LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Demography

Effects of in-utero influenza exposure on later life outcomes: Evidence from 19th century Scania

Filip Andersson
ake07fan@student.lu.se

Abstract: In this thesis the impact of in-utero influenza exposure on socioeconomic status and survival in adult life is studied. The data used is a subset of a longitudinal level database for Scania for 1815 to 1910. In this thesis the outbreak of influenza pandemic in 1833 is used as a natural experiment. The pathway model used in this thesis allows for both direct effect on the survival form the in-utero influenza exposure and indirect effects mediated through socioeconomic status in adult life for the individual. The analysis is done separately for women and men, and support for "Fetal origins hypothesis" as a negative effect on survival in later life is found for both sexes.

Key words: In-utero influenza exposure, adult health, socioeconomic status, dynamic path analysis, econometrics, economic demography, survival analysis

EKHM52
Master thesis, second year (15 credits ECTS)
June 2015
Supervisor: Jonas Helgertz
Examiner: Martin Dribe
Word Count: 22 944

Table of Contents

1. Introduction

1.1. The aim and goal of the thesis

Pathways to health and well-being is one of the big fields in the topics of demography and the work to identify critical periods over the life course that are more important compared to other periods to for example minimize the risk of certain diseases in old age. One of the well-known established theories that identify a specific critical period in the life course is the "Fetal Origins Hypothesis" (Barker, D. (1990, 1995)). This hypothesizes that the first period in life, namely the period as fetus, is a critical period for the development of the human being and that bad conditions during this period could lead to consequences in later life, for example higher risk of certain diseases. In this thesis the "Fetal Origins Hypothesis" and the related pathway model will be analyzed in a historical context.

For this thesis the analysis are divided into two main parts. The first part has the aim to investigate the effects on the adult life socioeconomic factors from being exposed to an influenza pandemic during fetal stage. The motivation for this part of the study is that almost every study that focuses on this particular problem has used similar natural experiment as the setup for the influenza exposure, namely the Spanish flu. These studies, like for example Almond, D. (2006), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010) are able to replicate each other's results for different countries, but do not touch upon the possibility that this particular pandemic, the Spanish flu, might have been significantly different to other influenza outbreaks throughout history. It is also worth pointing out that none of these author claims that the relationship found for the Spanish flu can be generalized for other influenza outbreaks. Furthermore, researchers have also to less extent worked with the Asian flu pandemic in 1957, see for example Kelly, E. (2011). So to furthermore test how well the outcomes for influenza outbreaks holds for another, previously not tested influenza pandemic, the first part of the analysis in this thesis will test if another influenza pandemic, in this paper the pandemic in 1833, also provide results for socioeconomic penalty for individuals affect in-utero.

For the second part of the analysis, this thesis moves on to focus on studying the survival in later adult life for individuals being exposed to influenza pandemic during the fetal stage compared to other individuals born within similar conditions except this specific feature. The question here being, how and in which way could this affect the survival chance in adult life for these individuals? There have been a few studies working with, at least partly, this type of setup for individuals affected in-utero by the Spanish flu, but the results have been inconclusive in regard of survival, see for example Cohen, A., Tillinghast, J. and Canudas-Romo, V. (2010) and Fletcher, J. (2014). On the other hand do studies by Almond, D. and Mazumder, B. (2005) and Garthwaite, C. (2008) find effects from the Spanish flu on later life health outcomes, which could indicate that there still exists an effect on later life health from being exposed in-utero of an influenza pandemic on survival.

The theoretical approach that is the fundamental ground for effects from early life conditions on later life health that will be considered in this thesis originates from the "Fetal origins hypothesis" (Barker, D. (1990, 1995)). Furthermore, to extend the analyzed model this thesis make use of the extended pathway model (Bengtsson, T. and Mineau G. (2009), Edvinsson, S. and Broström, G. (2012) and Palloni, A., Milesi, C., White, R. G. and Turner, A. (2009)). The data used in this thesis

consist of a subset of the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) for cohorts born between 1815 and 1835.

In line with the studies by Almond, D. (2006), Almond, D. and Mazumder, B. (2005), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010) the outbreak of an influenza pandemic will be seen as good model for the use of a natural experiment and will be handled as the treatment variable, but instead of using the Spanish flu, this thesis will take use of the pandemic outbreak in 1833.

Two important factors that could have had impacts on the outcomes in later life will be corrected for in both of the two analyzes. These factors are the socioeconomic status of the father at birth, which could have effects both for the analysis of the individual socioeconomic status in adult life, due to the generational inheritance of assets and status, and the survival, especially in adult ages. This control is also used in studies by for example Bengtsson, T. and Helgertz, J. (2013) and Quaranta, L. (2013). Furthermore, another factor that could cause bias in both the study of socioeconomic status and survival in adult life is other outbreaks of infectious diseases during fetal and early life stages for individuals. This field of study originates from the same theoretical background, the "Fetal origins hypothesis" (Barker, D. (1990, 1995)), and is used in a lot of studies, for example by Andersson, F. (2014), Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2013), that finds strong support for effects in later life survival from being born during periods of high infant mortality.

Since this thesis take the approach of a pathway model to explain the full life course events of how an influenza pandemic affects the individual during his or her life, the socioeconomic status of the individual is considered to be a mediating factor for the analysis of the adult life survival, namely the second part of the analysis. The reasons behind the choice of the mediating factor in this part of the analysis are that, as presented before, the strongest finding from the papers on the effects from the Spanish flu has been that those individuals that was affected in-utero had a socioeconomic penalty and that some other papers find support for health issues in later life, but that the analysis on direct effects on mortality form the Spanish flu has been inconclusive. Therefore, there may exist an indirect effect that has not been tested for, but that will be included as a part of this thesis.

As for this thesis the direct effect from being exposed to an influenza pandemic in-utero on the socioeconomic status in adult life is expected to be found, but due to pervious inconclusive results, the effects on survival is harder to predict. Furthermore, the indirect effect mediated through socioeconomic status could be a solution to measure the effect on survival from being affected of an influenza pandemic in-utero. In a more broad spectra one implication that is important to take into account from this thesis, and other papers focusing on the effects from in-utero and early life exposure to influenza pandemics, is that more studies finding negative effects for the fetus in later life would lead to more motivation for pregnant women to take vaccination against influenza and in general be extra carefully during periods of influenza outbreaks.

As stated before the analysis of this thesis consists of two different parts. The first part is the analysis of the effect on socioeconomic status. For this part the methodological approach will be to use logistic regression with random effects. This approach is quite similar to the approach by

Bengtsson, T. and Helgertz, J. (2013), which also uses random effects to measure the effects from the Spanish flu, but they do not use binary outcomes and instead uses linear regression.

The second methodological approach for the analysis of the direct and indirect effects on survival in later life will be taken from the dynamic path analysis, presented by Aalen, O., Borgan, Ø. and Gjessing, S. (2008). This model was later modified by Broström, G. and Edvinsson, S. (2012), to better fit analysis regarding data from a demographic context. Broström, G. and Edvinsson, S. (2012) presents a different linear hazard model compared to Aalen, O., Borgan, Ø. and Gjessing, S. (2008) called accelerated failure time instead of the Aalen´s additive hazard model to better fit analysis of survival in later life.

1.2. Summary of the findings in the thesis

The results found in the first part of the analysis of this thesis of the effect from in-utero exposure of influenza on the socioeconomic status in adult life is in line with results found in studies of the Spanish flu, like for example Almond, D. (2006), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010), namely that individuals exposed in-utero to an influenza outbreak had higher risk of belonging to lower socioeconomic class. The results were found both for men and women and could be seen as indication of the "Fetal origins hypothesis", but the mechanism cannot be dealt with in more detail, like studying the educational attainment, due to that this information do not exist in the used dataset. Anyhow the expected finding points to a direct scarring effect on influenza in-utero exposed individuals.

The second part of the analysis in this thesis focuses on analysis of a pathway model for the effects on survival in later life ages from the in-utero influenza exposure. From this analysis a direct negative effect on the survival in later life from being exposed in-utero of an influenza pandemic is found for women, which should be seen as support for the "Fetal origins hypothesis" and is an opposite result compared to the inconclusive results that studies by Cohen, A., Tillinghast, J. and Canudas-Romo, V. (2010) and Fletcher, J. (2014) found for the Spanish flu. Furthermore, for men indirect negative effect on the survival in later life is found from the in-utero influenza exposure mediated through the socioeconomic status of the individual in adult life. This should also be seen as support for the "Fetal origins hypothesis" even though not as clear as for women. It is hard to compare this result since no other study, to the knowledge of this thesis author, previously have tested for indirect effects within the setup of in-utero exposure to an influenza pandemic.

1.3. Outline of the thesis

There are six chapters in this thesis. The first chapter consists of the introduction, here the aims of the thesis are presented and the research questions are formulated, together with a short summary of the results.

Chapter two in this thesis consists of the previous results related to this thesis and the theory. This part is divided into six subparts, the first presents the earliest studies that lead to the development of the "Fetal Origins Hypothesis", the second is the direct presentation of the "Fetal Origins Hypothesis" which is the foundation of the theory for this thesis, the third deals with the later studies of the hypothesis, the forth presents the mechanisms for in-utero exposure for influenza, the fifth part is regarding the extended pathway model and the last part presents

the background of influenza pandemics in general and the outbreaks during the 1830s specifically.

The third chapter in this thesis is dealing with the methodology. This chapter consists of four parts. The first part presents the pathway model that the analyzes will be based on and the possible mechanism for this model. The second part presents the first statistical method used for the analysis of the direct effect on socioeconomic status in adult life, namely the logistic regression model. The third part presents the statistical model of dynamic path analysis, used in the second part of the analysis. Lastly the forth part of chapter three presents and discusses the assumptions and limitations of a natural experiment.

Chapter four presents the specific features of the data used in this thesis. Here, three parts are presented, the origin of the data, the structure of the data and the variables used in this thesis.

The fifth chapter presents the results from the analyzes done in this thesis. There are three parts in this chapter. First some descriptive statistics, then the analysis of the direct effect on socioeconomic status and lastly the analysis of direct and indirect effects on survival.

The last chapter is the conclusions. In this chapter the results are related to the theory presented in chapter two and the model presented in chapter three.

2. Previous research and theoretical considerations

2.1. Early related studies

To find the starting point of the research development that lead to the background theory behind this thesis, studies dating back to the early 20th century has to be considered. The main field that these studies did focus on was the general mortality decline that happened, at least, within the Western World as a starting point from the earlier part of the 19th century and onwards (Lee, R.D. (2003)). During this period researchers like for example Kermack, W., McKendrick, A. and McKinlay, P. (1934) presented results for England and Sweden that pointed to that cohorts of individuals that in general had better health during the period as young children in their first year of life had better health outcomes over the full life course. These results have also been seen in later studies done for different countries, for example France, by Preston, S. and van de Walle, E. (1978) and Switzerland, by Finch, C. and Crimmins, E. (2004). Since these studies focused on cohort based results no clear causality within the results can be identified due to the problem with periodic changes over time in the general mortality, still it can be seen as a start for the development of the important focus on the hypotheses that early life factor may have impacts on the individual not only at the that point in time, but also could pose as an important factor for outcomes later on in life. One of the first researchers to formulate a hypothesis in this manner was Preston, S. and van de Walle, E. (1978), who presented the idea that the exposure to higher disease load in early life could result in mortality reduction in later life.

Furthermore, Forsdahl, A. (1977), for Norway, and Barker, D. and Osmond, C. (1986), for Britain, found correlation between county specific infant mortality rates and an increased risk of heart disease in old age for individuals belonging to the cohorts born during years of higher infant mortality rate. Barker, D. and Osmond, C. (1986) also found similar results for other diseases, like for example bronchitis and emphysema. Even though these studies are independent of the problems with cohort periodic changes, these studies still suffers from the problem of controlling for cofounders and cannot identify critical periods for the individual development and can therefore not be seen as causal relationships (Bengtsson, T. and Helgertz, J. (2013)).

Another group of studies that originates from the studies by Kermack, W., McKendrick, A. and McKinlay, P. (1934) are a number of studies that focused on the Dutch "Hunger Winter" during 1944-1945 (Almond, D. and Currie, J. (2011)). What happened was that in November of 1944 the German Nazi forces cut the food shipments to the Netherlands, which resulted in a period of extreme nutritional problems for the Dutch people (Almond, D. and Currie, J. (2011)). Ravelli, G., Stein, Z. and Susser, M. (1976) found the risk of being obese for adult men to be double the normal risk if the individual was affected by the famine during fetal stage in the first trimester.

Furthermore, another important finding that had an important impact of the studies regarding the importance of conditions in the early stages of life happened in the early 1960s. During a couple of years in the late 1950s and the early 1960s a medication called thalidomide was prescribed to pregnant women to ease the problems with morning sickness Almond, D. and Currie, J. (2011). The studies by McBride, W. (1961) and Lenz, W. and Knapp, K. (1962) found that children to women who took the medicine had an extensive higher risk of extreme birth defects, like for example missing limbs. According to Dr. Philip Landrigan at the Department of Community and Preventive Medicine at Mount Sinai School of Medicine in New York, was this

the first time that researchers and medical doctors understood that the placenta did not work as a perfect protection for the fetus (Almond, D. and Currie, J. (2011)). Furthermore, other actions by the mother were also later found to have similar bad impact on the fetus, like for example smoking, alcohol or drug use (Almond, D. and Currie, J. (2011)).

2.2 Fetal origins hypothesis

As a result of the of the findings, both in his own work and by other authors, like these presented in the previous chapter, David Baker formulated the hypothesis of the importance of infant and fetal stage in life on the later life health outcome as a start in Barker, D. (1990). The hypothesis mostly goes under the name of the "Fetal Origins Hypothesis" (Barker, D. (1990, 1995)). The main hypothesis of the "Fetal Origins Hypothesis" is that fetuses can be affected by hazardous exposure to the mother during the period of in-utero and that this can be linked to health outcomes in the adult life of the individual (Almond, D. and Currie, J. (2011), Barker, D. (1995), Calkins, K. and Devaskar, S. U. (2011)). Furthermore, the hazard that the fetus experience during the period of in-utero could have an effect on the health of the individual both as a continuous effect and as a latent affect that could be "activated" at any time point during the individuals' life course (Almond, D. and Currie, J. (2011)). Another aspect of how the hazard that the "Fetal Origins Hypothesis" presents act though is fetal malnutrition. The effect from the malnutrition is according to the "Fetal Origins Hypothesis" that these fetuses could have a risk of getting an altered gene expression, which could lead to negative outcomes on the individual's physiological and cognitive abilities (Kelly, E. (2011)). Even though, as has been presented in the previous chapter, Baker was not the first researcher to present hypothesis and results for latent effects from early life factors on later life health, Baker hypothesized the general idea more specific and added three important aspects to the model of latent effects (Almond, D. and Currie, J. (2011)). These three aspects are the possible linking of early life effects on later life chronic degenerative diseases, like non-insulin dependent diabetic mellitus and cardiovascular disease (Rasmussen, K. (2001)), the use of better fitting data to analysis within this field and Barker worked hard to expand and highlight the importance and the focus on these hypotheses (Almond, D. and Currie, J. (2011)).

2.3. Recent studies and the widening of the fetal origins hypothesis

During the mid-1990s and onwards many researcher, both in the fields of economics, economic history, epidemiology and medicine have set up different ways to test and measure the possible effects from different in-utero and early life environments on later life outcomes for individuals and have found supports for the "Fetal Origins Hypothesis" by Barker, D. (1990, 1995). According to Almond, D. (2006) some of the most prominent epidemiological studies that have found support for the hypothesis are papers focusing on famine episodes and especially, as even touched upon previously in this chapter, the Dutch famine. Within this setup Roseboom, T., van der Meulen, J., Ravelli, A., Osmond, C., Barker, D. and Bleker, O. (2001) found results from individuals being in-utero affected by the famine on self-reported health in adult life, Roseboom, T., van der Meulen, J., Osmond, C., Barker, D., Ravelli, A., Schroeder-Tanka, J., van Montfrans, G., Michels, R. and Bleker, O. (2000) and Painter, R., Roseboom, T. and Bleker, O. (2005) found support for higher risk of coronary heart disease in adult life from in-utero exposure to the Dutch famine and Neugebauer, R., Hoek, H. and Susser, E. (1999) found results on psychological outcomes from the famine. Specifically the findings by Roseboom, T., van der

Meulen, J., Osmond, C., Barker, D., Ravelli, A., Schroeder-Tanka, J., van Montfrans, G., Michels, R. and Bleker, O. (2000) and Painter, R., Roseboom, T. and Bleker, O. (2005) follows the previous results by Forsdahl, A. (1977) and Barker, D. and Osmond, C. (1986) and should be seen as strong support to the "Fetal Origins Hypothesis" (Barker, D. (1990, 1995)).

Another important indicator of fetal stress that has been widely used in studies that focuses on latent effects in later life health is the birth weight of the individual (Bengtsson, T. and Helgertz, J. (2013)). This method of measuring the condition of the fetal stage is used by Barker, D. (1995), where the author measures the correlation between birth weight and adult life mortality and specifically the relation to heart conditions in adult life.

Yet another area of the measurement of the fetal condition and the early life condition that has been used by a number of researchers is the method presented by Bengtsson, T. and Lindström, M. (2000, 2003). In these studies the authors use of longitudinal data on individual level and try to account for both in-utero conditions and infant conditions. Bengtsson, T. and Lindström, M. (2000, 2003) uses the maternal mortality during the in-utero stage as measure for that period and the infant mortality rate during the year of birth as the measure of the disease load during infancy. The authors then combine these individual level measurements with aggregated price data during the two periods to measure the nutritional status of the individual during these critical periods. The main results from the studies by Bengtsson, T. and Lindström, M. (2000, 2003) show that the disease load in the year of birth had a high impact on the adult life mortality. From these findings other authors have either used the infant mortality rate in the year of birth as a direct measurement of the early life conditions or extended the measurement to combine it with the general disease load in specific years where outbreaks of whooping cough and small pox has played an important role, this studies are for example Bengtsson, T. (2004), Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2013).

The final presented and to this thesis most important setup of the "Fetal Origins Hypothesis" is the method of testing the hypothesis on a specific event as presented by Almond, D. (2006) and Almond, D. and Mazumder, B. (2005). Here the author focuses on the influenza outbreak in 1918 called the Spanish flu. This pandemic outbreak is in these studies used as a natural experiment to indicate if an individual was in fetal stage while the mother was high risk of being exposed to the influenza. The main motivations for Almond, D. (2006) to study the possible outcomes in later life from the risk of influenza exposure are presented to be that studies by Collier, R. (1974) and Ravenholt, R. and Foege, W. (1982) found indications that influenza exposure could increase the risk of sleeping sickness and Parkinsonism, although the results have been questioned. Furthermore, indications of higher risk of schizophrenia has also been suggested by some authors and Almond, D. and Mazumder, B. (2005) also found indications of higher risk for diabetes and stroke in later adult life for individuals that was exposed in-utero to the Spanish flu. The use of a specific event as an indicator of fetal stress can be designed with a big variation of possible setups; one that shows how different this setup can be is used in the study by Almond, D., Edlund, L. and Palme, M. (2009), where the authors investigate the effects on cognitive ability for individuals born in Sweden in 1986 from the nuclear fallouts caused by the disaster of the nuclear power plant in Chernobyl. From this study the authors finds lower educational attainments and cognitive ability scores for the affected individuals.

Also important within the field of economics in relation to the "Fetal Origins Hypothesis" is that numbers of researchers, with different setups as a starting point has found effects on not only the health, as the original hypothesis focuses on, but also other outcomes over the life course and especially socioeconomic and human capital outcomes have caught a lot of attentions. One of the studies that broke the ground for the study the fetal stress on economic outcomes is, according to Almond, D. and Currie, J. (2011), the study by Currie, J. and Hyson, R. (1999). In this study the authors uses the method of birth weight as the indication of the fetal stress and test the effects of low birth weight on educational and employment variables. The findings from this study was that individuals born with a birth weight under 2 500 grams experienced a high risk of preforming worse in school and had a lower chance of having a job in adult life. Following this study a lot of other studies, using more or less every in this chapter suggested different setups for the measurement of the "Fetal Origins Hypothesis", have tested the possible effects on different human capital outcomes.

Since the setup that will be used in this thesis is the method of testing the "Fetal Origins Hypothesis" on a specific event, in this study an influenza outbreak, some more on the findings from studies testing the effects in later life from influenza exposure in early life, especially in-utero, will be presented. One of the most important papers within this particular subcategory of the "Fetal Origins Hypothesis" is the paper by Almond, D. (2006), in this paper the author finds support for effects from in-utero exposure of the Spanish flu on human capital outcomes in later life for the United States. These effects are both found as worse educational outcomes of the affected individuals and lower earnings in adult life. On the other hand, this study has been criticized according to Bengtsson, T. and Helgertz, J. (2013) due to a shift in the socioeconomic structure of the parents during this period that was caused by how the drafting system worked during the First World War that coincided with the Spanish flu. This problem might have caused overestimation of the results found in Almond, D. (2006). Though this problem did exist for the study by Almond, D. (2006) other researchers have done similar studies on countries that were not involved in the war, like for example Nelson, R. E. (2010) for Brazil and Bengtsson, T. and Helgertz, J. (2013) for Sweden. From both the study for Brazil and Sweden the authors find similar effects in line with the study by Almond, D. (2006), which highly indicates that the effects from the Spanish flu seem to exist. Even the possibilities of multigenerational effects have been tested and Richter, A. and Robling, P. (2013) finds support for socioeconomic multigenerational penalty from the Spanish flu from data for Sweden, at least for women.

Furthermore, studies have been done for the outbreak of the Chinese famine in the later 1950s and early 1960s where authors have found significant negative effect on the chance of getting married for the in-utero affected individuals, see for example Almond, D., Edlund, L., Li, H., and Zhang, J. (2010) and Brandt, L., Siow, A. and Vogel, C. (2008). Other studies of the Spanish flu have focused on the health impact in later life from being affected by the influenza outbreak in-utero. Here, the most conclusive results has been on the health status in later life, seen as from example either self-reported health, as in Almond, D. and Mazumder, B. (2005) or as increased risk of specific health issues in later life, as in Garthwaite, C. (2008), where the author find support for higher risk of coronary heart disease, diabetes and kidney disorders. On the other hand, in contrast to most studies of the "Fetal Origins Hypothesis" studies that have focused on the direct impact from the in-utero exposure of an influenza pandemic on the mortality in later

life have been inconclusive, like for example Cohen, A., Tillinghast, J. and Canudas-Romo, V. (2010) and Fletcher, J. (2014). This will be further investigated in later part of this thesis, since according to the direct theory of the "Fetal Origins Hypothesis" and since health issues in later life has been found for the Spanish flu an effect on the mortality in later life as an outcome from an influenza pandemic still seems possible.

2.4. Mechanisms regarding the fetal origins hypothesis and influenza exposure

As presented in section 2.2. the "Fetal Origins Hypothesis" highlight the importance of the nutritional status during the fetal stage for the development of the individual. This section will focus a litter deeper from a medical perspective of the consequences of this hypothesis and specifically in regard of influenza affected mothers, since the main focus in this thesis lies within that field.

At first there are two main sources from where the nutrients that the fetus receives originate, namely the mother's diet and the stored nutrients in the body of the mother (Kelly, E. (2011)). According to Kelly, E. (2011) presented David Barker at the RAND Mini-Medical School for Social Scientists conference in 2009 that it is mainly the stored nutrients that have the most impact on the growth of the fetus. Furthermore, the fetus receives these nutrients by the placenta and how well this procedure works can be determine by the size of the placenta and the size of the placenta is then affected by the nutritional intake of the mother (Kelly, E. (2011)). Furthermore, if the mother would contract influenza there are three possible mechanisms that could affect the nutritional flow from the mother to the fetus. The first is that if the mother gets the influenza she could get lower appetite and lower her nutritional intake, which would lead to decrease of the store of nutrients in her body and as a consequence lower nutritional transfer to the fetus (Metzger, B., Vileisis, R., Ravnikar, V. and Freinkel, N. (1982)). The second risk is that the body during influenza could interfere with the absorption of fat, proteins and other nutrients of the mother due to higher rate of excretion (Kelly, E. (2011)). The third and last risk is that the fever, that influenza generally comes with, could increase the consumption of the nutrients in the mother's body, which then could lead to lower levels of nutrients. This effect could be large especially if the mother even before the infection had lower levels of nutrients in her body (Edwards, M. (2007)).

In regard of the development of the brain in connection to influenza exposure in-utero this part of the fetus is more sensitive to most types of hazard then other parts. Results for this have been shown from many of the empirical studies of the outcomes of the cognitive ability from exposure of influenza, presented in the previous section. According to Otake, M. and Schull, W. (1998) there exists two specific critical periods in the fetal development of the brain, where the fetus should be most at risk of damage that could be seen in the cognitive ability and other problems in later life. These two periods are first when the fetus is 8-15 weeks. During this period of time the start of the development of the brain takes palace and unfavorable conditions during this fetal stage could highly increase the risk of degenerative effects on the cognitive ability (Otake, M. and Schull, W. (1998)). Furthermore, the second period that the fetus might be the most susceptible to fetal stress in regard of the brain development takes place right after, namely week 16-25. During this period fetal stress could highly increase the risk of for example schizophrenia in adulthood (Otake, M. and Schull, W. (1998)). This result if for example highlighted in the study by (Rasmussen, S., Jamieson, D. and Bresee, J. (2008)).

2.5. The pathway model

An extended model that is often closely related to "Fetal Origins Hypothesis" is the pathway model (Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012). The main different between these two models, that has to be considered, is that when original framework of the "Fetal Origins Hypothesis" focuses on the direct latent effect from the conditions in early life on the later life health or, as presented in the previous section also the socioeconomic outcomes, the pathway model also includes the role that the adult environment has in regard of the outcomes in later life (Nelson, R. E. (2010)).

In the pathway model the effects from the fetal stress is seen to be able to act both as direct latent effect on the outcome, as in the "Fetal Origins Hypothesis", but also as an indirect effect that could be mediated though other different factors (Edvinsson, S. and Broström, G. (2012)). This could partly be seen as a combination of the assumption that the fetal stress would have an effect on the health in later life as in the studies by Barker, D. (1990, 1995) and the effect the fetal stress could have on the socioeconomic outcomes as the study by Currie, J. and Hyson, R. (1999), but here the effect of the socioeconomic variable also is seen as a mediator for the later life health outcome. There are a big variety of suggested mediators for possible indirect effects in a pathway model for example socioeconomic factors, educational factors, cognitive ability and marital status. Other variables could be seen as confounding factors that has to be taken into account within a pathway model. These factors could be environmental factor such as Gross Domestic Product per capita, which has been studied by (Van den Berg, G., Doblhammer, G. and Christensen, K. (2009) and Van Den Berg, G., Lindeboom, M. and Portrait, F. (2006)), but also the parental socioeconomic situation. Here, the parental socioeconomic status could pose as a big source of bias if not taken into account into the pathway model. There are two main reasons for this, firstly this variable could have an impact on the socioeconomic success of the individual in adult life, both due to educational investments and genetics (Almond, D. (2006) and Almond, D., Edlund, L. and Palme, M. (2009)). Also, Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012) finds that individuals with parents of higher socioeconomic status have a lower risk of having a low socioeconomic status themselves in adult life. Secondly, Rosenzweig, M. and Zhang, J. (2009) also find indications that the parental socioeconomic status can be correlated with the birth weight of the individual, meaning that controlling for the socioeconomic status of the parents could remove unwanted bias in a pathway model.

Furthermore, another factor that could work as mediating covariate within a pathway model, especially within studies of a more historical context is the marital status of the analyzed individual (Edvinsson, S. and Broström, G. (2012)). As also presented in section 2.3. this factor was directly affected in the context of "Fetal Origins Hypothesis" in the studies of the Chinese famine by Almond, D., Edlund, L., Li, H., and Zhang, J. (2010) and Brandt, L., Siow, A. and Vogel, C. (2008) and Edvinsson, S. and Broström, G. (2012) also argues that marriage posed as an emotional and financial support for individuals within a historical context and could therefore also have an impact on the mortality of the individuals.

As discussed partly in section 2.3. environmental external variation could also have an impact on the individual throughout the life course. Here, an important factor could be economic variation, but according to Bengtsson, T. and Lindström, M. (2000) this affect is mainly in the years as a child, but not the earliest part of the life course. Quaranta, L. (2013) explains this to be due to

that during the period that the child is still breastfed the most important factors for the individual development are the disease load and the lifestyle of the mother. On the other hand economic variation could change the lifestyle of the mother and in that situation an indirect effect on the young child working through the mother could still be present.

The last suggested mediating factor for a pathway model that will be presented in this thesis is the measurement of the height of an individual. This variable is often seen as an indicator of health in adult life that could have been affected by fetal stress, see for example Fogel, R. (2005) and Öberg, S. (2014). Öberg, S. (2014) argues that the height in adult life of an individual is affected by four different factors, genetic factors, environmental factors, nutritional factors and disease load. The important factors here that are closely related to the "Fetal Origins Hypothesis" are the nutritional factors and the disease load. Both these factors are seen, as presented in section 2.3., as factors that in early life could affect later life outcomes and if they also would affect the height of the individual then the height would be a suitable mediating factor in a pathway model. Furthermore, Öberg, S. (2014) argues that there is a direct negative effect on height, from the disease load in early life, but that there could also exist a selection process where only the strongest individuals survives the disease load and they could then be the individuals with the best genetics which would also become the tallest individuals.

As the pathway model is a more open than the "Fetal Origins Hypothesis", the model allows both for critical periods and periods of accumulated risk (Bengtsson, T. and Mineau G. (2009)). Therefore, when analyzing a pathway model, one important factor to take into account is to distinguish measures that is directly related to early life conditions and fetal stress and other factors (Edvinsson, S. and Broström, G. (2012)). As presented throughout this section of the thesis the often considered factors in the pathway model consist of cognitive and educational factors, socioeconomic factors, marital factors and health over the life course.

Within the structure of a pathway model the effects need to be taken into account as both effects and counter effects (Bengtsson, T. and Mineau G. (2009)). These are seen as either scarring or selecting effects. For example, there may exist a latent negative effect of disease load on later life health, a scaring effect, but there could also be a selection where only the strongest fetuses survive to even be born and those individuals may have better outcomes over the life course compared to the average population, a selection effect.

Closely related to the possible existence of both scarring and selection within studies of the pathway model, the risk of confounding factors has to be taken into account (Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012)). The problem here is that at each point in time both the scarring and the selection might take place and since these effects could work as counter effect and generate no results at all. Furthermore, during the evaluation of the results found in a pathway model both effects also has to be taken into account.

What is a clear picture from this part of the thesis is that the pathway model presents a lot of variation and possible models to test, but also due to this fact there do not exist a direct answer to how the pathway model should be structured and therefore makes the modeling of the effects complex. Even so a lot of the latest research within this field of studies takes the approach of a pathway model and that is also what will be done in this thesis as the second half of the analysis.

2.6. Influenza and the outbreak of pandemics during 1815 to 1840

Influenza is a highly infectious viral infection caused by the influenza virus, a rapidly mutating virus that causes outbreaks more or less every year at least somewhere in the world (C. Potter (2001)). The virus itself was first isolated in a laboratory environment in 1932, but due to its very distinct symptoms, including sudden fever lasting in general for three days, it has been fairly easy to identify outbreaks of influenza even in a historical perspective before the virus could be isolated (C. Potter (2001)). In general both the severity and the timing of the outbreaks has been very hard to predict, but from records of historical events the general tendency for an outbreak to proceed into a more sever situation, epidemic or pandemic, it seems to be that these often happens in the colder months of the year, when the immune system of individuals tend to be weaker (C. Potter (2001)). Furthermore, there seems to be a natural spreading pattern of the epidemics where it first shows up in Asia and then travels through Russia to Western Europe and North America (C. Potter (2001)). The spreading westwards from Russia into Europe can be seen since at least the 18th century (Patterson, K. (1985)). The last found characteristics of when the influenza virus tend to hit harder seems to be when the virus mutates into a new genetic form, since individuals that had developed anti-bodies against the old form now also would be a target to the mutated influenza virus (C. Potter (2001)).

The most severe outbreaks of the influenza virus can be defined as pandemics. For an influenza outbreak to be defined as a pandemic C. Potter (2001) presents some criteria that has to be fulfilled, these are that the outbreak must spread widely with a high risk of infecting a lot of individuals, result in higher mortality rates than normal and the version of the virus has to be of a specific subtype. Some of the more recent and famous pandemics include the 1889 Russian flu and the 1918 Spanish flu (Patterson, K. (1985)). Especially the second one is well known and defined as one of the most severe pandemics in the history and up to 50 million peopled died worldwide in this influenza pandemic (C. Potter (2001)). Due to the severity of the pandemic outbreak in 1918 it has posed as a well-used study case for a lot of different studies of possible latent risks for individuals infected by the influenza, for example Almond, D. (2006) investigates later life socioeconomic outcomes for individuals whose mothers were in risk of being infected when they were pregnant and Almond, D. and Mazumder, B. (2005) look at health outcomes for the same groups of individuals.

As this thesis relies on data from the 19th century, the pandemic influenza outbreaks that are of most interests are those during the first half of that century. The paper "Pandemic and Epidemic Influenza 1830-1848" by Patterson, K. (1985) highly focuses on this specific time period and some of the most important features of those outbreaks during that time period. Therefore, in the following paragraph the influenza outbreaks during the 1830s will be deeper presented and discussed. It is also important to point out that according to Patterson, K. (1985) the influenza activity in the previous decades had been low and that no pandemic outbreak of the influenza virus had happened since 1782 and that after the outbreak in 1848 the next outbreak of an influenza pandemic was the Russian flu.

The first wave of the influenza virus at bigger scale during the 1830s happened during the period of 1830-1831. During 1830 the influenza hit various places in Asia and in November 1830 it hit Russia and then reached Scandinavia in the spring of 1831 somewhat around March-May (Patterson, K. (1985)). The general characteristic of this influenza pandemic was that it had high

morbidity, but low mortality and for Europe the pandemic was considered mild, but widespread (Patterson, K. (1985)). Due to the high morbidity and the fast spread of the influenza during 1830-1831 it can clearly be seen that it had pandemic proportions.

The next strike of influenza during the 1830s happened in 1833. This influenza strike was more geographically limited, mostly to Europe compared to the wave in 1830-1831 which was widespread all over the world (Patterson, K. (1985)). On the other hand this influenza strike showed higher level of mortality and also high morbidity compared to the previous strike (Patterson, K. (1985)) might seem a bit contradictive due to the more limited area spread. One explanation to this could be the more limited connections between countries during the period of the outbreak. Another possible explanation could have been that due to the historical context of the outbreak the data sources form this time is less widespread. This pandemic influenza strike hit Denmark and the southern parts of Sweden in March to May 1833 and Gottlieb Gluge documents in his book "Die Influenza oder Grippe" (Gluge, G. (1837) p.143) a clear top in the mortality for these months during 1833 compared to the surrounding years within the period of 1820 to 1833 for Copenhagen. Furthermore, the high morbidity is validated from data for Stockholm where around a fourth of the individuals where infected and as high number as up to 80 percent of the individuals in London and Paris was infected during this influenza pandemic (Patterson, K. (1985)). The big variations within the levels of the morbidity of the influenza pandemic seems quite unlikely and could be due to the uncertainty of the data sources due to its historical context. This influenza pandemic showed indications that it may have been a second wave of the influenza in 1831, similar to what happened during the more well-known influenza pandemics in 1890, Russian flu, and 1918, Spanish flu, but the increased time lag between the two waves compared to these two other pandemics and that there cannot be found any evidence that individuals that were infected during the 1831 wave showed any immunity to the wave in 1833 indicates that it might instead had been two completely different viruses (Patterson, K. (1985)).

The last big wave of widespread influenza during the 1830s happened in 1836-1837. This wave of the influenza first hit Russia in November 1836 and then moved on to for example the southern and middle part of Sweden and Denmark in December the same year (Patterson, K. (1985)). This wave of the influenza had a more complex and different geographical spread than the earlier waves that more or less followed the common pattern of going from east to west, but instead from south to north including some other pattern as well (Patterson, K. (1985)). Furthermore, the specific characteristic of this influenza pandemic was a very high level of morbidity, with almost 50 percent of the population of Copenhagen becoming infected. Even though the morbidity of this influenza pandemic was high the level of mortality was low, but still the total number of deaths, due to high morbidity, was still high (Patterson, K. (1985)). Another specific characteristic of this influenza pandemic, which might indicate that the virus was not as potent as for example the virus in the wave of 1833, was that most of the deaths were among the elderly (Patterson, K. (1985))).

To summarize there were three main waves of the influenza virus during the 1830s that can be contributed to have been pandemics, the waves in 1830-1831, 1833, and 1836-1837. All of these influenza pandemic is characterized by high morbidity, but with different levels of individual level mortality. The waves in 1830-1831 and 1836-1837 had lower mortality compared to the wave in 1833, which seems to have been the most potent one of the influenza pandemics in the 1830s.

Even though the levels of total mortality was high for these pandemic outbreaks it is important to emphasize that compared to the 1918 Spanish flu, these pandemic outbreaks had much lower mortality (Patterson, K. (1985)).

3. Methodology

The chapter called methodology consists of three sections. In the first section the pathway model for the second part of the analysis is presented. This model is based on the theory and the connections to the theory will be discussed. The second section will present the statistical model used in the first part of the analysis, namely the logistic regression model and the third section will present the statistical model that is used for the analysis of the pathway model, namely the dynamic path analysis.

3.1. Mechanisms of the pathway model tested in the analysis

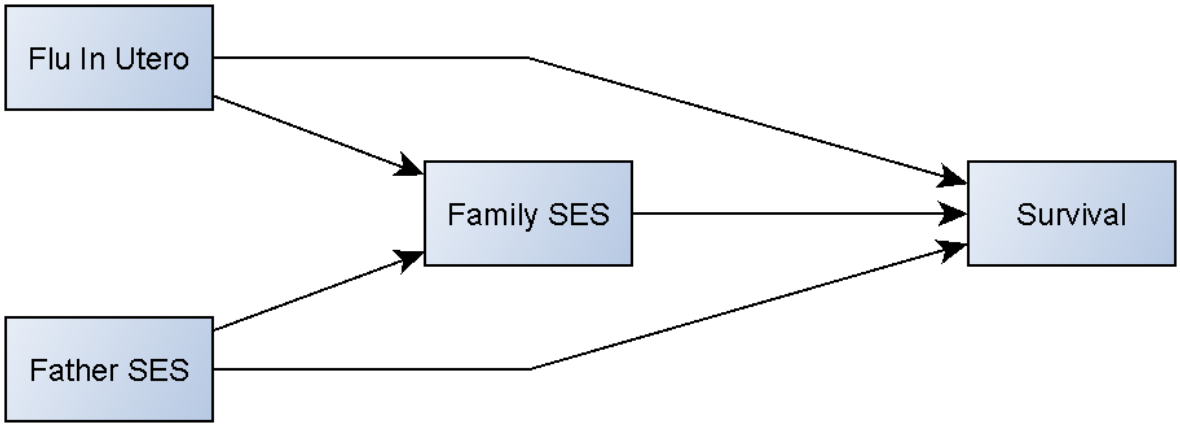

Figure 1 The pathway model used in this thesis

For this model there are two variables that are seen as the starting point of the model. These variables are assumed to be connected with the early life condition of the analyzed individual. The most important variable of these two are the variable that indicates if the individual was exposed to the 1833 influenza pandemic in-utero or not and how this would affect the survival in adult life as a direct latent effect as hypothesis by the "Fetal origins hypothesis" (Almond, D. and Currie, J. (2011) and Calkins, K. and Devaskar, S. U. (2011)). This can be seen in figure 1 as represented by the arrow that goes from flu in-utero to survival.

The second variable that is related to the early life condition is the socioeconomic status of the father at birth. The main function of this variable is to control for different variations in the early life and affect especially the socioeconomic situation of the individual in adult life as discussed by Almond, D. (2006) and Almond, D., Edlund, L. and Palme, M. (2009), but also early life environment as Bengtsson, T. and Lindström, M. (2000) and Quaranta, L. (2013) discuss that economic variation during childhood could have impacts in later life, and the parental socioeconomic status have a strong impact on how vulnerable the individual would have been to economic variation. In figure 1 the direct effect from the parental socioeconomic status is indicated by the arrow from father SES to survival.

Furthermore, the model used here also includes a variable for indirect effect that both the in-utero exposure of influenza and the father socioeconomic status are allowed to be mediated through. The mediating covariate in this model is the socioeconomic status of the individual in adult life. Similar mediating covariate are included in for example the analysis by Edvinsson, S. and Broström, G. (2012). Also, in regard of the models by for example Almond, D. (2006) and Bengtsson, T. and Helgertz, J. (2013), where the socioeconomic status are used as the outcome

variable, which also will be tested in the first half of the analysis of this thesis. The indirect effects can be seen in figure 1 as the arrows from the flu in-utero and the father SES to the family SES combined with the arrow that goes from the family SES to the survival.

The theoretical mechanisms that should be considered for the mediating effect in terms of scarring (Almond, D., Edlund, L. and Palme, M. (2009) and Palloni, A., Milesi, C., White, R. and Turner, A. (2009)) is that individuals that were exposed to the influenza pandemic in-utero had low cognitive ability or was physically weaker and that could lead to worse performance at work, which then would lead to lower socioeconomic status. As a consequence of this, lower socioeconomic status could lead to less possibility to live a healthy life, for example worse nutritional intake and could then lead to worse chance of survival in later life. As discussed in section 2.5. there could also exist a counter effect, or a selection effect (Bengtsson, T. and Mineau G. (2009)). Within this model that would mean that only the stronger individuals would even survive the influenza pandemic in-utero and those individuals would have a better ability to work then the average population, which then could lead to higher socioeconomic status and in the end to higher survival chance in later life.

3.2. Random effects logistic regression model

The first method used in the analysis to estimate the effects from the conditions in early life, the effects for an individual to have been affected in-utero by the influenza pandemic, takes the approach of random effects logit model. At first the use of this model compared to very similar probit model has to be addressed. The main reason for the use of logit model compared to probit model is that for the logit model, compared to the probit model, when estimating models with longitudinal data, which is used in this thesis, the logit model is the only one of the two models that yields consistent results. To achieve consistent results with a probit model within this setup the time variable has to go to infinity and for a fixed time variable, even when the number of individuals goes to infinity the model still would be inconsistent (Verbeek, M. (2012)). On the other hand this problem may not cause big issues within this study, due to that the time variable also is big, but there is no reason to risk inconsistent results when a better option is available.

The general way to define a binary choice model for longitudinal data is defined as (Verbeek, M. (2012)):

$$Y_{it}^{*} = X_{it}'\beta + \alpha_i + u_{it} \qquad (1)$$

Here, if the observation is $Y_{it} = 1$ if $Y_{it}^{*} > 0$ and $Y_{it} = 0$ in all other cases (Verbeek, M. (2012)). To derive the conditional maximum likelihood that proves consistent results for this estimation is quite straight forward for the situation when two points in time is considered, but for cases of lager dataset with a lot of time points, as in the case of this thesis, the derivation gets more complicated. Anyhow the proof for this is provided by for example Chamberlain, G. (1980) and Maddala, G. (1987).

It is worth to point out that the data structure in the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) is in the structure of an unbalanced longitudinal panel data, since every individuals are not recorded at every time points, but when a change in any of the parameters that the data consists of are recorded. This is however not a

problem since the statistical program STATA handles this issue by itself and the user do not have to worry about this problem at all.

3.3. Dynamic path analysis

The method that is used in this thesis to analyze the survival in later life is a modified version of the dynamic path analysis presented by Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006) and Aalen, O., Borgan, Ø. and Gjessing, S. (2008). This model is developed to analyze the survival form a path model, while allowing for time varying covariates and are able to analyze both the direct effects form a treatment variable and the indirect effects working through mediating covariates (Aalen, O., Borgan, Ø. and Gjessing, S. (2008)). As for this thesis the treatment variable will be, as presented before, individuals affected by an influenza outbreak during fetal stage, namely all individuals that was in risk of exposure of the influenza pandemic during the time of in-utero, and the mediating effect that will be tested for is the socioeconomic status of the individual in adult life, since the socioeconomic status in childhood is controlled for by the inclusion of the socioeconomic status of the father and that the characteristics of the individuals themselves do have an impact on the socioeconomic status before adulthood, for more detail see chapter 4.3.

The dynamic path analysis, by Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006), was originally made for analysis in medicine, especially the survival for treated patients with a specific medical problem. Therefore the direct model by Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006) is not proper to use, since the context here is instead later life survival in a demographic situation. This problem has been addressed by Broström, G. and Edvinsson, S. (2012), which has modified the model to instead fit in a demographic context. They have also taken into account the model presented by Lange, T. and Hansen, J. (2011), that is a combined model between the dynamic path analysis and the counterfactual model (Edvinsson, S. and Broström, G. (2012)). Broström, G. and Edvinsson, S. (2012) take the approach of accelerated failure time model that use the logarithmic value of survival in years as the response, compared to the Aalen´s additive hazard model, which is used by (Aalen, O., Borgan, Ø. and Gjessing, S. (2008)). The choice of an accelerated failure time model is proven in Broström, G. (2012) to fit will with demographic longitudinal data. Both in Broström, G. (2012) and Broström, G. and Edvinsson, S. (2012) it is shown that the best underline distributional assumption to use is the gompertz distribution. These articles shows that with this distribution the accelerated failure time model fits, at least almost, as well as the much more used Cox proportional hazard model for analysis of survival of individuals in older ages, but for younger adults and children Broström, G. (2012) concludes that the Cox proportional hazard model fits better.

The method of analyzing direct and indirect effects consists of two steps. The first is to analyze the direct effects from all covariates, including controls included in the model. This step is done with the accelerated failure time model (Broström, G. and Edvinsson, S. (2012)). The second step is to analyze the indirect effects working through the mediating covariate. This is done with linear regression of the treatment variable and the control variables on the mediating covariate (Lange, T. and Hansen, J. (2011) and Martinussen, T. and Scheike, T. (2006)). The linear equations are calculated simultaneously with respect to the accelerated failure time model. The structural equation system for a specific time, which in general is a time interval, follows as equation 2 and 3 (Broström, G. and Edvinsson, S. (2012)):

The accelerated failure time model:

$$\log(T) = \beta_0 + a_1 y_1 + \cdots + a_q y_q + \beta_1 z_1 + \cdots + \beta_p z_p + \varepsilon \quad (2)$$

The linear regression model:

$$z_p = \gamma_0 + \gamma_1 z_1 + \delta_1 y_1 + \cdots + \delta_q y_q + \eta \quad (3)$$

For the model z_1 is the treatment variable, y_1 to y_q is control variables, z_p are the mediating covariates and T is the survival time in years. The direct effect from the treatment variable is β_1 and the indirect effect is calculated by (Broström, G. and Edvinsson, S. (2012)):

$$Indirect\ effect = \beta_p * \gamma_1 \quad (4)$$

Here, β_p is the coefficient from the accelerated failure time model for the mediating covariate p. The total effect of the treatment is then calculated by (Martinussen, T. (2010)):

$$Total\ effect = Direct\ effect + Indirect\ effect \quad (5)$$

Finally, the results from the presented model can be interpreted as that the coefficient β_p is the logarithmic change in expected life compared to the baseline life expectancy. To express the effects in a normal scale the equation for is (Broström, G. (2012)):

$$Expected\ life = Baseline\ life\ expectancy * e^{\beta_i} \quad (6)$$

3.4. Natural experiment

For this thesis the main methodological setup that has to be addressed is the assumption that the risk of being exposed in-utero to influenza can be seen as a natural experiment or as it also is called a quasi-experiment. At first a definition of what a natural experiment is has to be addressed. The basis behind a natural experiment is that when doing social studies, especially historical once, it is impossible to set up a perfect experiment with good randomization mostly due to either, as in the historical perspective, not even an option, or it is unethical, compared to for example in the situation of natural sciences, where it is possible to control the environment within a laboratory setup (Durlauf, S. and Blume, L. (2008)). Furthermore, the costs, if a real experiment is accepted within social science, will be much higher compared to the use of a natural experiment and the researcher also have to deal with the problem of drop-outs, which may cause a big risk of bias if the drop-outs is not completely random (Durlauf, S. and Blume, L. (2008)). Instead to be able to find causal findings within social science one approach is to try design a natural experiment, namely to find a change or an event of interest that can be approximated to act as close as possible to if the event was controlled by the researcher. The main important assumptions that have to be fulfilled to make a good natural experiment are that the change or event has to have been close to random and unpredictable to the individuals within the study. Furthermore, the non-affected individual has to have similar characteristics compared to the individuals affected by the natural experiment. If the change or event would not be random and would be possible to predict there exist a risk that individuals with better knowledge could change their life pattern to prevent themselves from being affected by the natural experiment, which then could lead to worse outcomes for poor and poorly educated individuals.

Even if the setup of a natural experiment is done properly and would then seem as a good choice within social science to be able to find causal results within a study it still poses a number of problems that has to be taken into account by the researcher. In Durlauf, S. and Blume, L. (2008), three specific issues that have to be considered is presented as follows:

1. Even though the found result within a study proves causal relationship for the specific setup, if the test would be on a more broadly basis the results may differ. In other words the researcher has internal validity, but cannot prove external validity of the found results.
2. There is often a limitation of possible scenarios where it is possible to conduct a natural experiment and therefore there may be little variation within different studies and some "extreme" situations may be "over-used".
3. Results found from a natural experiment cannot be interoperated without preexisting theoretical explanations or assumptions.

In this following paragraph the focus will be to address how proper the setup within this thesis of natural experiment is and the three presented issues with the use of a natural experiment in the context of this thesis. As will be further discussed in section 4.3 the setup for the natural experiment in this thesis is the risk of being exposed to influenza infection during the time of in-utero for the individual at analysis.

First, the question if the strike of an influenza pandemic can be seen as random and unpredictable has to be discussed. As presented in section 2.6. according to C. Potter (2001) the timing of the outbreaks of an influenza pandemic has been very hard to predict, meaning that it can be assumed that outbreaks are unpredictable. Also, since the outbreak of the influenza pandemic in 1833 did not strike during the more normal period of the year, the winter, but instead in the spring (Patterson, K. (1985)), this outbreak was even more unpredictable than usual. Furthermore, it also stroke at a very high pace, meaning that individuals cannot plan their future around it, but there exist patterns of which individuals are more affected by it then others, though these differences is regarding the age of the individual and not the socioeconomic status, which would pose as a much bigger issue in that type of situation due to that women generally have children at a quite limited time in life.

Next, the three posed issues in Durlauf, S. and Blume, L. (2008) will be addressed. As for issue (1) of course the results found in this thesis can only be assumed to be valid for influenza outbreaks during fetal stage and no other types of bad early life conditions, but since influenza outbreaks still is happening the results have important interpretations. Furthermore, the validity of the results should at least to some extent be valid for all individuals born in similar life standards as the individuals tested in this thesis. Next, for issue (2) that seems to have been a general problem for similar studies because most of studies focusing on influenza outbreaks have used the 1918 Spanish flu outbreak, see section 2.6., as the natural experiment, but one important part of this thesis is to test the external validity of that natural experiment by using another outbreaks of influenza as the natural experiment, also since outbreaks of influenza happens on quite a regular basis testing for external validity should not pose as a big issue. As for the last issue (3) within this study there should be no problem since there is a well-established broad theory regarding the fetal origin hypothesis (Barker, D. (1990)) and its extensions see section 2.1. to 2.5. for a deeper presentation of the underling theory and related research.

4. Data

4.1. The origin of the data

The data that this thesis is based on originates from the database the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)). This database consists of longitudinal data regarding five parishes in western Scania, the southernmost county of Sweden (Öberg, S. (2014)). The five parishes are Kävlinge, Hög, Kågeröd, Sireköpinge and Halmstad. The database includes data from the middle of the 17th century up to the earliy 20th century, in regard of the public version used in this thesis, for the non-public version later data is also available (Quaranta, L. (2013)). As the data is of longitudinal structure the individuals included in the database are followed continuous over time and their demographic events are recorded continuously (Öberg, S. (2014)).

The different demographic records are collected though a number of different sources, for example the catechetical examination registers, tax registers and church books to be able to identify for example deaths, births, weddings, in- and out-migration and socioeconomic status (Öberg, S. (2014)).

The specific topography of Scania, in which the five included parishes belongs to, consist of fertile agricultural land and throughout the period of interest in this thesis the dominated jobs consisted of agricultural work, which also fits well with the five specific parishes (Öberg, S. (2014)). It is important to point out that these conditions cannot be seen as a generalization for 19th century Sweden, but instead for individuals living in an agricultural rural area during this time period (Quaranta, L. (2013)). In the first half of the 19th century the populations within the five parishes was more or less within the same numbers and not until the end of the 19th century there became a greater variation within the population numbers and the structure of the villages, were Kävlinge and Hög developed into small cities (Öberg, S. (2014)). This demographic change will possibly not have a big impact on this thesis, due to the time period that is analyzed.

In this thesis the analysis will be based on cohorts born between 1815 and 1840. The reasons for this specific time frame is at first the influenza pandemic, that is the basis for the natural experiment in this thesis, happened in 1833. As was presented in section 2.6. this was the strongest influenza pandemic during the outbreaks in the 1830s and also comparing to the other outbreaks that happened during this period the only one which did not coincide with high levels of outbreaks of other infectious diseases, like for example whooping cough or smallpox, see section 4.3. for more detail on the control of other outbreaks of diseases. Furthermore, since individuals are followed up until death or the age of seventy-five and the public version of the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) only ranges to 1910, to be able to follow at least most of the cohorts all the way, 1840 is the last year an individual can be born. This could pose as a problem due to the timing of the influenza pandemic, since it happens for individuals born in later end of the cohorts in the dataset. This could be a problem for example for the survival analysis if there exist a general increase in the survival for later cohorts in the dataset, which is quite likely due to the general mortality decline that took off in the first half of the 19th century. This could then lead to underestimation of the effects from the influenza on the survival. At last also there seems to occur a change in the conditions for infants in the middle of the 19th century within the Scanian Economic

Demographic Database, as the infant mortality rate suddenly drops and individuals born after this shift cannot be seen as having experienced equal conditions in early life compared to those born before (Quaranta, L. (2013)). As for the start cohort to be 1815, the main explanation is that at this point in time data for all the five parishes is included in the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)).

4.2. The structure of the data

The original structure of the Scanian Economic Demographic Database (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) is in the format of Intermediate Data Structure. To be able to properly use it within the analysis it is first transformed in STATA by a code written by Quaranta, L. (2012), which then restructures the data set into the standard version of longitudinal data with episodes. Apart from that also the socioeconomic status of the individual and the socioeconomic status of the father have to be restructured, since both these factors contain a lot of variables, due to different recordings from different sources. To restructure the socioeconomic status of the individual a code by Dribe, M. (2012) is used. Furthermore, to construct one variable of the socioeconomic status of the father, the code by Dribe, M. (2012) is reworked to work to fit the variables regarding the father instead. As stated before, the data consists of a lot of demographic characteristics of every individual, for example birth date, death date, data on marriage status, number of children and parish affiliation.

For this specific thesis the outcomes used are either on binary level, meaning that the variable takes on the value zero or one, or on factor level, meaning that the variable is on categorical level and are allowed to take on more values, but the values are only indicating a specific feature of the individual and are not actual values, but can be ranked. For example the socioeconomic status variables allows for individuals to belong to different socioeconomic classes, depending on the earnings and the occupation of the individual. The reasons to only use these types of simplistic variables are partly that the quality of data of this historical context does not provide the level of detail that is needed to use variables of continuous nature, but also that result from analysis on binary and factor level variables have clear and easy interpretations. Furthermore, even though the analysis is performed on these types of simplistic variables, every method used in the analysis of this thesis has the ability to also analyze variables of continuous nature.

Due to how the methods used in this thesis works individuals are allowed both to be in- and out-migrated, but the individuals are only allowed to be right censored, but not left censored. In general is right censored never a real problem, since the most important feature of an individual is to know when an individual is born and most methods regarding survival analysis are able to handle when the death of an individual is not known as long as it is known when the individual is not followed anymore.

The used part of the Scanian Economic Demographic Database (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) consist of 9 953 number of individuals born between 1815 and 1840. Of these individuals 5 217 is men and 4 736 is women. At last there are 3 067 men and 2 871 women in the data at the age of twenty and 1 256 men and 1 166 women in the data at the age of forty. There are a couple of explanations to why there are so much fewer individuals in the older ages. First of all individuals dies over the life course and for example, the life expectancy at birth for individuals born between 1815-1840 in Sweden was between around 42 to 46 years of age

(Mortality.org, (2015)). Furthermore, during this period of time the infant mortality rate was extremely high, so these two explanations could count for a substantial part of the lower numbers in older ages. Secondly, there could also exist a higher level of out-migration compared to in-migration within the parishes in the dataset. If this would be the major source to the lower numbers in the older age groups this could be a source of bias, due to that individuals that out-migrated might not be homogenous compared to the individuals who stayed and a selection of specific individuals could happen. These numbers are interesting because the analyses are divided into time intervals where the ages twenty and forty the starting points of the analyses, age twenty for the analysis of adult socioeconomic achievement and age forty for later life survival.

4.3. Variables used in the analysis

In this part of the thesis the variables used in the analysis will presented in terms of which they are, why they are used, what specific feature that has to be considered for the specific variable and how the variable is defined in the data.

As a start of this section in the thesis it should be presented that the analysis done in this thesis will be done for men and women separately. This is done since according to the theoretical assumptions that in-utero effects from influenza might act different for women and men, since there exit different fetal resistance towards infections which could affect the outcomes for men and women in different ways, see for example Bengtsson, T. and Helgertz, J. (2013).

As the main goal of this thesis is to study later life outcomes from the influenza pandemic in 1833 and try to evaluate the external validity of the much used Spanish flu pandemic in 1918, the first variable to present is how the influenza variable is handled in this thesis.

From the presented theory in section 2.6. it is known that during the period of March to May in 1833 Sweden was hit by a strong influenza pandemic with both high morbidity and relatively high mortality (Patterson, K. (1985)) and this pandemic outbreak is the basis for the analysis of this thesis, namely the treatment variable. As also presented in section 2.6. there was two other potential outbreaks of influenza pandemic during the 1830s, namely 1830-1831 and 1836-1837: There exist a couple main problems with these two outbreaks that makes them less useful compared to the outbreak of 1833. Firstly, both these two influenza outbreaks coincide with outbreaks of other infections, that also had effect on the infant mortality rates, which instead will be controlled for in the analyzes, more on this later in this part. Furthermore, when investigating the numbers of deaths within the five parishes by months during the period of 1815 to 1840 the general conclusion that can be drawn is that the influenza outbreak of 1833 is the only one that shows any increased number of deaths during its outbreak, compared to the other influenza pandemics. This goes in line with what is presented in Patterson, K. (1985) and Gluge, G. (1837), where both authors state that the influenza pandemic of 1833 was the strongest one. The numbers of deaths by month is presented in table 3 in section 5.1. Due to that almost all other comparable studies is done for the Spanish flu, the possible strongest influenza outbreak ever, it should be most convenient to use the strongest available outbreak as the treatment in this thesis. Since the theoretical approach that wants to be tested in this thesis is the "Fetal Origins Hypothesis" by Baker, D. (1990), see section 2.1. to 2.5. for details and background of this theory, the goal of this variable is to capture all individuals that was affected by the pandemic flu while being in the state of in-utero, which in this thesis, as in articles by for example Almond, D.

(2006) , Almond, D. and Mazumder, B. (2005) and Bengtsson, T. and Helgertz, J. (2013), is seen as a natural experiment, for further more detail on natural experiment see section 3.4. To properly design the variable for this natural experiment to act in the desired way, individuals born between March in 1833 to February 1834 are contributed the value one of this variable and all other individuals in the dataset the value zero. The choices of these dates are due to those individuals born from March and onwards are at least to some extent affected in-utero by the pandemic, but at different times during the fetal stage, and since the pandemic ended in May 1833 individuals born in February 1834 are nine months after the outbreak ended and those are the youngest individuals that can have been in-utero during the pandemic. Individuals born after February 1834 should have been conceived, if we assume that in general individuals are carried by their mothers at the normal of nine months, before the end of the pandemic. This way of setting up the influenza pandemic variable is in line with how it is done for the Spanish flu variables in for example Almond, D. (2006) and Almond, D. and Mazumder, B. (2005).

There are a couple of problems that should be discussed regarding the natural experiment of the influenza pandemic variable. First, as presented in section 3.4. the effects found from this variable can only be valid for individuals affected by an influenza outbreak during the time of in-utero and not seen as a generalization of bad in-utero conditions in general, since other factor also may cause bad in-utero conditions, like for example bad nutrition (Quaranta, L. (2013)). Furthermore, in general studies of natural experiment, and especially for studies of influenza outbreaks, suffer from the problem of external validity due to that almost every study so far have used the extreme 1918 Spanish flu as the natural experiment, but since this study focuses to use another outbreak and compare the results to studies of the Spanish flu, this study tries to put focus on the external validity of influenza outbreaks in a more general sense.

Another factor that could cause bias within the estimation of effects from the influenza pandemic is the fact that individuals within the dataset are allowed to be in-migrated into the five parishes at any point in time. This could cause a risk if these individuals did not experience the influenza pandemic outbreak at the same time, or even at all, as the individuals born within the parishes. Even though this problem may exist from a theoretical point of view, it should probably not pose as a big issue for this thesis. The reason here is that, firstly, the influenza pandemic was widespread over most of Europe and the level of the pandemic was high for most of the European countries and did hit most of Scandinavia mostly simultaneously, see section 2.6. Secondly, even though there was a quite high migration rates is Scania during the specific period in time, most of these migrants consisted of individuals born in parishes closely located to the five parishes in the dataset (Quaranta, L. (2013)), which was hit by the influenza pandemic at the same time. To summarize, even if there are a lot of individuals in the dataset that was born outside of the five parishes, they should not pose as a source of bias, since most of them consisted of migrants that were born in nearby parishes that should have been experiencing the influenza pandemic at the same time, only individuals more or less born in other countries or in the north of Sweden could to a bigger extent cause bias within the setup of this thesis. An alternative solution would be to restrict the sample in the analysis to individuals born within the five parishes, but since the dataset itself in not very big another problem would arise instead and that would be that the sample size would be too small and no significant effects could be expected to be found.

The second variable used in the second part of the analysis that will be addressed here is the variable that indicates when a specific individual dies. This variable is called death indicator in the dataset and takes the value of one at the time when the individual dies. This variable is only recorded for individuals that die within one of five parishes and if an individual dies elsewhere the death will not be recorded. Instead this individual will be seen as right-censored when he or she out-migrated from the parishes. Furthermore, an individual that dies after the age of seventy-five will not be registered as dead at any point in time for the analysis, but instead also right-censored at the age of seventy-five.

The next variable to be considered is the variable that controls for years with high infant mortality rate. This is done due to the extensive literature, see for example Andersson, F. (2014), Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and (Quaranta, L. (2013), that uses high infant mortality rate as an approximation to bad conditions in early life and find support for effects of later life survival from this variable. If this variable would not be used, even though the specific mechanism regarding the effects from this variable differs from the mechanism for the influenza exposure the results from the effects of being exposed to influenza outbreak in-utero could suffer from underestimation, since these individuals also may suffer from early life scarring. The control for high infant mortality rates is done in a similar way by Bengtsson, T. and Helgertz, J. (2013) in their study for effects from in-utero exposure to the Spanish flu. The variable for indicating if an individual was born during a year of high infant mortality rate is coded one if the individual was born during one of the years with high infant mortality rate this variable take the value of one and otherwise zero. The years that is used as the years of high infant mortality rate is taken from the study by Quaranta, L. (2013) and are 1816, 1821, 1826, 1831, 1832 and 1835. These years are directly calculated from the Scanian Economic Demographic Database and have been calculated by Quaranta, L. (2013). During these years it has been found that they coincide with a couple of different epidemical outbreaks, which consists of measles, smallpox and whooping cough, where whooping cough seems to have been the most common of these infections (Quaranta, L. (2013)).

The control variable for high infant mortality rate may partly suffer from some problematic causes. First, since the calculated values of the infant mortality rate are yearly averages, different individuals born during different parts of the year may have experienced the effect form the high infant mortality rate at different levels. Furthermore, there exist problems that could cause bias on the high infant mortality rate variable regarding individuals that was in-migrated and was not really experiencing high infant mortality rates in early life, in line with the same discussion as for the influenza pandemic variable previously. On the other hand, since this variable only works as a control variable in this thesis these problems could be seen as minor.

For this study, two more variables are considered, one as a control variable in line with the variable for high infant mortality rate and one that is seen both as an outcome, for the first part of the analysis and thereafter as a mediating factor within the analysis of survival. The main reason, at least partly, to present these two variables at the same time are that they are both regarding the socioeconomic status of individuals and are therefore closely related. These two variables are the socioeconomic status of the father, recorded when the individual is born, and the socioeconomic status in adult life for the individuals in the dataset.

As a starting point the coding of the two variables will be discussed since that procedure is similar for both variables and then the specific characteristics of both the variables will be presented separately. Both socioeconomic variables consists of a lot of different variables coded from different sources in the Scanian Economic Demographic Database and for the variable that indicates the socioeconomic status of the individual itself, Dribe, M. (2012) provides a code for STATA that sort all variables into one single variable. For the socioeconomic status of the father, a modification of the code by Dribe, M. (2012) is done to achieve the same type of single variable as for the socioeconomic status of the individual. For the socioeconomic status of the individual in adult life, the data from the variable is taken from the socioeconomic status of the family head, which in the normal situation would mean that for men it is his own socioeconomic status, but for women, if they are married it would be their husbands. The socioeconomic status has different possible ways to be defined and the code by Dribe, M. (2012) is designed so that it is possible to choose between some of these classifications. The one that is used in this thesis is called Social Power, SOCPO, by Van De Putte, B. and Miles, A. (2005). The SOCPO classification divide the socioeconomic status of the individual into five different categories, namely unskilled, semi-skilled, manual skilled, manual superskilled and nonmanual superskilled (Quaranta, L. (2013)). These categories works as a factor variable and each category are assigned a number from one to five, with one as the number for the group with the lowest socioeconomic status, namely unskilled. Furthermore, for the socioeconomic status of the father at the birth of the individual these categories are reassigned into a factor variable containing three different levels, unskilled (the same as for SOCPO), skilled (semi-skilled and manual skilled), middle- and upper-class (superskilled and nonmanual superskilled) and a variable controlling for indviduals where the socioeconomic status of the father at the birth is unknown. For the individual socioeconomic status in adult life the variable are instead transformed into a binary variable indicating if an individual is unskilled or not, namely that individuals with SOCPO unskilled has the variable one and all other individuals zero. For example Bengtsson, T. and Broström, G. (2009) use a similar approach in their study to categorize individuals to divide them into the groups landed and landless.

The main reason for using the socioeconomic status of the father at birth is that this might pose as a source for bias within the study due to that individuals from different socioeconomic families could experience different conditions in early life. One important factor here that could cause bias in the analysis of this thesis is the nutrition of the individual in early life, which has been found to have an impact on later life outcomes, see section 2.5. The nutrition might be closely related to the parents' socioeconomic status, since richer individuals should have been able to buy better food and especially in year where harvest was bad have better chances of still not starving and cause malnutrition of their children.

The socioeconomic status in adult life is used in two different ways in this thesis, first as the outcome variable. The reason for this part of the analysis is that the general findings form the studies made for the Spanish flu is that individuals experiencing that pandemic during fetal stages had worse outcomes for education and socioeconomic variables, see for example Almond, D. (2006), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010), and to be able to compare this thesis to those studies an analysis of the direct effect on the socioeconomic status in adult life is the most proper way to do that considering the dataset available for this thesis. Furthermore,

for the second part of the analysis in this thesis the socioeconomic status in adult life are considered to be a mediating effect of later life survival. This follows the study by Edvinsson, S. and Broström, G. (2012) where the authors consider socioeconomic status at age fifty as a mediating covariate for later life survival and find indirect effects from early life conditions mediated though socioeconomic status on later life survival. The same results are also achieved by Andersson, F. (2014).

5. Results

5.1. Descriptive statistics

Table 1 Weighted mean values for the individuals divided into time intervals

	Weighted mean value		Number of deaths
	Flu	Family SES	
Female 0-20	0.048	0.053	541
Female 20-40	0.048	0.095	95
Female 40-75	0.033	0.197	401
Male 0-20	0.017	0.122	576
Male 20-40	0.015	0.158	101
Male 40-75	0.033	0.177	360

As a first part of the presentation of the results for the analysis of the thesis a section that will discuss some descriptive statistics will be presented. The first statistics to be handled is the weighted mean values for the individuals in the dataset presented in table 1. These means have been weighted in regard to exposure time for each possible state of the variable and also relative to the frequencies (Broström, G. (2012)).

The first variable to shortly discuss is the variable called flu in table 1. This variable indicates the percentage of the population that was exposed in-utero by the 1833 influenza pandemic. What first can be seen is that in the earlier stages in life, a higher proportion of women compared to men were exposed in-utero and that for the different sexes the percentage increased for men and decreased for women. This could be an indication of a selection process even before birth and that more female fetuses survived the in-utero exposure and therefore had a higher percentage to start with. Furthermore, the fact that the percentage increases for men could also be explained by selection, which would mean that the male fetuses that was affected in-utero and survived was the strongest and that those then had a higher survival chance than the mean male population. As for the women the opposite explanation of scaring could be the reason behind the lower percentage in later life. In general the low percentage is expected since only individuals born during a short period compared to the period of all births in the dataset.

The other variable, in table 1 called family SES, which could have important implications, is the percentage of the population that belonged to the lowest socioeconomic class. Here, the percentage of individuals belonging to the lowest socioeconomic status increase with age. One explanation to this, which also could cause as a risk for bias in the following analysis, is that the increasing number of individuals belonging to the lowest socioeconomic class could be explained by out-migration. Here, that when individuals became richer they left the five parishes in the data and move for example to a bigger city. Another possible explanation is that at a certain age individuals would give their family farm to their children and would in the data be seen as landless, but their actual socioeconomic status would be the socioeconomic status of their children. Support for this problem has been found by Bengtsson, T. and Broström, G. (2009).

Table 2 Number of births by year and sex (Source: The Scanian Economic Demographic Database. Version 3.1)

Birth Year	No. of individuals	No. of men	No. of women	Sex Ratio
1815	347	175	172	1,02
1816	348	177	171	1,04
1817	368	199	169	1,18
1818	339	168	171	0,98
1819	354	186	168	1,11
1820	342	199	143	1,39
1821	370	214	156	1,37
1822	412	207	205	1,01
1823	423	213	210	1,01
1824	397	209	188	1,11
1825	444	251	193	1,30
1826	387	226	161	1,40
1827	367	171	196	0,87
1828	373	173	200	0,87
1829	394	187	207	0,90
1830	377	229	148	1,55
1831	365	184	181	1,02
1832	333	187	146	1,28
1833	438	211	227	0,93
1834	374	186	188	0,99
1835	424	231	193	1,20
1836	464	234	230	1,02
1837	384	197	187	1,05
1838	344	175	169	1,04
1839	372	197	175	1,13
1840	413	231	182	1,27
Total	9 953	5 217	4 736	1,10

To furthermore take on the discussion of the selection and scarring of the flu before birth the numbers of birth by year is presented in table 2. The most interesting numbers are the births in 1833. As can be seen in the table for this particular year the sex-ratio, normally 1.05 men per women, is a bit off. We see that more women than men was born during the year when the influenza pandemic outbreak happened. On the other hand the numbers are small and no conclusions can be drawn from this, but it could be seen as a weak indication of the effects.

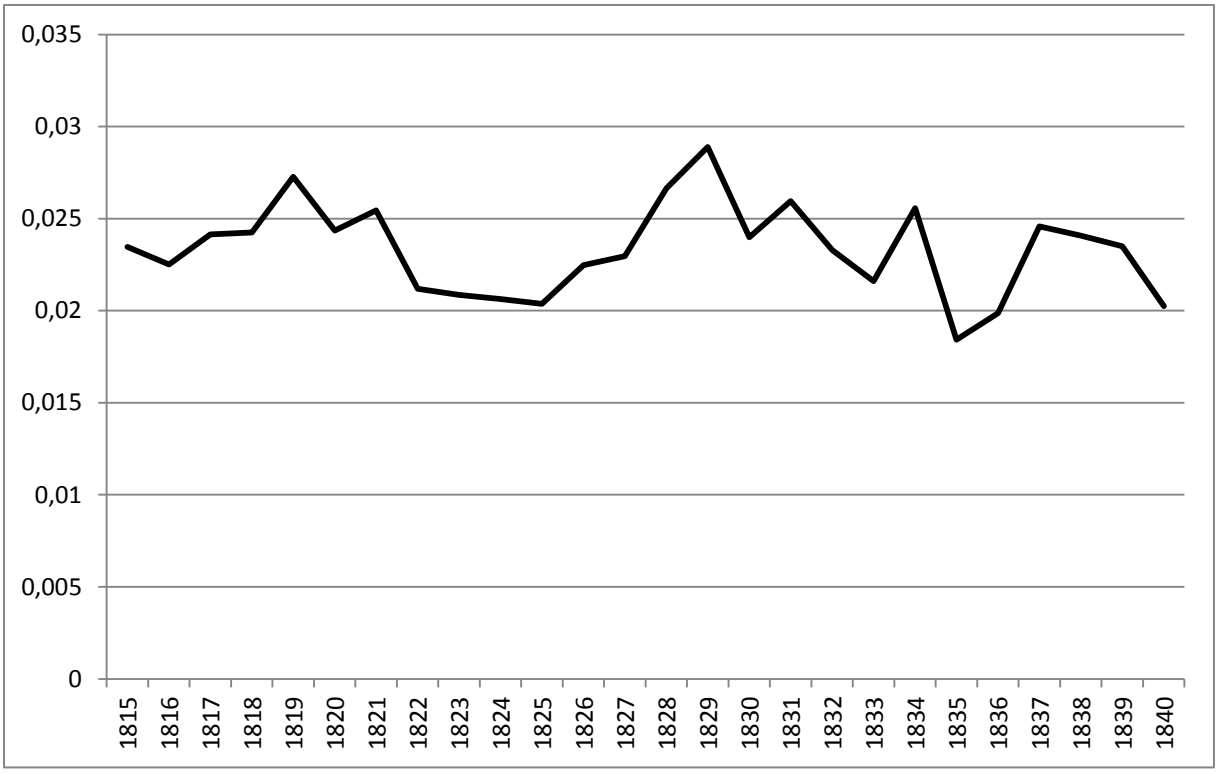

Figure 2 Crude death rates for Sweden 1815-1840 (Source: Statistiska Centralbyrån, (2015))

To try to investigate the mortality of the influenza pandemic of 1833 a little bit further the crude death rates for Sweden for the year 1815-1840 is presented in figure 2. From this figure no clear deviation from the general values can be seen for the year of the outbreak. It is important here to take into account that the influenza pandemic of 1833 did not have the extreme mortality as for example the Spanish flu, so a clear top might not be expected. Also the fact that a number of other outbreaks of infectious diseases, controlled for in the analysis see section 4.3. for details, happened during this period of time.

As a last part of this introduction to the analysis the numbers of deaths by months over the period 1815-1840 in the analysis is presented in table 3. As can be seen for the specific year of 1833 the numbers of deaths was highest during the period of March to May, which coincides with the period of the influenza pandemic outbreak. On the other hand, in line with the results seen in figure 2, the total deaths were not particularly high during the year of 1833. Here, instead the year of 1832 stands out as a year with high numbers of deaths and according to Quaranta, L. (2013) this was one of the years with high infant mortality rates and outbreaks of other infectious disease, for more detail see section 4.3. Furthermore, as can be seen in table 3, the year of 1831 shows in general high numbers of death both in the spring and in the autumn and winter. The extreme numbers in the winter do not coincide with the influenza pandemic outbreak that could pose as a problem for the results of this analysis, but according to Quaranta, L. (2013), 1831 was also a year of high infant mortality rates and outbreaks of other infectious disease, which is controlled for in the analysis. On the other hand the high numbers of deaths in April and May of 1831 might pose as the most problematic numbers for this thesis, since the influenza outbreak of 1831 happened during that period, for more detail see section 2.6. These individuals would then both risk of being exposed in-utero to the 1831 influenza and later experience high infant

mortality at birth and are therefore included in the group of individuals that are controlled for, born during periods of high infant mortality rate. One reason for this is that results from these individuals are more or less not interpretable since it is impossible to distinguish between if the effect they could show would be due to the influenza or the high infant mortality rate, since many studies including studies by Andersson, F. (2014), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2013) finds effects from years of high infant mortality rates.

Table 3 Deaths by month in the five parishes 1815-1840 (Source: The Scanian Economic Demographic Database. Version 3.1)

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1815	7	5	8	9	4	8	5	3	1	9	2	5
1816	3	7	11	12	5	2	4	9	6	7	8	6
1817	9	4	7	5	3	5	3	2	4	4	4	12
1818	10	12	9	4	7	6	4	3	4	3	4	6
1819	5	3	6	11	5	8	2	5	5	4	5	9
1820	8	4	6	4	6	2	8	4	2	2	4	11
1821	15	21	14	3	9	10	8	7	11	9	6	8
1822	6	5	11	5	9	3	5	4	3	5	9	10
1823	11	9	6	9	5	9	6	3	1	3	7	8
1824	11	4	6	6	4	5	2	3	4	3	5	8
1825	7	5	15	12	8	3	3	7	3	4	10	2
1826	7	9	7	6	9	7	5	9	8	8	14	9
1827	11	9	8	14	5	6	7	8	5	6	5	3
1828	5	6	11	11	6	5	8	2	8	3	11	3
1829	6	1	10	15	13	14	12	6	7	8	6	9
1830	9	10	17	14	7	9	5	6	7	6	2	10
1831	8	7	7	14	14	6	10	13	25	10	18	18
1832	8	18	26	18	14	7	12	6	8	3	8	8
1833	11	6	11	12	13	8	5	5	6	3	4	5
1834	8	9	7	9	9	4	3	7	3	9	5	4
1835	4	11	6	16	10	7	1	5	7	4	9	12
1836	6	11	20	11	13	8	3	4	4	0	3	5
1837	8	11	4	7	8	11	8	4	5	4	3	5
1838	12	7	11	12	12	9	8	2	5	8	6	10
1839	12	7	12	18	9	7	9	5	3	8	8	6
1840	6	9	14	13	9	4	0	6	5	8	5	8

5.3. Analysis of socioeconomic effects of early life exposure to influenza in adult life

The analysis in this thesis is divided into two parts; the first one is a model to analyze the effects from being affected by an influenza outbreak during fetal stage on the socioeconomic outcome for the individual in adult life and the second model is to analyze the direct and indirect effects on the health, here survival, in later adult life. Both these analyzes will be divided between the sexes with results first presented for women and then for men.

For this specific part of the thesis the focus will be on the direct effects on socioeconomic status in adult life for individuals affected by the influenza in-utero. All data management and all calculations of the analysis of this part are done in the program STATA. It is also important to point out that the level of significance that is chosen in the analysis is the 10% level. The main argument for the use of this level is the relatively small sample size in the data and that a p-value less than 0.10 generally is showing at least a good indication of significant relationship between variables.

The methodological approach that has been taken for this part of the analysis is the random effects logistic regression model for longitudinal panel data with a binary response variable, presented in section 3.2. The response variable for the analysis here is a dummy variable indicating if an individual is belonging to the lowest socioeconomic group or not. The variables used here to test the effect on the socioeconomic status are the variable indicating if the individual was in-utero during the 1833 influenza pandemic, a variable controlling for years of high infant mortality rate (in general controlling for outbreaks of other infectious diseases within the data) and a categorical variable that test for the effects of the father's socioeconomic status at birth of the individual (both controlling for early life environment and parental socioeconomic influences). For more detail of the model see section 3.1. For the analysis of the socioeconomic status in adult life all individuals in the ages twenty to seventy five is included in the dataset, but, as presented previously in this part, divided into one part for women and one part for men.

Results for women born between 1815-1840

Table 4 Effects on socioeconomic status from in-utero influenza exposure in adult life for women

Female age 20-75 born 1815-1840			
Covariate	Coefficient	Standard error	P-value
Flu	1.138	0.532	0.032
IMR high	-0.142	0.252	0.573
Father SES			
Unskilled	ref.		
Skilled	-1.587	0.529	0.003
Middle/Elite	-3.402	0.516	0.000
Unknown	-4.758	0.495	0.000
Constant	-3.134	0.243	0.000
Number observations	53 958		
Individuals	2 871	Wald chi2	187.73
Max. log Likelihood	-13 577.718	Overall p-value	0.000

The first category that is analyzed is the adult women in the dataset born between 1815 and 1840. The result from the random effects logistic regression on the socioeconomic status is presented in table 4. Here, significant result is found from the impact of have been exposed to the 1833 influenza pandemic on the odds of belonging to the lowest socioeconomic group as an adult, p-value equal to 0.032. The result indicates that women that were affected by the 1833 influenza pandemic during fetal stage had a higher risk of belonging to the lowest socioeconomic class. Furthermore, significant results are found for the impact from the socioeconomic status of the father at the time of birth. Here, it is found that having a father that belonged to any of the high socioeconomic classes reduce the risk of belonging to the lowest socioeconomic class in adult life. As can be seen by the coefficients in table 4, the higher the socioeconomic class of the father was the lower risk of belonging to the lowest socioeconomic class as an adult for a woman. The control for the effect of being born during a year of high infant mortality rate on the socioeconomic status in adult life is insignificant, with a p-value of 0.573. At last, as can be seen in table 4, the full regression is significant with p-value equal to less than 0.000. Furthermore, it can also be seen in table 4 that the analyzed group of women between age twenty to seventy five consisted of 2871 individuals.

Results for men born between 1815-1840

Table 5 Effects on socioeconomic status from in-utero influenza exposure in adult life for men

Male age 20-75 born 1815-1840			
Covariate	Coefficient	Standard error	P-value
Flu	1.163	0.630	0.065
IMR high	-0.385	0.256	0.132
Father SES			
Unskilled	ref.		
Skilled	-2.121	0.548	0.000
Middle/Elite	-3.323	0.547	0.000
Unknown	-5.838	0.520	0.000
Constant	-2.342	0.502	0.000
Number observations	54 663		
Individuals	3 067	Wald chi2	269.22
Max. log Likelihood	-12 739.679	Overall p-value	0.000

The next group that the analysis of the effect on the socioeconomic status is done for is the adult men in the dataset born between 1815 and 1840. The result from the random effects logistic regression on the socioeconomic status is presented in table 5. For the variable of the effect from being in the fetal stage during the outbreak of influenza the found effect is significant, with p-value 0.065, and similar for men compared to women when comparing the values of the coefficients. This result indicates that men affected by the 1833 influenza pandemic during fetal stage was experience higher risk of belonging to the lowest socioeconomic class. Next, there are significant results for the impact from the socioeconomic status of the father at the time of birth on the socioeconomic status in adult life for a man. Here, the results are that, similar as for women, individuals with fathers of higher socioeconomic status also faced a lower risk of belonging to the lowest socioeconomic class themselves. When comparing the results for men with the results for women, coefficients in table 4 and 5, they are in line with each other in the

way that the higher the socioeconomic class of the father the lower risk of belonging to the lowest socioeconomic class, but it can be seen that these effects are a little bit stronger for men than women. This could be a result of the fact that in a historical context the sons often inherited most of the assets from their parents. Furthermore, the effect of being born during a year of high infant mortality rate on the socioeconomic status in adult life is also insignificant for men, here with a p-value of 0.132. Lastly, the full regression model is significant with a p-value lower than 0.000. The total number on men in this analysis can be seen in table 5 to be 3067.

The results from the analysis of the effect from in-utero exposure of influenza on the socioeconomic status in adult life lies in line with results found in studies of the Spanish flu, like for example Almond, D. (2006), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010), though these studies often also have more detailed data and are able to analyze the underlying reason for this effect, namely that these studies finds direct effects from the in-utero exposure of the Spanish flu on the educational attainment. Since this variable not exists for the data used in this thesis the conclusions from this analysis is that the effects on the socioeconomic status is in line with the results for the Spanish flu, but the mechanism cannot be dealt with in more detail, like studying the educational attainment. Anyhow the expected finding points to a direct scarring effect on in-utero exposed individuals. These findings can be seen as support for the "Fetal origins hypothesis" and especially in connection to the mechanism presented in section 2.5. for the influenza exposure during fetal stage.

5.3. Analysis of direct and indirect effects from early life exposure of influenza on old age survival

For the second main part of the analysis in this thesis, the direct and indirect effects from the exposure of the influenza pandemic in-utero on the survival in adult life will be presented. In the same way as for the analysis on the socioeconomic status, the analysis will be carried out separately for men and women. The method that will be used for this analysis is the dynamic path analysis, presented in section 3.3. Here, this analysis will be carried out for one age group for each sex, namely the age group forty to seventy five. The main reason for this is that for younger age groups the gompertz distribution that the accelerated failure time model, used for the analysis of the direct effects on the survival rate fits best within analysis of older age mortality. Furthermore, as somewhere around the age of forty was the expected life at birth for the individuals within the dataset (Mortality.org, (2015)), this seems like an acceptable choice as the starting age on the analysis. To further check for the validity of the model, graphical comparison between predicted Cox proportional hazard model and the gompertz distributional model will be performed to show that the accelerated failure time model is appropriate to use for the analysis.

The results for this analysis will consist of the possible direct effects on survival for the accelerated failure time model and the indirect effects mediated though the socioeconomic status in adult life, as presented in chapter 3.1 and visualized in figure 1. The analysis of indirect effects will be presented as the linear regression of the influenza variable and the control variables for the socioeconomic status of the father and the years of high infant mortality rate on the socioeconomic status in adult life of the individual and after that the indirect effects will be calculated as presented in section 3.3. Furthermore, the mean differences in the baseline life expectancy, in this analysis the life expectancy from the age of forty for individuals that achieved that age, that is found to have the significant effects will be calculated.

In the same manner as for the first part of the analysis the data managements has been performed in STATA, but for this second part of the analysis the calculations of the accelerated failure time models has been done in the statistical program R with the package "eha" that has been written by Broström, G. (2014) and is well described in Broström, G. (2012). Also, in the same manner as for the first part of the analysis the significant level that is chosen in this part of the analysis is the 10% level, with the same motivation as presented before, in section 5.2.

Results for women born between 1815-1840

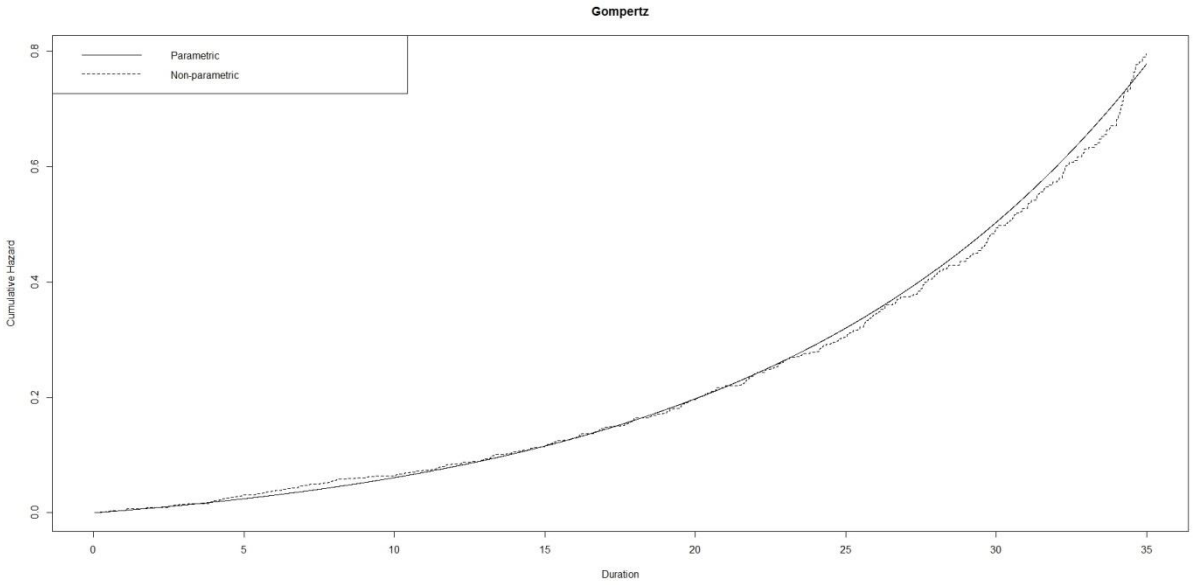

Figure 3 Graphical test of the validity of the accelerated failure time model for women

The first model to present and discuss is the model of direct and indirect effects on the survival in later adult life for women in age forty to seventy five. The first step is to check the underling distribution of the accelerated failure time model for the analysis of the direct effects. The reason to do this is to see if the model is correctly specified especially if the defined age group is a good choice. As can be seen in figure 3, the model, indicated by the dotted line, coincide well with the cox proportional hazard model. This is a good indication of that the model is well specified and that the results found in this analysis can be seen as a good representation of the effects.

Table 6 Effect from in-utero exposure of influenza on survival for women in age 40 to 75

Female age 40-75 born 1815-1840			
Covariate	Coefficient	Standard error	Wald p-value
Flu	-0.140	0.085	0.099
IMR high	-0.033	0.041	0.412
Father SES			
Unskilled	ref.		
Skilled	0.054	0.060	0.366
Middle/Elite	0.070	0.062	0.260
Unknown	0.019	0.063	0.766
Family SES	-0.121	0.040	0.002
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.522	0.086	0.000
log(shape)	-2.993	0.191	0.000
Baseline expected life	32.3		
Events	401	Total time at risk	24 175
Max. log Likelihood	-1 919.6	Overall p-value	0.026

Table 7 Effect from in-utero exposure of influenza on socioeconomic status for women in age 40 to 75

The effect of Flu, IMR at birth and Father on adult SES for female age 40-75			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.362	0.037	0.000
Flu	0.085	0.054	0.119
IMR high	0.001	0.026	0.959
Father SES			
Unskilled	ref.		
Skilled	-0.092	0.041	0.025
Middle/Elite	-0.169	0.041	0.000
Unknown	-0.119	0.041	0.003

The results for the direct effects on survival in the ages forty to seventy five for women are presented in table 6. Here the variable indicating if a woman did experience the influenza pandemic during the fetal stage is showing significance at the 10% level, with a p-value of 0.099. Furthermore, it can also be seen that the socioeconomic status of the women in adult life is significant. This result opens up for the existence of indirect effects on the survival mediated though the socioeconomic status in adult life for the woman. For the rest of the variables no significant direct effect is found on the survival for women. At last the full model, as expected from the result of the graphical analysis of the validity of the model, is significant with a p-value of 0.026.

The sign of the coefficient for the variable that is indicating if the individual was experiencing the 1833 influenza pandemic in-utero indicates that there is a negative effect on the chance of survival in the ages forty to seventy five for those women that was affected in-utero. When calculating the direct impact of this effect on the baseline life expectancy, the women that experienced the influenza pandemic in-utero had a life expectancy of 28.1 years compared to the baseline life expectancy of 32.3. If this is then calculated into months this means a loss of 50.4 months for the women affected in-utero by the influenza pandemic. This result is an indication of a latent effect on the survival in later life from an influenza pandemic, well in line with the "Fetal of Origins hypothesis" (Barker, D. (1990)), most often only found for cognitive ability and socioeconomic status when tested within a setup of influenza pandemics, like for example Almond, D. (2006). On the other hand latent effect on the survival in later life from infections in early life stages, including in-utero, is found by for example Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2014).

Furthermore, it can be seen that the coefficient for the socioeconomic status indicates that the poorest women in the dataset, belonging to the socioeconomic group of SOCPO 1, has the lowest survival chances in the ages forty to seventy five. Here, the life expectancy after age forty for the women in the lowest socioeconomic class is 28.6 years compared to the baseline life expectancy of 32.3. This is then calculated into months which mean a loss of 44.4 months for the women. The next focus in this analysis will be on the analysis of the indirect effects on the survival for women in the ages forty to seventy five.

As can be seen in the description of the method of dynamic path analysis in section 3.3. the next step is to analyze the linear effects of the covariates that could have an indirect effect on the survival on the mediating covariate, here the socioeconomic status in adult life. As can be seen in section 3.1. the theoretical indirect effects that are assumed to be tested in this model are the indirect effect from the in-utero exposure to influenza pandemic and the socioeconomic status of the father in early life of the individual mediated though the socioeconomic status in adult life. This analysis is in line with the first analysis performed in this thesis in section 5.2., but the difference is that to be able to calculate the indirect effects in regard of the accelerated failure time model, this time the effects on the socioeconomic status has to be done within the setup of a longitudinal linear regression model. Furthermore, compared to the analysis in section 5.2., which is done for the ages twenty to seventy five, here the analyzed group is instead the same as for the analysis of the direct effect on the survival, namely the ages forty to seventy five. The results for the linear regression of the socioeconomic status in the ages forty to seventy five is presented in table 7. From this model the found significant effect on the 10% level is the effect from the socioeconomic status of the father at birth of the individual. Here, it is found that if the father did not belong to the lowest socioeconomic class then the chance that the woman would belong to the lowest socioeconomic class in adult life was lower. This is in line with what was found in the analysis in section 5.2. On the other hand no significant results from the in-utero exposure of the influenza pandemic on socioeconomic status in the ages forty to seventy five is found here, as had been done for the ages twenty to seventy five in section 5.2., but as can be seen in table 7, the p-value of the variable for the influenza exposure was close to significant, p-value equal to 0.119 and the coefficient was following the results from the analysis in section 5.2. so the non-significance found here could be due to the lower number of women included in the analysis and even though the indirect effect is not found in this analysis the existence of an indirect effect on the survival mediated though the socioeconomic status cannot be out ruled.

To calculate the indirect effect on survival from the socioeconomic status of the father, the coefficient for individuals with a father belonging to the socioeconomic group of skilled individuals becomes 0.011, which is equal to a life expectancy of 32.7 compare to the baseline life expectancy of 32.3, namely an increase of almost 5 months. Furthermore, for women where the father belonged to the highest socioeconomic classes the coefficient becomes 0.02 and that gives a life expectancy after age forty of 33, which is equal to an increase of little over 8 months.

As there was found no significant indirect effect from the in-utero exposure of influenza mediated though the socioeconomic status, the total effect from the in-utero exposure of influenza is the same as the direct effect, namely a substantial decrease of 50.4 months in the life expectancy after the age forty.

Results for men born between 1815-1840

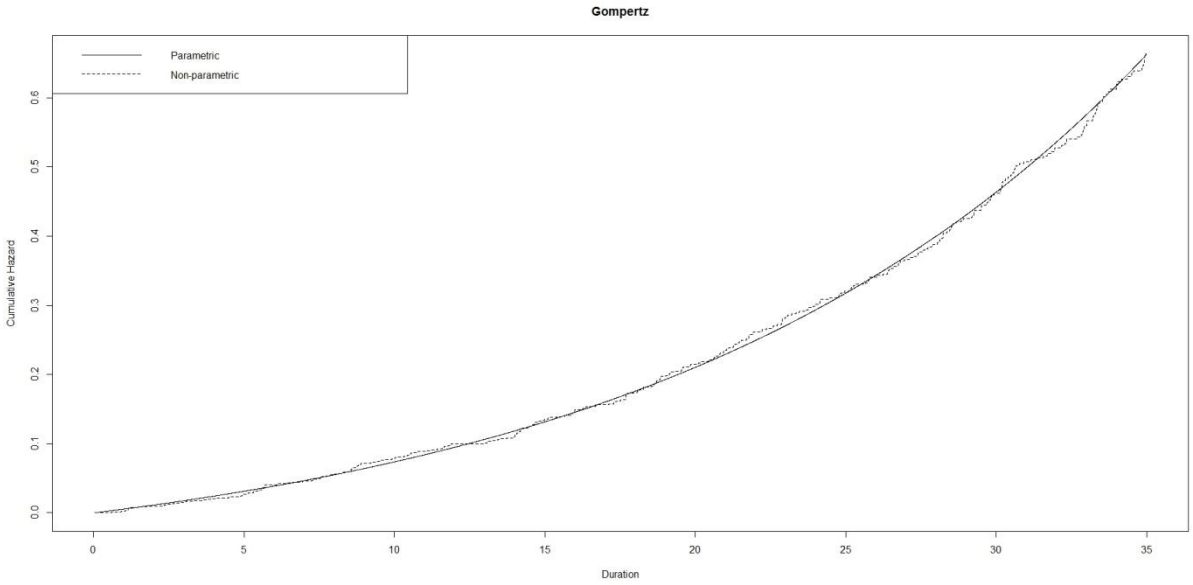

Figure 4 Graphical test of the validity of the accelerated failure time model for men

In the same way as for women the first step in the analysis of the survival for men in the ages forty to seventy five is to graphically check the underlying gompertz distribution of the accelerate failure time model in comparison to the cox proportional hazard function. As can be seen in figure 4, the model seems to fit well, more or less as the model for the women, which is a good indication that the accelerate failure time model is correctly specified and that the age group that is analyzed is a good choice in this analysis.

Table 8 Effect from in-utero exposure of influenza on survival for men in age 40 to 75

Male age 40-75 born 1815-1840			
Covariate	Coefficient	Standard error	Wald p-value
Flu	0.099	0.162	0.540
IMR high	-0.092	0.050	0.063
Father SES			
Unskilled	ref.		
Skilled	0.038	0.076	0.621
Middle/Elite	0.046	0.078	0.554
Unknown	0.094	0.084	0.260
Family SES	-0.180	0.052	0.000
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.819	0.114	0.000
log(shape)	-2.406	0.202	0.000
Baseline expected life	35.1		
Events	360	Total time at risk	24 260
Max. log Likelihood	-1 800.7	Overall p-value	0.008

Table 9 Effect from in-utero exposure of influenza on socioeconomic status for men in age 40 to 75

The effect of Flu, IMR at birth and Father SES on adult SES for male age 40-75			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.386	0.035	0.000
Flu	0.148	0.055	0.008
IMR high	-0.018	0.024	0.450
Father SES			
Unskilled	ref.		
Skilled	-0.129	0.039	0.001
Middle/Elite	-0.160	0.039	0.000
Unknown	-0.188	0.039	0.000

The results from the analysis of the direct effects on the survival for men in the age group forty to seventy five is presented in table 8. Here, the variable that is indicating the direct effect for men that has been affected by the 1833 influenza pandemic in-utero is insignificant, p-value 0.540. On the other hand the variable that is controlling for being born during a year of high infant mortality rate is significant at the 10% level, with p-value 0.063. Even though this thesis do not specifically focuses on the effects from high infant mortality rates, which is often used as an indication of bad early life conditions, it is worth pointing out that this effect is not unexpected, since other studies that focuses on the latent effects on survival from bad early life environment has found similar effects, see for example Andersson, F. (2014), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2014). These studies contribute this effect to be a support for the "Fetal of Origins hypothesis" (Barker, D. (1990)). For the specific effect from the variable controlling for bad early life conditions for men in this study the found expected life after age forty is 32 years compared to 35.1 for the baseline life expectancy after the age of forty. This can be recalculated to a decrease of around 37 months for individuals born during years of high infant mortality rate.

Moving on to the other significant variable for the accelerated failure time model for men in the ages forty to seventy five, presented in table 8, the variable indicating if a man is belonging to the lowest socioeconomic class or not has a significant effect on the survival in later life, p-value equal to or less than 0.000. The life expectancy for individuals belonging to the lowest socioeconomic group is then calculated to be 29.3 years compared to the 35.1 years for the baseline life expectancy. This substantial lower life expectancy represents a decrease of almost 70 months. For the variable that tests for the effects from the socioeconomic status of the father, no significant direct effects on the survival in later life is found.

The next step in this analysis of the survival in later life for men is to analyze the indirect effects that is mediated though the individual socioeconomic status in adult life, as can be seen in the model presented in section 3.1. Here, there are two specific indirect effects that is of most interest for this thesis, namely the indirect effects from the in-utero exposure to influenza pandemic and the socioeconomic status of the father in early life. As already defined in the analysis for the women, these effects are closely related to the analysis in section 5.2., but with a different model specification due to how the dynamic path analysis is defined, for a closer explanation see section 3.3. The results for the linear regression on the socioeconomic status in

the ages forty to seventy five is presented in table 9. Here, significant effect on the 10% level is found for the variable that is indicating if the individual had experienced the influenza pandemic in-utero or not. Furthermore, the effects from the socioeconomic status of the father at birth of the individual are also significant on the socioeconomic status in adult life for men. Both these results is in line with the results found in section 5.2. and also as for women in the previous part that if the father did not belong to the lowest socioeconomic class then the chance that the woman would belong to the lowest socioeconomic class in adult life was lower. Here, the p-value for the influenza variable was 0.008 and the p-values for the two interesting categories of the variable regarding the socioeconomic status of the father were 0.001 and below 0.000.

The indirect effect that is of most interest here is the indirect effect from the influenza variable on the survival, here the calculation gives a coefficient of -0.027. This coefficient gives a life expectancy of 34.2 years after the age forty. This is compared to the baseline life expectancy of 35.1 a decrease of about 11 months. Furthermore, there can also be found, in the same manner as for women, indirect effects from the socioeconomic status of the father on the survival mediated though the socioeconomic status in adult life for the men in the dataset. Here, the coefficient for a father belonging to the socioeconomic group of skilled individuals becomes 0.023 and that can be calculated to a life expectancy of 35.9 compared to the baseline life expectancy of 35.1, which is an increase of almost 10 months. Furthermore, when the father belonged to the highest socioeconomic classes a coefficient of 0.029 is found. This gives a life expectancy after age forty of 36.1 years, which is equal to an increase of 12 months in caparison of the baseline life expectancy.

For the study of the men in the ages forty two seventy five the found result in this study is an indirect effect on the survival for individuals affected by the influenza during the fetal stage. This effect is in line with the pre assumed effect on survival that is presented in section 3.1. and could furthermore be seen as support for the "Fetal of Origins hypothesis" (Barker, D. (1990)) in regard of the outbreaks of influenza. It is also worth pointing out that the negative effects on survival in adult life from being affected by influenza in-utero has not been found in previous studies, but authors like for example Almond, D. and Mazumder, B. (2005) and Garthwaite, C. (2008) has found effects both on specific health issues in later life and also a general tendency towards worse self-reported health in later life for individuals that was in-utero affected by the Spanish flu. This, together with the similar findings in this thesis compared to what other authors have found for the socioeconomic penalty of the influenza should be seen as strong indications both that the mechanism, presented in section 3.1., at least to some extent, is a good representation of the life-course risks individuals faces if affected by influenza in-utero, but also as support of the "Fetal of Origins hypothesis".

6. Conclusion

The main focus of this thesis has been to investigate the possible impact that being exposed to an influenza pandemic during the fetal stage on different outcomes for individuals during the life course. Furthermore, this has been done with a theoretical starting point of the "Fetal of Origins of Adult Disease" (Barker, D. (1990)) and the extended literature of the possible consequences from fetal influenza exposure. Furthermore, to be able to study the effects in later life on a broader spectrum the "Fetal of Origins hypothesis" has then been more generalized and the setup of a pathway model has been adopted, where the "Fetal of Origins hypothesis" is included as a starting point. The reason that makes the pathway model a good application within studies of life course effects is that it allows both for direct effects, in line with the "Fetal of Origins hypothesis", but also indirect effects that can be mediated through different covariates.

The analysis in this thesis is divided into three sections, where the first part focuses on descriptive statistics regarding the dataset and some more general mortality statistics that could have possible impact on the setup of the later analysis. From discussion regarding the presented weighted means of the individuals within the dataset in different age groups indications of a selection process even before birth is found. The indication is that more women fetuses survived the in-utero exposure, due to the sex-ratio of the children that was exposed in-utero, although a weak indication since other years also had skewed sex-ratio and could be an effect of a small sample size. Furthermore, as a consequence of that it can also be seen that of the total individuals in the youngest age group there were a higher percentage of women. Furthermore, the percentage of men increases over time, which could be explained by selection (Bengtsson, T. and Mineau G. (2009) and (Nelson, R. E. (2010)), which would mean that the male fetuses that was affected in-utero and survived was the strongest and that those then had a higher survival chance than the mean male population. This should also be closely connected to the skewed sex-ratio. As for women the percentage instead goes down in older age groups. This could be explained by scarring if the women survived the in-utero stage, but then instead became weaker in later life.

Furthermore, it is found from the weighted means that the proportion of individuals belonging to the lowest socioeconomic class increased in older age groups, with two possible explanations presented here. Either this is explained by out-migration of the richer individuals or the more probable explanation to this is that at a certain age individuals would give their family farm to their children and would in the data be seen as landless, but their actual socioeconomic status would be the socioeconomic status of their children, for which support also has been provided by Bengtsson, T. and Broström, G. (2009).

The next step in the analysis of the descriptive statistics handles the possible problems with the use of the influenza pandemic in 1833. In this part some of the most important implications to highlight are that for the specific year of 1833 the numbers of deaths was highest during the period of March to May, which coincides with the period of the influenza pandemic outbreak of 1833 and that during the other period of probable pandemic influenza outbreak in 1831 that the highest numbers of death both was in the autumn and winter, which did not coincide with the outbreak of the influenza pandemic that year, but instead with outbreaks of other infectious diseases, as presented by Quaranta, L. (2013). As discussed in section 3.4. a pandemic outbreak of influenza should fulfill the criteria for a natural experiment and therefore produce causal relationships. It is worth pointing out that the outbreak of the influenza in 1833 is not to a fully

extent possible to validate as pandemic and the causality of the results has to be taken with a bit of caution. On the other hand even if the influenza outbreak of 1833 did not hit the five parishes in the used dataset as hard as some other parts of Europe, it is possible that this outcome could only be true for the mortality of the pandemic and in that situation all parameters for a natural experiment would be well fulfilled. Most of the findings points to that at least an outbreak happened and the results should therefore have interesting interpretations.

The second part of the analysis focuses on the effects from in-utero exposure to the 1833 influenza outbreak on the socioeconomic outcomes in adult life for women and men separately. This is done with the approach of a logistic regression model and the socioeconomic status of the father at birth and years with high infant mortality rates and other outbreaks of infectious diseases are controlled for. As the results, both for men and women, significantly indicates that individuals that was exposed during fetal stage had a direct negative effect on the socioeconomic status in adult life. This result is in line with results found in studies of the Spanish flu, like for example Almond, D. (2006), Bengtsson, T. and Helgertz, J. (2013) and Nelson, R. E. (2010) and which also has been addressed by the explanation of the mechanisms of fetal problem in regard with if the mother would have the flu, for detail see section 2.5. This result could furthermore also be seen as an indication to that the use of the influenza outbreak of 1833 are a valid setup for the analysis. Since variables measuring the cognitive ability of the educational attainment not exists for the dataset used in this thesis the conclusions from this analysis is that the effects on the socioeconomic status is in line with the results for the Spanish flu, but the mechanism cannot be dealt with in more detail. Anyhow the expected finding points to a direct scarring effect on in-utero exposed individuals.

The third part of the analysis deals with the full pathway model, presented in section 3.1. Here, the effects from the exposure in-utero from the influenza pandemic in 1833 is tested on the survival in adult life, specifically the ages forty to seventy five. This effect is tested both as a direct effect and as an indirect effect mediated though the socioeconomic status in adult life. The choice of mediator of indirect effect is partly motivated by the previous findings and theoretical considerations of the direct effect on socioeconomic status from fetal influenza exposure. Furthermore, since the findings from the analysis of the direct effects on the socioeconomic status in second stage of the analysis in this thesis also highly support the theory of socioeconomic penalty form fetal stress, the choice of socioeconomic status as the mediating covariate seems appropriate.

From the analysis of the pathway model for women it is found a direct negative effect on survival from being exposed to the influenza pandemic in-utero. Here, the effect can be calculated to a loss of 50.4 months in life expectancy from the age of forty an onward compared to individuals not effected in-utero by the influenza outbreak. This result can be seen as an indication of a latent effect on the survival in later life from an influenza pandemic, well in line with the "Fetal of Origins hypothesis" (Barker, D. (1990)), in general only found for cognitive ability and socioeconomic status when tested within a setup of influenza pandemics. The few studies that has tried to evaluate some kind of direct penalty on the survival or mortality previously has only gotten inconclusive results, like for example Cohen, A., Tillinghast, J. and Canudas-Romo, V. (2010) and Fletcher, J. (2014) that has tested for mortality effect from the Spanish flu. On the other hand for example the study by Fletcher, J. (2014) do not specifically focuses on later life

mortality or survival, which is the period as for previous studies related to the "Fetal Origins Hypothesis" and in line with the original hypothesis, though the setup of infant mortality rate as an approximation for bad early life conditions, where the effects has been found, see for example studies by Andersson, F. (2014), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2014). This could also be related to the pervious discussion regarding the descriptive statistics for women where there also was found minor support for the scarring of women in later life and also the connection, although a weak indication, to the skewed sex-ratio is worth to take into account here.

For the same analysis performed for men, no significant direct effect on the survival for individuals affected in-utero by the influenza pandemic in 1833 is found. On the other hand, compared to the women, here a negative indirect effect mediated through the socioeconomic status in adult life on the survival in the age forty to seventy five is found. This effect represents loss of 11 months compared to the baseline life expectancy. The mechanism here should be seen as, men that was exposed in-utero by the influenza pandemic had, as also seen in the analysis of the direct effect on socioeconomic status, a decrease in their working ability which then lead to a lower socioeconomic status in adult life. This in the end then led to a lower survival chance in later life. This mechanism would have been expected to find for women too, since from the analysis of the direct effect on the socioeconomic status the effect from being exposed to influenza was found and that from the pathway model it was found that the lowest socioeconomic group had lower survival rates even for women. One simple possible explanation to this could be that the sample size in this thesis is a bit too small and that with bigger dataset this effect could be found for women too, since if the coefficient in the analysis of the pathway model would have been significant for the effect from influenza on the socioeconomic status then the same effect as for men would have been found. Also the p-value was nearly significant at the 10% level, with p-value of 0.119. All these facts indicate a too small sample. Furthermore, the finding of the indirect effect is in line with the "Fetal of Origins hypothesis", and especially in regard of what has been found from the previous studies of in-utero influenza exposure.

This thesis contributes to the literature of the "Fetal of Origins hypothesis" and especially the part focusing on influenza exposure in-utero. Firstly, this thesis deals with a completely different influenza outbreak, compared to previous studies, which also is set in a different time period compared to the other literature. Even so, the scarring effect on the socioeconomic capital that both the theory predicts and that has been found by the researchers with other setups is also found in this thesis. Furthermore, this thesis finds support for negative effects on the survival in later life for individuals exposed to influenza in-utero. This has not been found in other studies dealing with the Spanish flu, but increases of specific diseases in adult later life have been found. In a broader spectra the specific study of in-utero influenza exposure is important due to the continues outbreaks of the disease, and even if the outbreaks seldom is of pandemic proportions it still puts the fact that pregnant women faces a risk during influenza periods and probably should highly consider vaccination and in general be extra cautious during influenza periods. Furthermore, if a pregnant women is infected the knowledge of the risk of damage could at least to some extent maybe be avoided. At last it is worth pointing out that even though the method the dynamic path analysis used for the analysis of the pathway model is not unique for this thesis a small rework of it has been done in this thesis and only a few other studies have used similar

setups of the method before within a demographic context and the possibility of measuring indirect effects straightforward poses for great improvements of analyzing pathway models.

References

Aalen, O., Borgan, Ø. and Gjessing, S. (2008). *Survival and event history analysis*. 1st ed. New York: Springer.

Almond, D. (2006). Is the 1918 Influenza pandemic over? Long-term effects of in utero Influenza exposure in the post-1940 US population. *Journal of Political Economy*, 114(4), pp.672--712.

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives*, pp.153--172.

Almond, D., Edlund, L. and Palme, M. (2009). Chernobyl's subclinical legacy: prenatal exposure to radioactive fallout and school outcomes in Sweden. *The Quarterly Journal of Economics*, 124(4), pp.1729--1772.

Almond, D., Edlund, L., Li, H., and Zhang, J. (2010). "Long-Term Effects of the 1959–61 China Famine: Mainland China and Hong Kong." In *The Economic Consequences of Demographic Change in East Asia,* NBER-EASE Vol. 19, ed. T. Ito and A. Rose, 321--50. Chicago: University of Chicago Press.

Almond, D. and Mazumder, B. (2005). The 1918 Influenza Pandemic and Subsequent Health Outcomes: An Analysis of SIPP Data. *American Economic Review*, 95(2), pp.258--262.

Andersson, F. (2014). The effect of conditions in early life on health in adult life: A study of 19th century Scania. *Master Thesis Lund University*.

Barker, D. (1990). The fetal and infant origins of adult disease: The womb may be more important than the home. *BMJ: British Medical Journal*, 301(6761), pp.1111.

Barker, D. (1995). Fetal origins of coronary heart disease. *BMJ: British Medical Journal*, 311(6998), pp.171--174.

Barker, D. and Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. *The Lancet*, 327(8489), pp.1077--1081.

Bengtsson, T. (2004) Mortality and social class in four Scanian parishes, 1766–1865 T. Bengtsson, C. Campbell, J.Z. Lee (Eds.) *et al.*, *Life under pressure. Mortality and living standards in Europe and Asia, 1700–1900*, MIT Press, Cambridge, MA (2004), pp. 135--172

Bengtsson, T. and Broström, G. (2009). Do conditions in early life affect old-age mortality directly and indirectly? Evidence from 19th-century rural Sweden. *Social science & medicine*, 68(9), pp.1583--1590.

Bengtsson, T., Dribe, M. and Svensson, P. (2012). *The Scanian Economic Demographic Database. Version 3.1* (Machine-readable database). Lund: Lund University, Centre for Economic Demography.

Bengtsson, T. and Helgertz, J. (2013). Long-term income effects from early-life exposure to the 1918/1919 influenza pandemic: the case of southern Sweden. *Seminar paper 103, Lund University, Department of Economic History.* Paper presented in Lund in October 2013.

Bengtsson, T. and Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760-1894. *Population Studies*, 54(3), pp.263--277.

Bengtsson, T. and Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766--1894. *International Journal of Epidemiology*, 32(2), pp.286--294.

Bengtsson, T. and Mineau G. (2009). Introduction: Early-life effects on socio-economic performance and mortlity in later life: A full life-course approach using contemporary and historical sources. *Social science & medicine*, 68(9), pp.1561--1564.

Brandt, L., Siow, A. and Vogel, C. (2008). Large Shocks and Small Changes in the Marriage Market for Famine Born Cohorts in China. *Working Paper University of Toronto, Department of Economics*.

Broström, G. and Edvinsson, S. (2012). A parametric model for old age mortality in mediation analysis. Paper presented at the IUSSP 2013 Conference in Busan, Korea.

Broström, G. (2012). *Event history analysis with R*. 1st ed. Boca Raton, Fla.: CRC Press.

Broström, G. (2014). eha: Event History Analysis. R package version 2.4-1

Calkins, K. and Devaskar, S. (2011). Fetal origins of adult disease. *Current problems in pediatric and adolescent health care*, 41(6), pp.158--176.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1), pp.225--238.

Cohen, A., Tillinghast, J. and Canudas-Romo, V. (2010). No consistent effects of prenatal or neonatal exposure to Spanish flu on late-life mortality in 24 developed countries. *DemRes*, 22, pp.579--634.

Collier, R. (1974). *The plague of the Spanish lady*. New York: Atheneum.

Currie, J. and Hyson, R. (1999). Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight. *American Economic Review*, 89(2), pp.245--250.

Dribe, M. (2012). Socal class in SEDD. *http://extract.sedd.ed.lu.se/ExtractionFileList.aspx*. Lund University, Center for Economic Demography.

Durlauf, S. and Blume, L. (2008). *The new Palgrave dictionary of economics*. Basingstoke, Hampshire: Palgrave Macmillan.

Edvinsson, S. and Broström, G. (2012). The effect of early-life and mid-life factors on old age mortality. Paper presented at the IUSSP 2013 Conference in Busan, Korea.

Edwards, M. (2007). Hyperthermia in utero due to maternal influenza is an environmental risk factor for schizophrenia. *Congenital Anomalies*, 47(3), pp.84--89.

Finch, C. and Crimmins, E. (2004). Inflammatory Exposure and Historical Changes in Human Life-Spans. *Science*, 305(5691), pp.1736--1739.

Fletcher, J. (2014). Examining the Long Term Mortality Effects of Early Health Shocks. *US Census Bureau Center for Economic Studies Paper No. CES-WP-14-19*

Fogel, R. (2005). The Escape from Hunger and Premature Death, 1700-2100. *Population*, (3), pp.367--370.

Forsdahl, A. (1977). Are poor living conditions in childhood and adolescence an important risk factor for arteriosclerotic heart disease?. *Journal of Epidemiology & Community Health*, 31(2), pp.91--95.

Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006). Dynamic path analysis—a new approach to analyzing time-dependent covariates. *Lifetime data analysis*, 12(2), pp.143--167.

Garthwaite, C. (2008). The Effect of In-Utero Conditions on Long Term Health: Evidence from the 1918 Spanish Flu Pandemic. *Working Paper University of Maryland, Department of Economics*

Gluge, G. (1837). *Die Influenza oder Grippe*. Minden: Essmann.

Kelly, E. (2011). The Scourge of Asian Flu: In utero Exposure to Pandemic Influenza and the Development of a Cohort of British Children. *Journal of Human Resources*, 46(4), pp.669--694.

Kermack, W., McKendrick, A. and McKinlay, P. (1934). Death-Rates in Great Britain and Sweden: Expression of Specific Mortality Rates as Products of Two Factors, and some Consequences thereof. *Journal of Hygiene*, 34(04), p.433.

Lange, T. and Hansen, J. (2011). Direct and indirect effects in a survival context. *Epidemiology*, 22(4), pp.575--581.

Lee, R.D. (2003) "The Demographic Transition: Three Centuries of Fundamental Change". *Journal of Economic Perspectives.* , 17(4), pp. 167--190.

Lenz, W. and Knapp, K. (1962). Die Thalidomid-Embryopathie. *Dtsch med Wochenschr*, 87(24), pp.1232--1242.

Maddala, G. (1987). Limited Dependent Variable Models Using Panel Data. *The Journal of Human Resources*, 22(3), pp.307--338.

Martinussen, T. and Scheike, T. (2006). *Dynamic regression models for survival data*. 1st ed. New York, N.Y.: Springer.

Martinussen, T. (2010). Dynamic path analysis for event time data: large sample properties and inference. *Lifetime data analysis*, 16(1), pp.85--101.

McBride, W. (1961). Thalidomide and Congenital. *The Lancet*, 278(7216), p.1358.

Metzger, B., Vileisis, R., Ravnikar, V. and Freinkel, N. (1982). "Accelerated Starvation" and The Skipped Breakfast In Late Normal Pregnancy. *The Lancet*, 319(8272), pp.588--592.

Mortality.org, (2015). *HMD Country*. [online] Available at: http://www.mortality.org/cgi-bin/hmd/country.php?cntr=SWE&level=1 [Accessed 31 May 2015].

Nelson, R. E. (2010). Testing the fetal origins hypothesis in a developing country: evidence from the 1918 Influenza Pandemic. *Health economics*, 19 (10), pp. 1181--1192.

Neugebauer, R., Hoek, H. and Susser, E. (1999). Prenatal Exposure to Wartime Famine and Development of Antisocial Personality Disorder in Early Adulthood. *JAMA*, 282(5), pp.455--62.

Otake, M. and Schull, W. (1998). Review: Radiation-related brain damage and growth retardation among the prenatally exposed atomic bomb survivors. *International Journal of Radiation Biology*, 74(2), pp.159--171.

Painter, R., Roseboom, T. and Bleker, O. (2005). Prenatal exposure to the Dutch famine and disease in later life: An overview. *Reproductive Toxicology*, 20(3), pp.345--352.

Palloni, A., Milesi, C., White, R. and Turner, A. (2009). Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. *Social science & medicine*, 68(9), pp.1574--1582.

Patterson, K. (1985). Pandemic and Epidemic Influenza, 1830-1848. *Social Science & Medicine*, 21(5), pp.571--580.

Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012). Lifelong Socio Economic Position and biomarkers of later life health: A formal comparision of the critical period, accumulation and chains of risk hypotheses. Paper presented at the Gerontological Society of America 2012 in San Diego, USA.

Potter, C. (2001). A history of influenza. *Journal of Applied Microbiology*, 91(4), pp.572--579.

Preston, S. and van de Walle, E. (1978). Urban French Mortality in the Nineteenth Century. *Population Studies*, 32(2), pp.275--297.

Quaranta, L. (2012) STAT code to transform data extracted from the Intermediate Data Structure into rectangular episodes tables. *http://extract.sedd.ed.lu.se/ExtractionFileList.aspx.* Lund University, Center for Economic Demography.

Quaranta, L. (2013). *Scarred for life: how conditions in early life affect socioeconomic status, reproduction and mortality in Southern Sweden, 1813-1968*. 1st ed. Lund: Lund University.

Quaranta, L. (2014). Early life effects across the life course: The impact of individually defined exogenous measures of disease exposure on mortality by sex in 19th- and 20th-century Southern Sweden. *Social Science & Medicine*, 119, pp.266--273.

Rasmussen, K. (2001). The "Fetal Origins" Hypothesis: Challenges and Opportunities for Maternal and Child Nutrition. *Annual Review of Nutrition*, 21(1), pp.73--95.

Rasmussen, S., Jamieson, D. and Bresee, J. (2008). Pandemic Influenza and Pregnant Women. *Emerg. Infect. Dis.*, 14(1), pp.95--100.

Ravelli, G., Stein, Z. and Susser, M. (1976). Obesity in Young Men after Famine Exposure in Utero and Early Infancy. *New England Journal of Medicine*, 295(7), pp.349--353.

Ravenholt, R. and Foege, W. (1982). 1918 Infulenza, Encephalitus Lethargica, Parkinsonism. *The Lancet*, 320(8303), pp.860--864.

Richter, A. and Robling, P. (2013). Multigenerational effects of the 1918-19 influenza pandemic in Sweden. *Working Paper Stockholm University*.

Roseboom, T., van der Meulen, J., Osmond, C., Barker, D., Ravelli, A., Schroeder-Tanka, J., van Montfrans, G., Michels, R. and Bleker, O. (2000). Coronary heart disease after prenatal exposure to the Dutch famine, 1944-45. *Heart*, 84(6), pp.595--598.

Roseboom, T., van der Meulen, J., Ravelli, A., Osmond, C., Barker, D. and Bleker, O. (2001). Effects of Prenatal Exposure to the Dutch Famine on Adult Disease in Later Life: An Overview. *Twin res.*, 4(05), pp.293--298.

Rosenzweig, M. and Zhang, J. (2009). Do Population Control Policies Induce More Human Capital Investment? Twins, Birth Weight and China's "One-Child" Policy. *Review of Economic Studies*, 76(3), pp.1149--1174.

Statistiska Centralbyrån, (2015). *Befolkningsutveckling; födda, döda, in- och utvandring, gifta, skilda 1749–2014*. [online] Available at: http://www.scb.se/sv_/Hitta-statistik/Statistik-efter-amne/Befolkning/Befolkningens-sammansattning/Befolkningsstatistik/25788/25795/Helarsstatistik---Riket/26046/ [Accessed 31 May 2015].

Van den Berg, G., Doblhammer, G. and Christensen, K. (2009). Exogenous determinants of early-life conditions, and mortality later in life. *Social Science & Medicine*, 68(9), pp.1591--1598.

Van Den Berg, G., Lindeboom, M. and Portrait, F. (2006). Economic conditions early in life and individual mortality. *The American Economic Review*, pp.290--302.

Van De Putte, B. and Miles, A. (2005). A social classification scheme for historical occupational data: partner selection and industrialism in Belgium and England, 1800-1918. *Historical Methods*, 38(2), pp.61--92.

Verbeek, M. (2012). *A guide to modern econometrics*. 4th ed. Chichester: Wiley.

Öberg, S. (2014). *Social Bodies: Family and community level influences on height and weight, southern Sweden 1818–1968*. 1st ed. Gothenburg: Unit for Economic History, Department of Economy and Society, School of Business, Economics and Law, University of Gothenburg.