LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Demography

The effect of conditions in early life on health in adult life
A study of 19th century Scania

Filip Andersson
ake07fan@student.lu.se

Abstract: In this thesis the impact of conditions in early life on survival in adult life is studied. This is done for longitudinal level data from Scania for 1815 to 1910 and conditions in early life are measured by infant mortality rate. In the thesis a model that allows for both direct effect on the survival form the infant mortality rate and indirect effects mediated through socioeconomic status and marriage is presented. The analysis is done separately for women and men, and support for "Fetal origins hypothesis" is found for both sexes. Furthermore, also a counter effect mediated through socioeconomic status is found both for men and women. This counter effect support the theory of survival of the strongest individuals during bad early life conditions.

Key words: Adult health, dynamic path analysis, early life conditions, econometrics, economic demography, survival analysis

EKHM51
Master thesis (15 credits ECTS)
June 2014
Supervisor: Tommy Bengtsson
Supervisor: Luciana Quaranta
Examiner: Patrick Svensson

Table of Contents

1. Introduction

1.1. The aim and goal of the thesis

The aim of this thesis is to study survival in adult life for individuals being exposed to undesirable conditions in early life in contrast to those who were not. How and in which way could this affect the survival chance in adult life for these individuals? The paper will take the theoretical approach from the literature of the "Fetal origins hypothesis" (Barker, D. (1990, 1995)) and 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 study will be performed on data from the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)) for cohorts born between 1815 and 1850.

In this paper the measure of bad condition in early life for the individuals will be the infant mortality rate during the year of birth for the individual. This approach has been taken in previous studies by for example Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012), Quaranta, L. (2013) and Öberg, S. (2014). A possible problem, and a limitation in the thesis, with using the cohort based mortality rate for the year of birth as an approximation for the early life conditions is that it is impossible to distinguish between if the cause of the high mortality rate is due to bad conditions during the stage of fetal or infancy for the individual (Quaranta, L. (2013). This is due to that individuals born in the first part of the year will experience most of the year as infants and those born during the latter part of the year will experience most of the year in-utero.

Furthermore, the effects from being married or not and the socioeconomic status for an individual on the survival will be studied. Here, the variables for marriage and SES will be seen as possible mediator for indirect effect from bad early life conditions.

In many previous studies, for example Almond, D. (2006), Bengtsson, T. and Broström, G. (2009) and Quaranta, L. (2013), strong support for the "Fetal origins hypothesis" have been found, as a direct effect from bad early life conditions on mortality, but there is a problem with distinguishing between the effect being a latent direct influence or being mediated through a pathway model with indirect effects. The paper by Edvinsson, S. and Broström, G. (2012) is trying to evaluate the existence of indirect effects in a pathway model, but finds only weak support for indirect effects. As for this paper the direct effect from bad early life condition on survival in adult life is expected to be found. Furthermore, the indirect effect mediated through socioeconomic status could be found and also the indirect effect mediated through marriage.

The methodological approach in this paper will be taken from the dynamic path analysis, presented by Aalen, O., Borgan, Ø. and Gjessing, S. (2008). This model was designed to analyze survival rate in biomedicine, but has been extended to demographic survival analysis by Broström, G. and Edvinsson, S. (2012), where they use the model of accelerated failure time instead of the Aalen´s additive hazard model. This approach has then been used in the paper by Edvinsson, S. and Broström, G. (2012), where they, with a data set similar to the SEDD, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)), but for the region around the city Sundsvall instead, have analyzed survival in old age in a historical context.

1.2. Summary of the findings in the thesis

From the analysis in the thesis support for the "Fetal origins hypothesis", namely that a latent effect of being born in a year of bad early life conditions with a negative effect on survival in later adult life is found for both females and males. If this latent effect is due to bad conditions during the fetal stage or during the first year of life is not possible to distinguish between. Furthermore, an indirect effect from being born during a year with bad infant conditions is found mediated through the socioeconomic status. This indirect effect operates as a counter effect to the direct negative effect on survival and is support for the theory of selection between individuals where the strongest individuals survive. This indirect effect is found both for women and men and are mostly found in later adult life, but some support for the theory of selection are also found for both men and women in the early adult life.

1.3. Outline of the thesis

The thesis consists of six chapters followed by a list of references. The first chapter is the introduction, where the aim of the thesis is presented and the research question is formulated.

The chapter that follows is the theory. It consists of two parts, the first describes the "Fetal Origins Hypothesis" which could be seen as the base of this thesis and the second part is about the extended pathway model.

Chapter three is about the methodology. Here, the first part formulates the pathway model that the analysis will be based on and the possible mechanism for this model. The second part of chapter three presents the statistical model of dynamic path analysis, which the results are based on.

The following chapter is about the data used in the analysis. This chapter contains of three parts, the origin of the data, the structure of the data and the variables used in this thesis.

Following the presentation of the data is the results as chapter five. First some descriptive statistics of means are presented and then the full analysis of the pathway model.

The last chapter is the conclusions. Here the results are connected with the theory presented in chapter two and the model presented in chapter three.

1.4. Acknowledgements

I would like to thank my supervisors Tommy Bengtsson and Luciana Quaranta for the support and the suggestions they have given me during the work of the thesis. I also like to thank Tommy for introducing me to the subject and Luciana for helping me to get started with the data management. At last I like to thank Göran Broström for answering questions regarding interpretation of the results.

2. Theory

2.1. Fetal origins hypothesis

Before the mid-20th century researchers commonly thought that the period of in-utero was a safe environment for the human fetus and that the placenta was completely protecting the fetus from harmful events (Almond, D. and Currie, J. (2011)). It was a general belief that mothers could have a drink or smoke during pregnancy and that the fetus still would be safe (Almond, D. and Currie, J. (2011)). The theory that the fetus would not be affected by the lifestyle of its mother is nowadays considered to be wrong (Almond, D. and Currie, J. (2011)). Instead researchers try to explain what hazards that the mother can be exposed to that are affecting her child during in-utero conditions. In this thesis one of these theories will be discussed, namely the "Fetal Origins Hypothesis", presented by Barker, D. (1990, 1995). Barker tries in this theory to link the fetal in-utero conditions, for example if the mother smoked, with the health of the individual in adult life (Barker, D. (1995), Calkins, K. and Devaskar, S. U. (2011)). The basis for the theory of the "Fetal Origins Hypothesis" is that hazards that the individual experience during in-utero should be affecting the health in later life as continuous and that it can be latent for a long time (Almond, D. and Currie, J. (2011)). The idea is that the effects from a specific hazard that the individual was exposed to during in-utero can affect at any time and that the hazard event trigger a specific biological mechanism (Almond, D. and Currie, J. (2011)). For example the theory has been used to explain the risk of having heart problems in old age (Almond, D. and Currie, J. (2011), Barker, D. (1995)).

In the early studies presented by Barker the empirical work that was presented was correlation coefficients (Almond, D. and Currie, J. (2011). In Barker, D. and Osmond, C. (1986) the correlation was between infant mortality rates and adult mortality rates grouped geographically and in Barker, D. (1995) between birth weight and adult life mortality. The ways of how to measure the early life conditions have later been extended to a number of technics. In Bengtsson, T. and Lindström, M. (2000, 2003) they combine longitudinal data with infant mortality rate to measure the impact of exogenous variation in early life conditions. Furthermore, one technic that has been used in a number of recent studies, like for example Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2013), is the infant mortality rate for the year of birth first presented in Bengtsson, T. (2004). Another possible measure is to use a natural quasi-experiment, like for example an outbreak of highly contagious disease or a natural disaster, presented by Almond, D. (2006) and also used in for example Almond, D., Edlund, L. and Palme, M. (2009) and Nelson, R. E. (2010).

The "Fetal Origins Hypothesis" was originally formulated for use in epidemiology, but has later been adopted by economists that have included other outcomes then risk of diseases, for example socio-demographic outcomes like educational attainment, wages and socioeconomic status (Almond, D. and Currie, J. (2011), Almond, D., Edlund, L. and Palme, M. (2009) and Bengtsson, T. and Broström, G. (2009)). Also, the researchers that have presented the best evidences for the existence of the "Fetal Origins Hypothesis" are economists (Almond, D. and Currie, J. (2011)).

2.2. The pathway model

A different approach, bur closely related to the "Fetal origins hypothesis", is the pathway model (Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012)). In general the pathway model, in comparison to the "Fetal Origins Hypothesis", predicts that the adult environment plays a larger role for future adult outcomes instead of just the early life conditions (Nelson, R. E. (2010)). The model assumes that the treatment variable could affect the outcome, not only as a latent direct effect, like in Bakers theory, but also as indirect effects that is mediated by different covariates (Edvinsson, S. and Broström, G. (2012)). These indirect effects can be from social and economic sources like socioeconomic status over the life-course, education or income, but could also be environmental factors, for example Gross Domestic Product per capita (Van den Berg, G., Doblhammer, G. and Christensen, K. (2009) and Van Den Berg, G., Lindeboom, M. and Portrait, F. (2006)). Furthermore, Almond, D., Edlund, L. and Palme, M. (2009) also state the importance of the parental influences for the outcome of educational attainment for the individual. Similarly, Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012) present that individuals from a family with a more desirable parental background have a better chance of achieving a good socioeconomic status.

Marriage is described to be a strong social protective factor in a historical context (Edvinsson, S. and Broström, G. (2012)). The arguments are that it secure individuals with emotional and financial support and should be seen as an important factor for old age survival (Edvinsson, S. and Broström, G. (2012)).

In the pathway model it is important to separate between conditions and events that is significant related to early life conditions and those that are not (Edvinsson, S. and Broström, G. (2012)). The pathway model allows for different types of risks over the life-course, both critical period, like "Fetal Origins Hypothesis", but also accumulated risks, like different types of exposure that take place over time (Bengtsson, T. and Mineau G. (2009)). Some of the most considered effects in a pathway model are, education, income, socioeconomic status, marital status and number of children (Edvinsson, S. and Broström, G. (2012)). One important argument of why the socioeconomic status is affected by the early life conditions is the possible existence of a scarring effect on cognitive ability. This is highlighted in studies by for example Almond, D., Edlund, L. and Palme, M. (2009) and Palloni, A., Milesi, C., White, R. and Turner, A. (2009). The idea is that an individual born during bad conditions will not develop in a desirable way and therefore may not develop basic skills, perform well in school or participate in social activities (Palloni, A., Milesi, C., White, R. and Turner, A. (2009)).

External level influences on the individual are also an important factor to take into consideration. Bengtsson, T. and Lindström, M. (2000) presents that economic variation could affect the health in childhood years, but that this effect is not present in the stage of infancy. This is due to that during the stage when a child is still breastfed the child is primarily affected by the lifestyle of its mother and possible outbreaks of epidemic diseases (Quaranta, L. (2013)).

Furthermore, there are scholars that are considering height as a mediating factor for early life conditions on later life health, for example Fogel, R. (2005) and Öberg, S. (2014). The argument is that there is a direct negative effect on height from being exposed to diseases in early life (Öberg, S. (2014)). But, there could also be a counter effect, that comes from selection due to that the strongest infants survive during periods of high infant mortality rate and the surviving individuals would then be taller (Öberg, S. (2014)). In general the influences from early life that would affect health is described in Öberg, S. (2014) as the nutritional status of the mother and by living conditions during the early life. Also, Öberg, S. (2014) argues that there is a possible effect between socioeconomic status and the height of an individual and this effect would be an indicator of the influence of environmental factors on growth.

In a pathway model both effects and counter effects are allowed, for example having access to medical treatment could have a counter effect to having been exposure to a specific type of health hazard, while the hazard itself could have a scaring effect, making it harder to acquire the socioeconomic status needed to afford the medical treatment (Bengtsson, T. and Mineau G. (2009)). The indirect effect through socioeconomic status for health in adult life from bad fetal conditions together with the possibility of the opposite effect due to selection is also highlighted in Bengtsson, T. and Broström, G. (2009). Furthermore, Bengtsson, T. and Broström, G. (2009) ague that poor socioeconomic status in childhood could lead to poor conditions in later life and that would finally lead to lower chance of survival in adult life.

In general the risk of cofounding is always present when performing pathway model analysis (Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012)) due to the existence of scarring, the "negative" effect of disease, and selection, the "positive" effect of disease and when evaluating results from analysis both these effects has to be taken into account. Though the complicity and endless possibilities of the pathway model makes the modeling of these effects hard, most of the recent papers on analysis of early life conditions on outcomes in later life take this approach.

In this thesis the approach of the pathway model will be taken. Here the critical period for the individual will be seen as being born during bad early life conditions or not, in line with the "Fetal Origins Hypothesis" (Barker, D. (1990)). Furthermore, the chains of risk over the life course will be seen as time spent as married or not and the changes is socioeconomic status over the life-course.

3. Methodology

The following chapter of the paper consists of two parts. The first part describes the model, based on the theory, which will be used in the analysis of this thesis. Here a connection between the theory and the model will be presented. The second part will describe the statistical models used in the analysis.

3.1. Mechanisms of the pathway model tested in the analysis

The theoretical approach taken in this thesis is the pathway model, discussed in chapter 2.2. In this chapter a presentation of the pathway model used in this thesis will be presented both graphically in figure 1 and in the following paragraphs.

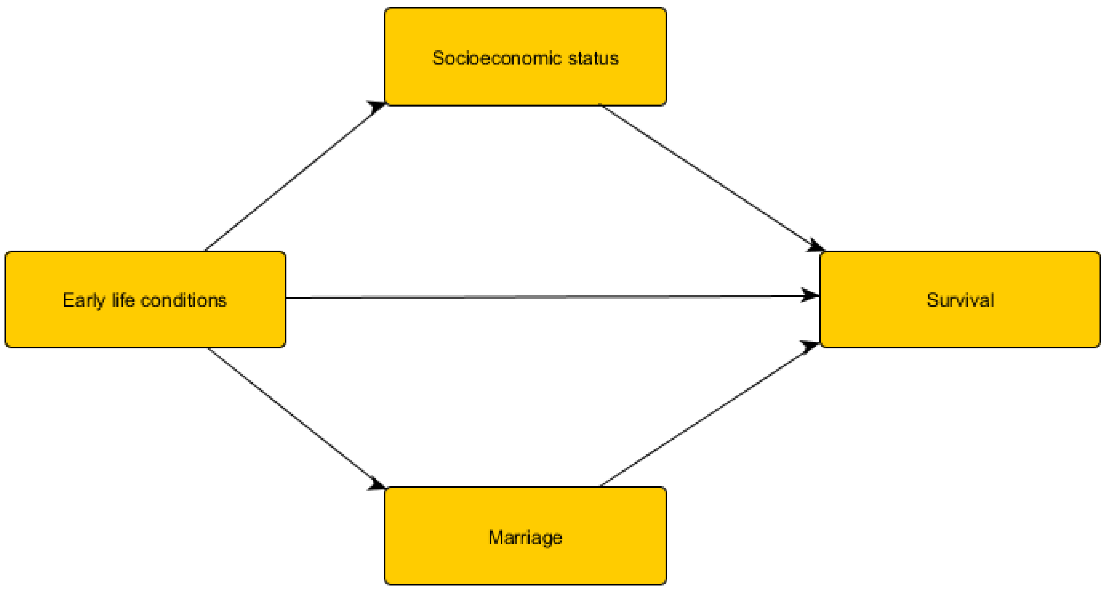

Figure 1 The pathway model used in this thesis

First, the situation of having experienced bad early life conditions or not could affect the survival in adult life directly as a latent effect in line with the "Fetal origins hypothesis" (Almond, D. and Currie, J. (2011) and Calkins, K. and Devaskar, S. U. (2011)). In figure 1 this represents the arrow from early life conditions to survival. Furthermore, the model allows for two ways of indirect effect, either mediated through socioeconomic status or marriage, compare with for example the model by Edvinsson, S. and Broström, G. (2012). In figure 1 the indirect effects are represented as the arrows from early life conditions to marriage and socioeconomic status together with the arrows from socioeconomic status and marriage to survival. The mechanism for the mediating effect through socioeconomic status could be, if seen as scarring (Almond, D., Edlund, L. and Palme, M. (2009) and Palloni, A., Milesi, C., White, R. and Turner, A. (2009)), that bad early life conditions would lead to less ability to perform in school or work and would lead to a lower achieved socioeconomic status. This would then lead to less possibility to live a healthy life, which then would lead to a lesser chance of surviving in adult life. There could also be a counter effect on survival (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)). In that situation the strongest individuals, with higher ability to work, would survive bad early life conditions and that would lead to better achieved socioeconomic status, which then could lead to better survival in adult life.

The same as for the indirect effect mediated through socioeconomic status, the possibility of scarring and selection is also present for the mediating factor marriage (Edvinsson, S. and Broström, G. (2012)). The mechanism here for scarring is that the bad early life condition would lead to weaker, less wanted individuals that would have a lesser chance of finding someone to marry. Then those unmarried individuals would have a lesser chance of having children, which, at least in a historical context, would reduce the chance of having any support in old age and would then reduce their chance of surviving (Edvinsson, S. and Broström, G. (2012)).

Furthermore, the idea for the opposite selection effect, is that the individuals with more attractive abilities survive and would have a better chance of getting married and therefore experience the advantages from marriage on chance of surviving (Edvinsson, S. and Broström, G. (2012)). Here, also the possibility of correlation between marriage and socioeconomic status arise, since higher achieved social class could be seen as an attractive ability and if those individuals that survived were the strongest then there could be a close correlation between marriage and socioeconomic status.

3.2. Dynamic path analysis

The method used in this thesis to analyze the data originates from the dynamic path analysis presented by Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006) and further described in Aalen, O., Borgan, Ø. and Gjessing, S. (2008). The main idea with the dynamic path analysis is to analyze the chance of survival for a path model, with the ability of the affecting covariates to vary over time. The variation in time is allowed both for the direct effect from the treatment variable, in this thesis, being born during a year of high infant mortality rate, and for the indirect effects that operates through mediators (Aalen, O., Borgan, Ø. and Gjessing, S. (2008)). In this thesis the mediating effects that will be studied are the socioeconomic status and a variable for being married or not, for more detail see chapter 4.3.

The original model, by Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006), is constructed to be used for survival analysis in a medical context, for example the effect of survival over time for a patient treated with a specific medicine, while here the context is instead of a demographic nature. In Broström, G. and Edvinsson, S. (2012), this matter is addressed and they present a more suitable model for mortality and survival analysis. They have as a starting point the method presented by (Lange, T. and Hansen, J. (2011)), which should be seen as a mixture of the dynamic path analysis and the counterfactual model (Edvinsson, S. and Broström, G. (2012)). Instead of using the Aalen´s additive hazard model, where the response variable is an intensity function for a counting process (Aalen, O., Borgan, Ø. and Gjessing, S. (2008)), which is generally used in the epidemiology and medicine context (Fosen, J., Ferkingstad, E., Borgan, Ø. and Aalen, O. (2006), Lange, T. and Hansen, J. (2011) and Martinussen, T. and Scheike, T. (2006)), Broström, G. and Edvinsson, S. (2012) take the approach of accelerated failure time model, a model that use the logarithmic value of survival in years as the response. This type of response variable fits well with how demographic longitudinal data are structured.

The model consist of two parts, one for analysis of the direct effect from different covariates, here the accelerated failure time model, and one part that analyze the indirect effect from the treatment variable (Broström, G. and Edvinsson, S. (2012)). The second part is handled through linear regression of the treatment variable on the mediating covariate (Lange, T. and Hansen, J.

(2011) and Martinussen, T. and Scheike, T. (2006)), which are included in the accelerated failure time model. The linear equations are calculated simultaneously with respect to the accelerated failure time model. The structural equation system for a specific time, in reality interval, is as follows (Broström, G. and Edvinsson, S. (2012)):

The accelerated failure time model:

$$\log(T) = \beta_0 + \beta_1 z_1 + \cdots + \beta_p z_p + \varepsilon$$

The linear regression model:

$$z_p = \gamma_0 + \gamma_1 z_1 + \eta$$

Here z_1 is the treatment variable, 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)):

$$Indirect\ effect = \beta_p * \gamma_1$$

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

$$Total\ effect = Direct\ effect + Indirect\ effect$$

Finally, the way to interpret the results from the presented model is that the coefficient β_p is the change in expected life compared to the baseline life expectancy on a logarithmic scale. To recalculate this onto the normal scale the equation will be for the direct effect (Broström, G. (2012)):

$$Expected\ life = Baseline\ life\ expectancy * e^{\beta_1}$$

Furthermore, for the indirect effect and the total effect the same equation is used, but the coefficient for the direct effect is replaced by the coefficient for the indirect effect or the total effect.

4. Data

4.1. The origin of the data

The analysis in this thesis is performed on a dataset that is a subset of the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)). The SEDD is a longitudinal database collected from five rural parishes in the western part of Scania (Quaranta, L. (2013)). The included parishes are Hög, Kävlinge, Halmstad, Sireköpinge and Kågeröd and the data dates back to the middle of the 17th century (Bengtsson, T. and Broström, G. (2009)). The five parishes are located nearby each other and had a normal variation in regard of socioeconomic status and topography for an agricultural society in the historical context (Quaranta, L. (2013)). The data cannot be seen as a generalization of the conditions in Sweden, but as an approximation for people living in a rural setting (Quaranta, L. (2013)).

The database is constructed from a number of sources. The births and deaths are mostly based on population registers, while occupations and socioeconomic status are obtained from poll-tax and income registers (Quaranta, L. (2013)).

Around the year 1840 the population in the five parishes were about four thousand, but increased during the second half of the 19th century and in 1900 the population total had exceeded six thousand (Quaranta, L. (2013)).

The followed cohorts that are used in the analysis are the once born between 1815 and 1850. There are a couple of reasons for the choice of this subset of the Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)). First, since the analysis concern following the individuals from birth until the age of seventy and the last year of data is for 1910, to be able to follow at least most of the individuals until death or right censoring, the youngest cohort has to be born in the first half of the 19th century. In the data set used, the last cohort that is possible to follow the complete course until the age of seventy is those born in 1840. For the last ten cohorts the oldest ages when the individuals are possible to observe are between sixty to sixty-nine depending on the year of birth. Furthermore, Quaranta, L. (2013) presents that 1850 can be seen as a breaking point in the SEDD, because the infant mortality rate is much higher before than after that period. The choice of individuals born in 1815 as the first included cohort is due to the fact that the data for the parish that dates back the shortest time starts in the early 19th century and in 1815 all perishes are included.

4.2. Structure of the data

When receiving the data from Scanian Economic Demographic Database, (Bengtsson, T., Dribe, M. and Svensson, P. (2012)), the structure of the data follows the Intermediate Data Structure. The data is then transformed into rectangular episodes tables in STATA by using to code by Quaranta, L. (2012). Furthermore, the socioeconomic status of the individuals contain a lot of variables, to simplify this, the code by Dribe, M. (2012) is used. The data follows individuals from when they are born or in-migrate until they die or out-migrate. Important life course events and other information are also registered, for example births of children, marriage or changes in socioeconomic status (Quaranta, L. (2013)).

The analysis in this thesis is performed with bivariate outcomes, zero and one, on the used variables. For example the treatment variable indicates if an individual is born during a year of high infant mortality rate or not. This is partly due to the simplicity of this types of analysis in regard to evaluate the results, but also that the quality of data, due to its historical context, is not detailed enough. To allow the use of fully numerical values for all variables there is a need of a more detailed data set. It is worth pointing out that the methods used in the analysis of this paper allows for fully numerical data and a similar analysis on contemporary, more detailed, data could achieve better results.

The individuals in the data are allowed to be in- and out-migrated. The reason that this is possible is that the accelerated failure time model, used is the analysis, allows the use of right censored data (Broström, G. (2012)). Furthermore, an individual is kept in the data set only if they do not have missing values for their socioeconomic status and their civil status. If individuals with missing values in these variables are kept they could be a source for unnecessary bias in the analysis.

The total number of individuals born between 1815 and 1850 in the data set is 12971, of which 6355 are women and 6616 are men. Furthermore, there were 4167 men and 4076 women in the data that at any point in time was alive in the parishes in the age of twenty and 1603 men and 1505 women in the parishes alive in the age of forty. The reason for presenting the number of individuals in the ages twenty and forty is due to that the analysis will be divided into adult life intervals and that the first interval will start at age twenty and end at age forty and the second will start at age forty and end at age seventy.

4.3. Variables used in the analysis

As presented previously in this paper, the main goal of this thesis is to test for the "Fetal Origins Hypothesis" and possible pathways that bad early life conditions could act through. In this paragraph the variables used in the analysis is presented and discussed, both from this thesis point of view and also in comparison to work by other authors.

Before the presentation of the variables, it has to be pointed out that the analysis will be done separately for females and males. In regard to how the statistical model (see chapter 3.2) is defined, this would be the simplest way to be able to identify differences between the sexes. In the study by Edvinsson, S. and Broström, G. (2012), which uses a similar methodological approach, the same consideration is done.

The first variable to discuss is the treatment variable. That is if an individual in early life is experience bad life conditions. The choice in this thesis of how to measure the early life conditions is by the infant mortality rate the year that the individual was born. This measure is used in studies by for example Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and (Quaranta, L. (2013). In for example Quaranta, L. (2013) and Bengtsson, T. and Broström, G. (2009) also the variation in price index are used as indicator of bad nutrition during the fetal stage. There, the prize index for the year before birth is linked to the individual. The reason for that is that it has been shown that economic variation effect childhood death (Quaranta, L. (2013)).

As said previously in this study, the variable infant mortality rate will be used as an approximation for early life conditions (Bengtsson, T. (2004)) and is set to be a dummy variable equal to one if the individual is born during a year with high infant mortality rate and otherwise equal to zero. In Edvinsson, S. and Broström, G. (2012) instead of using a dummy variable indicating high IMR or not they use the fully mean-centered value of IMR. Here, the years with high infant mortality rate that are chosen as indicator years are the same as the ones presented in Quaranta, L. (2013), namely the years 1816, 1821, 1826, 1831, 1832, 1835, 1838 and 1846. During these years with high infant mortality rate there were outbreaks of different epidemics which consisted of outbreaks of measles, smallpox and whooping cough (Quaranta, L. (2013)). Of these diseases the most common one was whooping cough. The calculated infant mortality rate used to find the years with high IMR comes from calculation on the data from SEDD done by Quaranta, L. (2013). Furthermore, Quaranta, L. (2013) also shows that IMR can be seen as exogenous in respect to later life health. With respect to that, results found in the analysis could be seen as causal findings.

One possible problem with the variable for infant mortality rate could be the allowing of in-migration in the analysis. The problem would then be that an individual that was born during year of high IMR but not in one of the five parishes would be seen as born during an exposure, even if there were no high IMR during that year at the place of birth for the specific individual. This could be a cause of bias in estimation of the survival in old age. During the studied period there were high migration rates within Scania and around half of the population was not born in the same parish as their current residence, but most of the migrants were born in a parish located close (Quaranta, L. (2013)). Since most of the migration was within Scania only 15% of the total population was born in another county or country (Quaranta, L. (2013)). Furthermore, Quaranta, L. (2013) checks if the outbreaks in Malmöhus county, a larger area of Scania in which the five parishes is located, show similar patterns to the five parishes itself and the result is that it does. The conclusion is that it seems that if the in-migrated individuals come from the regions nearby the five parishes then the problem may not exist, but it they have migrated from further away, they are probably a higher risk of bias.

Furthermore, another cause of bias in the IMR is that it is yearly averages, namely cohort based IMR (Quaranta, L. (2013), meaning that for a complete cohort all individuals are seen as being exposure equally long time to the high infant mortality rate. This could lead to underestimation if the individual are born early in the period of outbreaks and overestimation if the individual is born during the end of an outbreak. There is also the possibility that, if high infant mortality rate would indicate bad in-utero conditions, those born late in a year of high IMR did experience bad in-utero conditions and those born yearly in a year of high IMR did experience bad infant conditions.

The next variable to discuss is the response variable in this study which is the event of death. The variable is called death indicator and indicates if and when the individual dies. This is done as the last spell for the individual. If the individual died in another place than one of the five parishes then the death would not be recorded, but the individual will be right-censored when they out-migrated. Individuals that die before the age of twenty will not be considered in the analysis since the analysis start for individuals from the age of twenty. Individuals that in-migrate after the age of twenty will be included in the analysis, but they are left-censored. As for individuals that survive the age of seventy they are of course included too. Individuals in the last ten cohort will not be followed until the age of seventy, but if they survive, until the year 1910. For example the last cohort (1850) can only be followed until the age of sixty.

In the study there are two covariates with the possibility of being mediators. The first one is socioeconomic status. This is in line with the mediating analysis in the study by Edvinsson, S. and Broström, G. (2012), who considers socioeconomic status at age fifty. The socioeconomic status variable is, as said before, presented in many different variable in SEDD and the code for STATA written by Dribe, M. (2012) sort it into one variable. The socioeconomic status for the individual is in the data taken for the head of the family that the individual belongs to. For men this generally means that when they live at home it is the SES of their father that is considered and when they move away from home it is their own SES. For women in general the SES of their father is considered until they get married, then their husbands SES would be considered. There are different standards of how to classify socioeconomic status, in the code by Dribe, M. (2012) two classifications are possible to choose from. In this thesis the one called Social Power, SOCPO, by Van De Putte, B. and Miles, A. (2005) is used. The variable contains of five different levels of socioeconomic status, 1, 2, 3, 4 and 5. Here, 1 is defined as individuals that are unskilled, 2 is semi-skilled, 3 is manual skilled, 4 is manual superskilled and 5 is nonmanual superskilled (Quaranta, L. (2013)). These classes are later transformed into a dummy variable indicating if an individual belongs to the lowest social class or not. Technically this means that if an individual has a SOCPO of 1 then the variable socioeconomic status, in the result presented as SES.low, will be equal to one and otherwise the individual will have the number zero for socioeconomic status. For the socioeconomic variable Bengtsson, T. and Broström, G. (2009) use a similar approach since they categorize families into the groups landed and landless. In Edvinsson, S. and Broström, G. (2012), they use an integer scale with six steps instead. The socioeconomic variable used in this thesis is a time dependent variable that is allowed to change value at any time during an individual's life-course. As said before only the individuals with known socioeconomic status are used in the analysis. The choice of excluding individuals without known SES are in line with what has been chosen by different scholars like Bengtsson, T. and Broström, G. (2009), Edvinsson, S. and Broström, G. (2012) and Quaranta, L. (2013).

The second covariate that is seen as a possible mediator in the analysis is the variable marriage. This variable is produced from the variable civil status in SEDD (Bengtsson, T., Dribe, M. and Svensson, P. (2012)), and is a dummy variable that indicates if the subject is or have ever been married. In Edvinsson, S. and Broström, G. (2012), their mediating variable for being married is used in a similar way. Edvinsson, S. and Broström, G. (2012) argue that marital status can be measured either as being married or not or by analyzing the age at which the individual is getting married. In this paper, since the use of longitudinal data with episodes, the timing of marriage is allowed to have an impact on the result from the analysis. There is a possible problem with having civil status as a time varying covariate which will be discussed in the end of this chapter.

The paper by Edvinsson, S. and Broström, G. (2012) also consider number of children as a mediator. Edvinsson, S. and Broström, G. (2012) describe the reason for the use of this variable to be that being exposed to a high disease load in infancy could lead to reduced fecundity or even infertility. More, having fewer children could be a risk of not having someone that could take care of the individual in old age and therefore lead to lower chance of survival in those ages especially in a historical context when the public social security was non-existing (Edvinsson, S. and Broström, G. (2012)). The variable number of children will not be considered in the analysis of this paper. The reason for this is the high risk of being too closely related to the civil status variable. As will be discussed further in the next paragraph, there is already a possible problem with correlation between the mediating variables. If an inclusion of the variable number of children should be done in the analysis it might have to be seen as a mediating effect of marital status on the survival. To be able to perform this analysis a much bigger data set is probably needed.

Furthermore, there is a possible problem concerning the variables socioeconomic status and marriage. Due to how these two variables are collected and recorded into SEDD (Bengtsson, T., Dribe, M. and Svensson, P. (2012)), there is a risk that these variables are too closely recorded. This problem emerges because when an individual is registered as married in the church books their socioeconomic is also recorded simultaneously and for females this often means that their socioeconomic status is just changed from their father´s to their husband´s. This would create two variables that record the same effect. But, there is also another possible risk because for men the problem could be that before they get married they work as farmhand, even for their own family and when they get married they get their own share of the family land and then they will be recorded as freeholders at the same point as they get married even if their real socioeconomic situation has not changed at all. This problem is also discussed in Ploubidis, G., Grundy, E., Benova, L., Laydon, D. and De Stavola, B. (2012), where it is argued that age, gender, marital status and number of children pose as variables with risk of cofounding in regard to the relationship between socioeconomic status and health in later life. To take a closer look at this problem, the correlations between the socioeconomic status and marriage for different ages are presented in table 1. Here, the correlation could be seen as a measure for how closely related getting married is to have a change in the variable for belonging to the lowest socioeconomic class or not. These correlations are calculated for the dummy variables for SES.low and marriage that are used in the analysis. There is a risk of bias by over counting some individuals in the correlation, since the structure of the data are in the form of longitudinal episodes and individuals do not need to have any change in their socioeconomic status, then they would be counted one time less than an individual with a change in the socioeconomic status.

Table 1 Correlation between socioeconomic status and being married

	Correlation between SES.low and Marriage
Female full population	0.23
Female 20-40	0.31
Female 40-70	0.05
Male full population	0.25
Male 20-40	0.45
Male 40-70	0.21

The first thing to notice from the correlations, presented in table 1, is that for the full population the correlation between SES and marriage is 0.23 for women and 0.25 for men. Here one can see that there is a possible relation between the variables. Next, a comparison between different stages in the life course is performed. Here the ages 0-20 are left out, partly because it is not in use in the analysis, but also because relatively few people get married before age 20 and without the existence of marriage this problem would not exist.

For female in the age of twenty to forty the correlation between SES and marriage is 0.31. Here the discussed problem may to exist and the use of both variables has to be done with a bit of caution. On the other hand, for women of older ages the problem seems to disappear, correlation only 0.05. This could be due to much fewer women getting married in those ages and for this group the analysis can be carried out without taking the problem into consideration.

For the male population the correlation between SES and marriage is surprisingly higher. For the ages twenty to forty the correlation is 0.45. Furthermore, the correlation do not go down as much as for women and for the ages forty to seventy the correlation is still 0.21. An explanation to this could be that men generally are getting married at an older age. Furthermore one explanation to the fact that the correlation generally is higher among men could be, as said before, that the men generally have a change from farmworkers to freeholders when they marry and inherit land for themselves.

5. Results

5.1. Descriptive statistics

Table 2 Weighted mean values for the individuals in the different time intervals

	Weighted mean value (percentage of total population)			Number of deaths
	IMR.high	SES.low	Married	
Female 20-40	0.199	0.230	0.568	126
Female 40-70	0.210	0.314	0.877	327
Male 20-40	0.229	0.192	0.481	122
Male 40-70	0.232	0.257	0.901	336

As an introduction to the analysis some simple descriptive statistics are first presented in table 2. The calculated mean values here are the weighted means against exposure time and weighted relative to frequencies of levels of factors (Broström, G. (2014)). Here one can see that the percentages of the individuals, present in the analysis, that are born during a year with bad conditions are about 20%. This is expected since 8/36 (number of years with high infant mortality rate/total number of years) is equal to 22%.

Next, one can observe that the proportions of individuals born during a years of high infant mortality rate are increasing, both for men and women, between the early adult life and older ages. This could be an indication of selection of the strongest individuals as discussed previously in the theory and will be further evaluated in the analysis.

Another interesting, and somewhat strange, observation is that there seems to be an increase in the number of individuals surviving to older age that are of the lowest socioeconomic class. This can be seen both for the female and the male population. One possible explanation to the higher proportion of individuals belonging to the lowest socioeconomic class could be that when the individual get to a certain age his or her children would inherit the family farm, thus the individual itself would be landless, but would still in reality belong to the socioeconomic class of his or her children. Further, a more expected observation is that the older the population gets, the bigger fraction are married. There seems to be a difference in the timing of getting married between women and men, where men seems to get married at an older age then women. Also the number of deaths increases with higher age, which is expected.

5.2. Analysis of direct and indirect effects

The analyses are divided between the sexes and first the results from the analysis of the females will be presented and after that the analysis for the men. The analysis for each sex is divided into two parts, the time intervals presented in the statistical theory, for detail see chapter 3.2; first the analysis for individuals in age twenty to forty and then for individuals between age forty to seventy. The analysis consist of the results for the possible direct effects on survival with an accelerating failure time model and analysis of indirect effects with a linear regression model for the infant mortality rate on the mediating covariate socioeconomic status and marriage, as described in chapter 3.1 and visualized in figure 1.

A limitation of the results in the thesis, already presented in chapter 1.1 and 4.3, is the use of cohort based mortality rates as an approximation for the early life conditions. This makes it impossible to distinguish between if the effect on survival is due to bad conditions during the stage of in-utero or infancy, but instead the results reflect a combination of both factors. If the individual is born in the first part of the year it will experience the most of the year as infant or if it is born during the latter part of the year will experience the most of the year in-utero.

As said before the data management is done in the program STATA, but the analyses are performed in the statistical program R with the package "eha" by Broström, G. (2014). Furthermore, the level of significance that is chosen in the analysis is the 10% level. The reason behind this 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.

Table 3 Effect from high infant mortality rate on survival for female in age 20 to 40

Female age 20-40 born 1815-1850 in 1830-1890			
Covariate	Coefficient	Standard error	Wald p-value
IMR.high	0.114	0.152	0.454
SES.low	0.047	0.131	0.720
Married	0.044	0.127	0.731
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.915	0.361	0.000
log(shape)	-2.727	0.522	0.000
Baseline expected life	43.6		
Events	126	Total time at risk	22132
Max. log Likelihood	-771.99	Overall p-value	0.836926

The first category that is analyzed is the female between twenty to forty years of age. First the accelerated failure time model for the direct effects is analyzed. The result from this analysis is presented in table 3. Here, no significant result is found from the impact of being born during a year of high infant mortality rate on the survival, p-value equal to 0.454. The same results are found for the two possible mediating covariates, belonging to the lowest socioeconomic group or not, p-value equal to 0.720, and being married or not, p-value equal to 0.731. Further, as an analogy to the results of no significant covariates the overall p-value for the whole model show no significance, see table 3.

Table 4 Effect from high infant mortality rate on socioeconomic status for female in age 20 to 40

The effect of IMR.high on SES.low for female age 20-40			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.311152	0.002385	<2e-16
IMR.high	-0.062511	0.005515	<2e-16

Table 5 Effect from high infant mortality rate on being married for female in age 20 to 40

The effect of IMR.high on Married for female age 20-40			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.574117	0.002584	<2e-16
IMR.high	-0.066790	0.005974	<2e-16

There are significant effects from being born during a year of high infant mortality rate both on the socioeconomic status and being married, the results are presented in table 4 for the socioeconomic status and in table 5 for being married. When studying the sign of the coefficient one can see that being born in a year of bad early life conditions reduces the risk of belonging to the lowest social class, here SES.low=1, but it also decreases the chance of being married, here Married=1. Then, as a direct analogy from the results of no significance from the direct effect analysis, the analysis of the indirect effects from being born with bad early life conditions through the mediating variables socioeconomic status and marriage shows no significant indirect effects on the survival.

A possible explanation to why the accelerated failure time model on survival did not produce any significant results is that the sample size in the analysis is too small and as a consequence the number of deaths for the individuals in this stage of life is too few. Another explanation to the result being insignificant could be the existence of both effects and counter effects, namely scarring and selection (Bengtsson, T. and Mineau G. (2009)). When continuing to analyze the results from the linear regressions of early life conditions on socioeconomic status and marriage, possible support for this theory can be found. The result of having less chance of getting married, if born during a bad year, is following the theory of scarring. Here, the result of having a lesser risk of belonging to the lowest social class is a support for the selection of the strongest individuals (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)).

Table 6 Effect from high infant mortality rate on survival for female in age 40 to 70

Female age 40-70 born 1815-1850 in 1865-1910			
Covariate	Coefficient	Standard error	Wald p-value
IMR.high	-0.088	0.047	0.058
SES.low	-0.077	0.042	0.069
Married	0.085	0.056	0.128
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.396	0.093	0.000
log(shape)	-3.016	0.215	0.000
Baseline expected life	28.7		
Events	327	Total time at risk	19603
Max. log Likelihood	-1576.4	Overall p-value	0.0507795

The next model to present the results for is the women in age forty to seventy, presented in table 6. Here the covariate high infant mortality rate, being born during a year of bad early life conditions or not, and socioeconomic status, belonging to the lowest social class or not, show significant results at the 10% significant level from the accelerated failure time model on survival. For the last covariate for being married or not there is no significance found. At last the full model shows significant results at the 10% significance level, with p-value equal to 0.0508.

The sign for the coefficient from the infant mortality rate shows that there is a negative effect on the chance of survival for the women born during a year of high infant mortality rate within the ages forty to seventy. When computing the impact of being born during bad early life conditions to the baseline life expectancy, the individual born during a bad year in mean, after age forty has a life expectancy of 26.3 years compared to the baseline life expectancy 28.7 years. The difference here gives a loss of 29 months in mean for those women born during bad early life conditions. This result shows that in this environment there is a good indication of the existence of the "Fetal of Origins hypothesis" (Barker, D. (1990)) since for women in the early adult life we do not find any significant differences between those born during a bad early life conditions or not, but here we find a latent negative effect on survival in later adult life for those women.

Furthermore, the coefficient for the socioeconomic status, belonging to the lowest social class or not, indicates that the poorest individuals have worse health compared to the rest of the female population in the ages forty to seventy. After computation of the coefficient we here find that those women belonging to the lowest social class in the ages forty to seventy compared to the baseline life expectancy has over 25 months shorter life expectancy then the rest of the female population. Furthermore, here a possible indirect effect for being born during a year with high infant mortality rate mediated though socioeconomic status could be found, which we will be investigating in the following analysis.

Table 7 Effect from high infant mortality rate on socioeconomic status for female in age 40 to 70

The effect of IMR.high on SES.low for female age 40-70			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.481147	0.003298	<2e-16
IMR.high	-0.013546	0.007428	0.0682

Table 8 Effect from high infant mortality rate on being married for female in age 40 to 70

The effect of IMR.high on Married for female age 40-70			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.858027	0.002249	< 2e-16
IMR.high	0.039008	0.005065	1.39e-14

The results from the regression of being born during a year with high infant mortality rate as independent variable on the mediating covariate for belonging to the lowest socioeconomic class or not is presented in table 7. The finding here is that there is a small negative effect that is significant at the level of 10%. This may seem to be an opposite effect of what would be expected since this means that if an individual is born during bad early life conditions it would have a little better chance of avoiding to belong to the lowest socioeconomic class. Here, this could be an effect from selection (Bengtsson, T. and Mineau G. (2009)).

Furthermore, a positive significant effect from being born during a year with high infant mortality rate on marriage is found, presented in table 8. This effect shows that those women born during a year of high infant mortality rate had a higher chance of being married after the age of forty. Since the effect from being married on the survival are insignificant, no mediating effect through marriage can be found for being born during a year with high infant mortality rate. However an indirect effect through the socioeconomic status can be found. This effect has a sign indicating a counter effect to the direct effect on survival from the infant mortality rate. This may be explained by selection of the strongest individuals in early life during bad conditions. This indirect effect is actually also found by Edvinsson, S. and Broström, G. (2012). The coefficient of this indirect effect is calculated, as described in chapter 3.2, to 0.00104, which gives the indirect effect for the female individuals with bad early life conditions in mean, compared to the baseline life expectancy, less than one month longer survival in the ages forty to seventy.

When calculating the total effect of being born during a year of high infant mortality rate the result is (-0.088 + 0.00104) = -0.08696. After calculations, the life expectancy is found to be 26.3 years and compared to the baseline life expectancy this represent a loss of 29 months for the women born during a year of high infant mortality rate from the age forty. Here, the results from the analysis of the indirect effect seem to be negligibly, since the expected life after age forty for those women born during a year of high infant mortality rate is almost the same while including or excluding the indirect effect.

Table 9 Effect from high infant mortality rate on survival for male in age 20 to 40

Male age 20-40 born 1815-1850 in 1830-1890			
Covariate	Coefficient	Standard error	Wald p-value
IMR.high	-0.038	0.133	0.777
SES.low	-0.249	0.142	0.080
Married	0.422	0.122	0.001
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.860	0.354	0.000
log(shape)	-2.650	0.521	0.000
Baseline expected life	40.2		
Events	122	Total time at risk	20978
Max. log Likelihood	-742.93	Overall p-value	0.0118753

Next, the analysis for the male population will be presented. First, for the time interval representing the age twenty to forty will be studied. Here, from the accelerated failure time model presented in table 9, no significant effect from being born in a year with high infant mortality rate on the survival of the individuals can be found, but the socioeconomic status show a significant effect at the significance level 10% on the survival. This effect equals that the expected life form age twenty would be 31.3 years compared to the baseline life expectancy of 40.2 years from age twenty. If recalculated into months, this represents a loss of 106 months for the individuals that belong to the lowest socioeconomic class. Here, the socioeconomic status seems to matter a lot for the survival of the male population.

Furthermore, there is also a significant positive effect, p-value equal to 0.001, from being married on the survival for the male individuals. This effect, when calculating with respect to the baseline life expectancy of 40.2, gives a life expectancy of as much as 61.3 years. Even if it is expected that socioeconomic status and marriage could be variables that matter a lot for the survival, these high numbers, especially for marriage, seems to be overestimated. This could be due to the fairly small proportion of events in the accelerated failure time model.

Here, the full model is showing significance, with p-value equal to 0.012, see table 9. Next the mediating effects from being born in a year with high infant mortality rate from the two covariates socioeconomic status and marriage will be presented.

Table 10 Effect from high infant mortality rate on socioeconomic status for male in age 20 to 40

The effect of IMR.high on SES.low for male age 20-40			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.258152	0.002411	<2e-16
IMR.high	-0.005061	0.005133	0.324

Table 11 Effect from high infant mortality rate on being married for male in age 20 to 40

The effect of IMR.high on Married for male age 20-40			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.482600	0.002758	< 2e-16
IMR.high	0.024012	0.005870	4.31e-05

From the linear regression model of being born during a year with high infant mortality rate on the socioeconomic status, presented in table 10, no significant effect from being born in a bad year is found, but from being married, presented in table 11, a positive effect of being born in a year with high infant mortality rate is found with a p-value equal to less than 0.001. The calculated indirect effect of being born in a year with high infant mortality rate on survival is here (0.024*0.422) = 0.0101, meaning that there is an advantage of being born in a year of high infant mortality rate on survival in the ages twenty to forty comparing to being born during a year with good conditions. This coefficient represent, with respect to the baseline life expectancy an increase of almost five months better survival for the men born during a year with high infant mortality rate. A possible explanation to this is the idea of natural selection, that the infants that survive bad conditions would be the strongest once. This theory is discussed in for example Nelson, R. E. (2010). Here, the result has to be taken with caution since it is based on the result from the accelerated failure time model for the marriage variable, which, as discussed earlier, could be overestimated.

Table 12 Effect from high infant mortality rate on survival for male in age 40 to 70

Male age 40-70 born 1815-1850 in 1865-1910			
Covariate	Coefficient	Standard error	Wald p-value
IMR.high	-0.093	0.049	0.057
SES.low	-0.107	0.048	0.026
Married	0.078	0.070	0.261
Basline Parameters	Coefficient	Standard error	Wald p-value
log(scale)	2.553	0.108	0.000
log(shape)	-2.687	0.216	0.000
Baseline expected life	29.9		
Events	336	Total time at risk	20260
Max. log Likelihood	-1642.9	Overall p-value	0.0316306

The last group to present the results of the path analysis for is the results for the men in age forty to seventy. From the accelerated failure time model for the survival, presented in table 12, there is a significant result, with p-value equal to 0.057, from being born in a year of high infant mortality rate or not. This effect can be seen as being born during those years has a negative impact on the chance of surviving in the ages forty to seventy. Compared to the baseline expected life from age forty of 29.9 years, if being born during bad early life conditions the expected life form age forty would instead be 27.2 years. In terms of month this is a loss of almost 32 months. This could be seen as support for the "Fetal of Origins hypothesis" (Barker, D. (1990)), with the latent effect on survival in older ages.

Furthermore, the effect from socioeconomic status is significant and here there is a negative effect on survival for the male individuals belonging to the lowest social class, p-value equal to 0.026. This effect represents, when compared to the baseline expected life of 29.9 years, instead an expected life of 26.9 years or a loss of over 36 months. For the last covariate, being married or not, in the accelerated failure time model no significant effect is found, p-value equal to 0.261. The total accelerated failure time model is showing significance, with p-value equal to 0.032, see table 12.

Table 13 Effect from high infant mortality rate on socioeconomic status for male in age 40 to 70

The effect of IMR.high on SES.low for male age 40-70			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.413075	0.003234	< 2e-16
IMR.high	-0.048576	0.006667	3.26e-13

Table 14 Effect from high infant mortality rate on being married for male in age 40 to 70

The effect of IMR.high on Married for male age 40-70			
Covariate	Coefficient	Standard error	P-value
(Intercept)	0.902157	0.001994	< 2e-16
IMR.high	-0.015771	0.004110	0.000125

Next, the results for the analysis of the indirect effects are presented. The first regression to analyze is the high infant mortality rate on the socioeconomic status, presented in table 13. Here we find a significant effect that being born during a year with bad early life conditions gives a lower risk of belonging to the lowest social class, socioeconomic status equal to 1 in the data. Further, this gives a positive indirect effect on the survival from being born in a year with high infant mortality. The value of this effect is (-0.049*-0.107) = 0.0052. When comparing this effect with the baseline life expectancy of 29.9, this indirect effect increase the life expectancy to 30.1 years instead, calculating this value into effect in months the result is that this effect represents almost 2 months longer survival. The indirect effect here can be seen as a counter effect to the direct effect of being born during a year with high infant mortality rate and support the theory for selection (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)).

At last the regression of being born during a year with high infant mortality rate on being married or not is presented in table 14. Here, a significant effect with p-value less than 0.001 is found. This effect is representing that for the ages forty to seventy the individuals not born during a year of bad early life conditions have a higher chance of being married. Since the direct effect from the variable for marriage on the survival in the accelerated failure time model is insignificant, the indirect effect for being born in a year of bad early life conditions mediated through marriage is not considered here.

When calculating the total effect of being born during a year of high infant mortality rate the result is (-0.093 + 0.0052) = -0.0878. After calculations, the life expectancy is found to be 27.4 years and compared to the baseline life expectancy this represent a loss of 30 months for the men born during a year of high infant mortality rate in later adult life.

6. Conclusion

In this thesis the impact of early life conditions on adult survival has been studied. As a starting point the "Fetal of Origins hypothesis" (Barker, D. (1990)) has been taken. Here conditions in early life are seen as a crucial point for the adult health of an individual (Calkins, K. and Devaskar, S. (2011)) and the effect from these early life conditions can be latent for a long time (Almond, D. and Currie, J. 2011). The theory from the "Fetal of Origins hypothesis" has then been more generalized and extended and included in a pathway model. For this model the effect from the early life conditions are not only seen as a latent direct effect, but also allows for indirect effect (Bengtsson, T. and Mineau G. (2009) and Edvinsson, S. and Broström, G. (2012)). The tested model in this thesis is then based on the theory of the pathway model and described in chapter 3.1. The basis of this model is that the early life conditions can have a direct effect on the survival of the individual and also that there could be indirect effects on the survival mediated through the socioeconomic status and the marital status of the individual.

The first analysis that is performed on the data is the calculation of the weighted means in chapter 5.1, called descriptive statistics. Here, the first result found is that that the proportion of individuals born during a year with high infant mortality rate are higher within the ages forty to seventy compared to the ages twenty to forty for both sexes. The possible explanation to this, found in the theory, is the idea of the selection of the strongest individuals, also referred to as selection (Bengtsson, T. and Mineau G. (2009) and (Nelson, R. E. (2010)), and that those individuals would be the strongest once in the early adult life. Öberg, S. (2014) also discusses the possibility of this effect but in the context of individuals' height and that during periods of high infant mortality rate the surviving individuals could be taller. Another observation of the results of the weighted means is that in the older age there are a bigger fraction of individuals belonging to the lowest socioeconomic class. The explanation found to that is that when the individual gets old his or her children would inherit the family farm, but would still in reality belong to the socioeconomic class of his or her children. Support for this is found in Bengtsson, T. and Broström, G. (2009). At last the proportions of the population that are married during the different period in their life course are calculated and no strange findings can be found there at the first stage of the analysis.

Next, the analysis of the data is performed as a dynamic path analysis, for detail see chapter 3.2. This method allows for the coefficient of the covariates to vary during different periods of time and also for both direct and indirect effect, which is needed to fully analyze the pathway model. In the analysis the women and men are separated and for both sexes the adult life is divided into two time periods that are analyzed. These periods are age twenty to forty and age forty to seventy.

There is a possible limitation of the results due to the use of cohort based infant mortality rates as the indication of the early life condition for a specific individual. Here, it is impossible to distinguish between if the effect from infant mortality rate on survival is due to bad conditions during in-utero or as a newborn. Due to this limitation the results should be seen as representation for a combination of bad conditions during in-utero and infancy periods.

The first results from the dynamic path analysis are for women between twenty to forty years of age. The found result from the analysis of the direct effect on the survival finds no significant result for being born during a year of high infant mortality rate. Neither for the two possible mediating covariates, belonging to the lowest socioeconomic group or not and being married or not any significant result is found. One explanation to why the analysis of direct effects on survival within the ages twenty to forty for females did not produce any significant results is that the sample size of individuals in the analysis is too small which leads to a low number of events in the data. Furthermore, one logical explanation is that there are no differences within these ages on the survival. For the variable being born during a year of high infant mortality rate or not, this could be theoretically explained by that the effect of bad early life conditions are latent, as described by Almond, D. and Currie, J. (2011). This does not however explain the insignificant results for the variables for being married or not and the socioeconomic status. At last the results could be explained by the existence of both effects and counter effects, namely scarring and selection (Bengtsson, T. and Mineau G. (2009)). From the analysis of the results for the linear regressions of being born during a year of high infant mortality rate or not on socioeconomic status and being married or not there is support found for the theory of scarring and the counter effect of selection of the strongest individuals simultaneously (Nelson, R. E. (2010)). The result form the regression on the chance of getting married is that being born during a year of high infant mortality rate reduces the chance of getting married, which is a support for the scarring theory, that individuals born during bad early life conditions would have less attractive abilities and would have a lower chance of getting married (Edvinsson, S. and Broström, G. (2012)). The results from the linear regression on the socioeconomic status is that if a woman is born during a year of high infant mortality rate then her chance of belonging to the lowest social class during her young adult life would reduce. Here, this could be seen as a support for selection of the strongest individuals with better abilities which would lead to better achieved socioeconomic status (Bengtsson, T. and Mineau G. (2009), Nelson, R. E. (2010) and Öberg, S. (2014)).

The results for the second model in the analysis are for women in the ages forty to seventy. From the analysis of the direct effects a negative effect on survival in later adult life for those women born during a year of high infant mortality rate is found. This result shows a good support for the "Fetal of Origins hypothesis" (Barker, D. (1990)), since we here find a latent effect on survival in older adult age that was not found in the analysis of the younger age of adulthood. This result follows the theory that the effects from the early life conditions can latent for a long time (Almond, D. and Currie, J. 2011), here at least until the age of forty for women in this data set. Furthermore, the covariate for the socioeconomic status also has a significant direct effect on the survival for the women here. This allows for an indirect effect of the early life conditions mediated through the socioeconomic status. From the linear regression model of being born during bad early life conditions or not on the mediating covariate for the socioeconomic status a small significant negative effect is found. The conclusion is that if a woman is born during a year of high infant mortality rate she would have a little better chance of avoiding to belong to the lowest socioeconomic class in her later adult life. This result could be seen as support for the theory of selection of the strongest individuals where the surviving individuals would have better abilities and achieve a higher social status (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)). Next, from the analysis of the effect from being born during a year with high infant mortality rate on marriage a significant effect showing that those individuals born with bad early life conditions had a higher chance of being married after the age of forty. As discussed in chapter 4.3, with results presented in table 1, there could exit a problem with the possibility of close correlation between the socioeconomic status and marriage, here the result of those variables indicating the same result could be a sign of correlation. At last for the analysis of the female model the total effect of being born during a year of high infant mortality rate on the survival in the ages forty to seventy in calculated. Here, the total effect is equal to -0.0870. When comparing this to the baseline life expectancy for females in this model those individuals that were born during high infant mortality rate have a mean loss in life expectancy of 29 months, which were almost the same as the result for the direct effect on its own.

The conclusions for the female pathway model is that support for a latent direct effect on the survival in later adult life is found for individuals born during bad early life conditions in line with the theory for the "Fetal of Origins hypothesis" (Barker, D. (1990)). Furthermore, an indirect counter effect on survival that is mediated through the socioeconomic status that provide for some support for selection of the strongest individuals is found and some indications of selections is also found in the results for being married or not (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)).

For the male individuals, that will be discussed next, the same analysis as for women is performed, namely dynamic path analysis, for detail see chapter 3.2, with the time divided into the ages twenty to forty and then forty to seventy in the same way as for women.

At first, the results from the analysis of the males in the ages twenty to forty will be discussed. From the analysis of the direct effects no significant effect on survival for individuals born during a year of high infant mortality rate is found. This result is the same as for the result for the females in the same ages. Further, for men, significant results are found for direct effects from both the covariate for being married or not and for the socioeconomic status. This is different to the results for the females where none of these covariates were significant. Both the significant effects seem to have a very high impact on the survival for men in the ages twenty to forty and even if there are good explanations in the theory for these effects the high values, especially for marriage, may be overestimated. One reason for this is that the sample size might be too small and the proportion of events in the accelerated failure time model has randomly estimated bias. The same theoretical explanation to the non-significant result from the early life conditions as presented for the women in the ages twenty to forty, that the effect of bad early life conditions could be latent and would not be measureable until late in life, as described by Almond, D. and Currie, J. (2011), is applicable here. The next analysis performed on the data for males in the ages twenty to forty is the linear regression model with the early life conditions on the two mediating covariates socioeconomic status and being married or not. From the model for the socioeconomic status, no significant effects are found. One theoretical explanation to this is the existence of both selection and scarring effects for the individuals born during a year of high infant mortality rate and that these effects are counter effects (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)). Further, from the results for the linear regression on being married or not, a positive effect of being born in a year with high infant mortality rate is found. A possible theoretical explanation to this is the idea of natural selection (Bengtsson, T. and Mineau G. (2009), Edvinsson, S. and Broström, G. (2012), Nelson, R. E. (2010) and Öberg, S. (2014)). Here, the mechanism would then be that the individuals with more attractive abilities survive and would have a better chance of getting married and would experience the advantage of marriage on chance of surviving (Edvinsson, S. and Broström, G. (2012)). The total effect here from being born during a year of high infant mortality rate would then be the indirect effect mediated though marriage, since the direct effect was insignificant in the accelerate failure time model. This result has to be taken with caution since it is based on the result for the direct effect from marriage on the survival and that result could be biased, as discussed previously.

The second part of the analysis for the male individuals is for the ages forty to seventy. From the analysis of the direct effects the first finding is that the variable for high infant mortality shows a negative significant effect on survival. This result support the theory of "Fetal of Origins hypothesis" (Barker, D. (1990)) in the same way as for the female individuals in the data, namely that there seems to be a latent effect from being born during a year of high infant mortality rate on the survival in later adult life. Next, for the covariate socioeconomic status, a negative direct effect on the survival for the individuals belonging to the lowest social class is found, but for the variable being married or not, no significant result is found. This allows for an indirect effect from being born during bad early life conditions on survival mediated through the socioeconomic status. These results also coincide well with the results from the analysis for the women in the same ages. When analyzing the linear regression model for being born during a year of high infant mortality rate on the socioeconomic status an effect that can be seen as a counter effect supporting the theory for selection is found (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)). The mechanism explaining this would here be that the surviving individuals would have better abilities and achieve a higher social status which could increase their chance of survival. At last the effect of being born during a year of bad early life conditions indicates a lower chance of being married. This result shows support for the theory of scarring, with the mechanism that individuals born with bad early life condition would be weaker, less wanted individuals that would have a lesser chance of getting married (Edvinsson, S. and Broström, G. (2012)). As the result from the direct effect of being married or not on the survival is insignificant the indirect effect for being born in a year of high infant mortality rate on the survival, mediated through marriage, is not considered here.

The total effect for men from being born during a year of high infant mortality rate on the survival is found to be -0.0878 and is similar to the total effect for women in the same ages which were -0.0870. The total effect for men corresponds to a loss of 30 months compared to the baseline life expectancy for men and for women a loss of 29 months compare to the baseline life expectancy for women. The mediating variable that is found for both the women and the men is socioeconomic status and for this indirect effect a counter effect to the direct scarring is found. The mechanism for selection of the strongest individuals here should be seen that the strongest individuals, with higher ability to work, would survive when being born during bad early life conditions and that would lead to better achieved socioeconomic status, which then could lead to better chance to survival in later adult life (Bengtsson, T. and Mineau G. (2009) and Öberg, S. (2014)).

Although the results for the total effect of being born during a year of high infant mortality rate on the survival for men and women are similar there are differences between the sexes in the analyzed models. First, when comparing the results from the analysis for the age twenty to forty socioeconomic status and marriage have a direct effect on the survival for the men, but not for the women. For men there is a negative effect on survival from belonging to the lowest socioeconomic class and a positive effect on survival from being married. Furthermore, the effect from being born during a year of high infant mortality rate on the socioeconomic status and being married or not are different for men and women in the age twenty to forty. For women there is a positive effect from being born with bad early life conditions on the socioeconomic status, but for men there is no effect at all. Furthermore, women have a negative effect on being married while men have a positive effect on being married from being born with bad early life conditions. From the analysis for the age forty to seventy the results for men and women are generally more similar. Here the main difference is that for men the effect of being born during a year of high infant mortality rate on marriage has a positive effect while for women this effect is negative.

When comparing the results in this thesis with the results from Edvinsson, S. and Broström, G. (2012) they also find a negative effect from infant mortality rate on survival in later life, but compared to this study where it is about the same for men and women, Edvinsson, S. and Broström, G. (2012) find a stronger effect for women than men on old age survival. Furthermore, Bengtsson, T. and Broström, G. (2009) also finds a direct effect on survival in old age from the conditions in early life, but they do not find any impacts from indirect effects compared to this study where a small counter effect to the direct effect is found.

Finally, in the thesis the importance of early life conditions on the survival in adult life has been addressed and direct negative effect from being born during a year of high infant mortality rate on the survival has been found. Also the possibility of indirect effect mediated through the socioeconomic status and the marriage status has been studied. Here, counter effects, both for men and women, have been found. These counter effects could be explained by a selection of the individuals born during bad early life conditions, but it is hard to verify this completely. Also, the problem with the use of cohort based infant mortality rate has been addressed and to fully understand the mechanism behind the early life conditions, models that can distinguish between the effect of bad in-utero conditions and bad infancy conditions is needed. With the possibility of a bigger data set, better measure of early life conditions and more detailed data then even better understanding on the pathways from early life to adult health could be found.

References

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

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. (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.

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 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.

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.

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

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.

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

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.

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

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.

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.

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.

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.

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.

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.

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.