1]. When the immune system encounters pathogens, cytokines signal immune cells and activate them, which trigger further production of cytokines. This positive feedback loop is normally kept in check by the body, but in some instances the reaction becomes uncontrolled and tissues and organs are damaged. For example, if a cytokine storm occurs in the lungs, fluids and immune cells may accumulate and eventually block off the airways, potentially resulting in death. The precise reason for this is not entirely understood but it may be caused by an exaggerated response when the immune system is exposed to a new and highly pathogenic invader.

4.3 Research implication

The influenza pandemic of 1918 is the propitious place to test the hypothesis that TB increases the risk of influenza death, and our study results suggest that this hypothesis holds. If another deadly influenza pandemic were to occur, the TB hypothesis would be useful as TB is still prevalent in some developing countries and the prevalence is highly variable worldwide [25]. Further investigation of flu casualties among TB and non-TB-infected individuals from other countries during 1918 might pose a similar challenge given the limited existence of historical epidemic records. An alternative way to test the TB-influenza selection hypothesis using individual data would be to collect data from TB-prevalent countries today [6] where influenza is still a major problem, bearing in mind that this would be complicated by the presence of HIV as a risk-enhancer for many diseases.

Chapter 5

Conclusions

The influenza pandemic in 1918 was the most devastating flu epidemic known in human history and most of the deaths occurred among the tuberculous population. Three separate analyses showed that there was an underlying co-morbidity that elevated the risk of influenza death among young adults, and the causal association between TB infection and influenza death, given infection, provided evidence that the risk of dying from influenza is higher among TB-infected individuals compared with non-TB-infected individuals. The epidemiological trends of age, period and cohort effects were consistent with TB mortality being 'washed out' during the influenza pandemic, confirming our hypothesis. It should be noted that our results do not refute other hypotheses that have been put forward to explain the W-shaped age-specific mortality pattern, but our study has presented findings in a systematic fashion suggesting that tuberculosis may have been responsible for this pattern. If another deadly influenza pandemic were to occur, testing the role of the TB hypothesis would be useful as TB is still prevalent in some developing countries and its prevalence is highly variable worldwide. An alternative way to test the TB-influenza hypothesis would be to collect individual data from TB-prevalent countries today where influenza is still a major problem, bearing in mind that it would be complicated by the presence of HIV as a risk-enhancer for many diseases.

List of Acronyms and Symbols

α natural logarithmic values of A, see eq. 2.11

β natural logarithmic values of P, see eq. 2.11

χ^2 chi square values

γ natural logarithmic values of C, see eq. 2.11

λ predicted mortality, see eq. 2.14

NLL   negative log-likelihood, see eq. 2.14 and 2.16

AIC   Akaike's Information Criterion

d.f.   degree of freedom

OR   odds ratio

μ natural logarithmic values of K, see eq. 2.11

A age variables, see eq. 2.6, 2.8, and 2.10

C cohort variables, see eq. 2.9 and 2.10

i index

j index

K parameter in Poisson regression model, see eq. 2.6, 2.7, 2.8, 2.9 and 2.10

k index

P period variables, see eq. 2.7, 2.8, and 2.10

x observed mortalities

Bibliography

[1] Abdel-Ghafar AN, Chotpitayasunondh T, Gao Z, Hayden FG, Nguyen DH, de Jong MD, Naghdaliyev A, Peiris JS, Shindo N, Soeroso S, Uyeki TM, Writing Committee of the Second World Health Organization Consultation on Clinical Aspects of Human Infection with Avian Influenza A (H5N1) Virus. Update on avian influenza a (h5n1) virus infection in humans. *N Engl J Med.*, 358:261–73, 2008.

[2] Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research.* Blackwell Science, 2002.

[3] Centre of Statistics Bureau (Centraal Bureau voor de Statistiek). Annual numbers of the Netherlands (jaarcijfers voor Nederland). 1901-1940.

[4] Collins SD. Age and sex incidence of influenza and pneumonia morbidity and mortality in the epidemic of 1928-29 with comparative data for the epidemic of 1918-19. *Public Health Rep*, 46:1909–37, 1931.

[5] Depart of Hygiene, Japanese Ministry of Interior. Influenza (ryukousei kanbou). ministry of interior, tokyo, japan. 1922.

[6] Dolin P Pathania V Raviglione MC for the WHO Global Surveillance Dye C, Scheele S and Monitoring Project. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. *JAMA*, 282(7):677–686, 1999.

[7] Ferrie JP, Troesken W. Water and chicago's mortality transition, 1850-1925. *Explorations in Economic History*, 45(1):1–16, January 2008.

[8] Frost WH, Sydenstriker E. Influenza in maryland: preliminary statistic in certain localities. *Public Health Rep*, 34:491–504, 1919.

[9] Glezen WP. Emerging infections: pandemic influenza. *Epidemiol Rev.*, 18:64–76, 1996.

[10] Gottfredsson M. Lessons from the 1918 spanish flu epidemic in iceland. pages 115–122, 2008.

[11] Guggenheim R. About influenza in tuberculosis (ueber influenza bei tuberkuloesen)(in germany). *Contribution of specific clinical tuberculosis(Beitr Klin Tuberk Spezifisch Tuberk Forsch)*, 44:237–50, 1920.

[12] Haque A, Hober D, Kasper LH. Confronting potential influenza a (h5n1) pandemic with better vaccines. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 13:1512–1518, 10 2007.

[13] Hashimoto K, Okayasu M, Nomura Y, Hayashi A, Kasai T, Nakagawa O, Saito S, Takamura T. Pulmonary tuberculosis and its complication: The influence of influenza on pulmonary tuberculosis(in japanese with english abstract). *Kekkaku (Tuberculosis)*, 37:230–242, 1962.

[14] Hobcraft J, Menken J, Preston S. Age, period, and cohort effects in demography: a review. *Population Index*, 48(1):4–43, 1982.

[15] Holford, T.R. *Analyzing the temporal effects of age, period and cohort.* Edward Arnold, 1992.

[16] Japan Anti-tuberculosis Association. Statistics for the tuberculosis in japan from 1900-92. *Japan Anti-tuberculosis Association, Tokyo, Japan.*, 1900-92.

[17] Kilbourne, E.D. Influenza pandemics of the 20th century. *Emerg Infect Dis.*, 12:9–14, Jan 2006.

[18] Kongsomboon K, Singhasivanon P, Kaewkungwal J, Nimmannitya S, Mammen Jr MP, Nisalak A, Sawanpanyalert P. Temporal trends of dengue fever/dengue hemorrhagic fever in bangkok, thailand from 1981 to 2000: an age-period-cohort analysis. *The Southeast Asian Journal of Tropical Medicine and Public Health*, 35:913–917, 12 2004.

[19] Bryder L. 'not always one and the same thing': the registration of tuberculosis deaths in britain, 1900-1950. *Soc Hist Med*, 9(2):253–265, 1996.

[20] Langford C. The age pattern of mortality in the 1918-19 influenza pandemic: an attempted explanation based on data for england and wales. *Med Hist.*, 46:1–20, 2002.

[21] Thompson WW. Luk J, Gross P. Observations on mortality during the 1918 influenza pandemic. *Clin Infect Dis.*, 33:1375–78, 2001.

[22] Cinatl J. Michaelis M, Doerr H. An influenza a h1n1/09 virus revival pandemic h1n1/09 virus. *Infection*, 37(5):381–389, Oct 2009.

[23] Murphy TJ. Postinfluenzal tuberculosis. *Boston Med Surg J.*, 181:266–70, 1919.

[24] Noymer A. Testing the influenza-tuberculosis selective mortality hypothesis with union army data. *Social Science and Med.*, 68:1599–1608, 2009.

[25] Noymer A, Garenne M. The 1918 influenza epidemic's effects on sex differentials in mortality in the united states. *Popul Dev Rev.*, 26:565–81, 2000.

[26] Pearl M. Preliminary note on the incidence of epidemic influenza asmong the actively tuberculous. pages 536–540, 1919.

[27] Presanis AM, De Angelis D, The New York City Swine Flu Investigation Team, Hagy A, Reed C, et al. . The severity of pandemic h1n1 influenza in the united states, from april to july 2009: a bayesian analysis. *PLoS Med 6(12): e1000207. doi:10.1371/journal.pmed.1000207*, 12:15–22, 6 2009.

[28] Reid A. Neonatal mortality and stillbirths in early twentieth century derbyshire, england. *Population Studies*, 55:213–232, 2001.

[29] Reid AH, Taubenberger JK, Fanning TG. Evidence of an absence: the genetic origins of the 1918 pandemic influenza virus. *Nat Rev Microbiol.*, 2:909–14, 2004.

[30] Shimao T. Tuberculosis and its control: Lessons from the past and future prospect. *Kekkaku (Tuberculosis)*, 80:481–489, 2005.

[31] Stivelman B, Hills B. Effect of influenza on pulmonary tubeculosis. *N Y Med J.*, 110:20–21, 1919.

[32] Taubenberger JK, Morens DM. 1918 influenza: the mother of all pandemics. *Emerg Infect Dis.*, 12:15–22, 2006.

[33] Taubenberger JK, Reid AH, Lourens RM, Wang R, Jin G, Fanning TG. Characterization of the 1918 influenza virus polymerase genes. *Emerg Infect Dis.*, 12:15–22, 2006.

[34] United States Department of Commerce. Vital statistics rates in the united states 1900-1940. *United States Government Printing Office, Washington DC, USA.*, 1943.