Intergenerational transfers of infant mortality in 19th century northern Sweden

Göran Broström, Sören Edvinsson, and Elisabeth Engberg

2018-02-13

Abstract

This contribution is part of an international comparative initiative with the aim to assess the analytical power of the Intermediate Data Structure (IDS) in a study of possible intergenerational transmissions of death in infancy. An evaluation of the data in applied research will be useful for further development of the IDS structure and for its future use in comparative research. An additional methodological aim for this part of the study is to evaluate and compare different models for statistical analysis of intergenerational transfers. The analysis is based on a cohort of mothers born 1826–1854, whose experiences of infant mortality are compared to the ones of the previous generation, that is, the grandmothers. Data are collected from Swedish parish records, available in the database POPUM at the Demographic Data Base in Umeå. The analysis shows a clear association between infant mortality among mothers and grandmothers. The probability of an infant death for a woman is increased if her mother also had experienced an infant death. Testing different approaches for analysis, simple models with few restrictive assumptions gave similar results as more complicated models. Since it is easy to feel confident in the models with the weakest assumptions, we argue that such models are preferred for this type of analysis.

Contents

1 Introduction

This article has a threefold aim, including one demographic query and two methodological issues. The first and main purpose is to apply the Intermediate Data Structure (Alter and Mandemakers, 2014) and the statistical modelling to a demographic study of the possible intergenerational transmission of death in infancy. More specifically, the hypothesis of no intergenerational transfer is tested. The second (and first methodological) aim is to assess the power of the IDS database on data from DDB (Demographic Data Base, Umeå University) for demographic analysis. An evaluation of the data in applied research will be useful for further development of the IDS structure and for its future use in comparative research. These two aims are common for all contributions in the present volume. The third (and second methodological) objective is however unique for our contribution. There are several considerations needed to be taken when analysing intergenerational transfers and it is therefore important to discuss alternative approaches in order to find the most fruitful one(s). We therefore evaluate and compare different models for statistical analysis of intergenerational transfers. In particular we argue in favour of using of what we call *simple and sound* models.

Access to historical micro-data has increased our knowledge of the revolution in human life, described in the demographic transition theory. We know much more about the mortality and fertility decline. Infant mortality has been thoroughly investigated in a multitude of studies representing different historical and spatial contexts. The scholarly debate on the determinants continues however. One important aspect is the important role of family dependency determining survival of children, an observation that has led to an interest in following patterns over generations. Studies of intergenerational effects require

individual data covering long periods, extending several generations, data that have not been available until the recent development of historical population databases. The study we present here is the contribution concerning northern Sweden to the comparative project described by Quaranta (ressb), where intergenerational transfers of infant mortality on the maternal side are analysed in different contexts (Donrovich et al., ress; Quaranta, ressa; Sommerseth, ress; Van Dijk and Mandemakers, ress), but with similar methods. While restricting the analysis to the female side, we still acknowledge the possible impact from the male side (Murphy and Knudsen, 2002; Pembrey et al., 2006). In our case the analysis is based on a cohort of mothers born 1826–1854, for whom their experiences of infant mortality is compared to the one of the previous generation—the grandmothers. Thus, we analyse the family patterns of infant mortality in a certain region, and during a restricted period. Even though the present study alone cannot resolve the question about the generalizability of the transmission of infant mortality in families or about possible changes over time, the standardised approach used in the project substantially increases our understanding of these issues.

Observed family dependencies has led to an increased attention to a possible intergenerational transmission of demographic patterns (Edvinsson and Janssens, 2012), sometimes even considering epigenetic effects (Pembrey et al., 2014). It is however not obvious that the transmission has been equally strong across times and between different places – this is still an open question. Concerning transmission of fertility behaviour, studies indicate that the association between generations was negligible in pre-transitional societies but became more substantial over time (Murphy, 1999). Kohler et al. (1999) suggest that the role of genetics for fertility outcomes increased in importance during later stages of the transition. In a study on the same region as in the present study, Kolk (2014) finds an association between fertility across generations and that this is only partly mediated by education or occupational class. They are instead connected to other circumstances, such as transmission of values and fertility intentions.

Studies about trans-generational effects on infant mortality are however rare. It is still an under-researched topic. There are results showing major regional differences in the strength of the intergenerational transmission of infant mortality (Vandezande, 2012). While in certain regions the mortality history of infants is strongly correlated with the survival of infants in the previous generation, in other regions this effect is weak or completely absent (Brändström et al., 2008). Vandezande (2012) suggests that this can be attributed to differences in local culture and family systems, but also proposes the hypothesis that the variations might be related to strong local variants in gene defects as well. Most studies of regional differences are however based on a limited number of rather small regions. Different studies are hard to compare due to differences in methodology, both in terms of database management and statistical analysis. Our study, together with the others in the same project, aims to contribute to an increased understanding of the familial transmission of infant mortality in different contexts.

2 The Skellefteå region

2.1 Geography and population

The area under study is the 19th century Skellefteå region in the province of Västerbotten in the northern part of Sweden. At the outset of the study it consisted of one large rural parish, Skellefteå. By 1900 three new parishes had been detached into separate units, but their populations are still included in the study until 1900. The region was large, both in terms of area and of population. With an area of about 1700 square miles, Skellefteå was considerably larger than most rural parishes in Sweden. It was considered a one-day journey to travel from the northern to the southern border, and a ride from the coast to the more remote and sparsely populated parts of the parish in the west could take even longer, especially in wintertime. The main part of the population was, however, concentrated in the coastal area and in river valleys. In the early 19th century the population size was around 6900, and it increased rapidly during the first half of the century. By 1850 it had reached to about 17 000 and at the turn of the century it had further doubled. Despite the large increase in population, which was mainly the result of a high natural growth, the population density on the whole remained low (Alm Stenflo, 1994).

2.2 Economy

Skellefteå was a rural area with a mixed economy, based on animal husbandry, forestry and sidelines such as tar and saltpeter production. By the mid-19th century export of tar and lumber became an increasingly important part of the economy. Most farmers in the region were smallholders and there were no large estates. Some small sawmills were established early in the century, but before 1900, industrialisation had little impact on the local economy. In 1835, approximately 85 percent of the population made their living from farming. Although the distribution of economic resources was more uniform than in several other Swedish regions, the social stratification became more pronounced throughout the 19th century. The increasing proletarianization was mainly a consequence of rapid population growth. The number of farming households remained fairly stable, while the number of landless households increased. The socio-economic development was also influenced by two devastating subsistence crises in the region, in the 1830s and in the 1860s (Engberg, 2005).

2.3 Fertility and mortality

Infant mortality was comparatively low during the 19th century (Edvinsson et al., 2005; Edvinsson, 2004b). Fertility was high, not only by Swedish standards, but also in European comparison and there are no indications of family planning. Total fertility fluctuated around five children per woman and, although fertility did decline during the

nineteenth century, the actual fertility transition occurred late in the region (Alm Stenflo, 1994; Coale and Watkins, 1986). The rate of illegitimacy was low in comparison with many other parts of Northern Sweden, where frequent pre-nuptial conceptions and illegitimate births were common (Brändström et al., 2002). The illegitimacy rate fluctuated between three and six per cent during the nineteenth century (Alm Stenflo, 1994).

3 Data sources

3.1 The parish records

The data in this study are based on information found in the Swedish parish records, consisting of birth, marriage, death, migration, and longitudinal catechetical registers (Nilsdotter Jeub, 1993). These registers are part of the same church book system, being the responsibility of the clergy that they were accurately kept. As they from the 18th century were used for the collection of national population statistics, the registers were intended to report births and deaths, and not only baptisms and burials.

In the catechetical registers the clergy kept continuous records of all demographic events for all individuals residing in a parish. These detailed records make it possible to follow individuals over time and also to identify their relatives and kin over time and throughout the life course. Furthermore, a rich variety of additional information, for example on occupation, was recorded. The historical parish records have been digitised by the Demographic Data Base (DDB) at Umeå University, and individual level data from more than one hundred parishes are at present available for research in the databases POPUM and POPLINK.

3.2 POPUM and POPLINK

The POPUM and POPLINK databases differ in two respects: The time dimension and restrictions of data access. The data collection in POPUM as well as POPLINK of the Skellefteå data starts in 1699 with birth records. Catechetical registers are available from 1720. The end year of the POPUM database is 1900. Since the database does not include personal data referring to now living persons, it is not subject to any legal restrictions in regard to confidentiality (Personal Data Act 1998:294). During recent years, the POPLINK database has been created with an extension of the time span to the 1950s for the Skellefteå and Umeå regions and a generational depth of up to 12 generations (Westberg et al., 2016). The information in POPLINK is subject to legal restrictions. So far, only POPUM has been converted into IDS, and hence also forms the basis of this study. Data in POPLINK will be available in IDS format in 2018.

3.3 Restrictions

The start of the mother cohort of mothers born 1826–1854 is motivated by the ambition to avoid the under-registration caused by missing death registers as much as possible but still being able to have a sufficiently large population to analyse. By having end year set to 1854 we can follow the full reproductive history of the mother cohort. Restricting it to the chosen mother cohort, we situate our study to a fairly consistent context. In the following we discuss a couple of important circumstances we have had to address when choosing birth cohort.

The special characteristics of the Swedish parish registers, makes it necessary to evaluate available information and to cautiously decide on the best way to analyse the data. Traditional family reconstitution methods usually lack population registers such as the Swedish catechetical registers. Families and life courses are created from registers of events, and children born as well as other individuals are considered alive and present in the studied area if there is no information on them having died and the family is still present in the studied area. A correct family reconstitution requires access to death registers.

Although the Swedish system of catechetical registers reported the complete population, making it possible to analyse mortality in most age groups, this is unfortunately not the case when it comes to infants. When death records are missing we have no reliable information on their death or survival, which might result in an over-estimation of infant survival. The Swedish catechetical registers otherwise provide alternative and better information about life courses for other age groups. Every individual should in principle be recorded in these registers with information on date of entrance and date of exit.

In many parishes, newborn children were not entered into these registers immediately after birth. The time elapsed between actual birth and it being entered in the register varied in time and between parishes. In previous research on Sundsvall Edvinsson (1992) found that newborn children were not immediately recorded in the catechetical registers, which had the consequence that births and infant deaths were severely under-registered for the period before 1860 when birth and death registers are missing. We have had to address the same problem for the Skellefteå region. There are flaws in the 18th and early 19th century data about infant deaths: All death registers are missing before 1815 and 1823—1830. Births are on the other hand completely recorded since birth registers are preserved from 1699 onwards. Thus, it has been necessary to evaluate the consequences of missing data and the possibility to rely on death information in the catechetical registers for the earlier time as an alternative.

For the 18th century, we conclude that it is problematic to use death information from catechetical registers for the analysis of infant mortality. An analysis of the 18th century registers in Skellefteå indicate that it could take several years before a child was entered in the catechetical register, maybe in connection to when the child was expected to participate in the yearly church examinations. Consequently we find many children being

born but with no more information on their life courses. A reasonable assumption is that children with only a birth entry and no further information, despite the fact that the rest of the family is still resident in the parish, died quite young. This cannot be established with complete certainty. Furthermore we do not have any date and consequently no age for the assumed death. These problems are however good illustrations of the need to carefully evaluate the sources and what implications possible deficiencies may have on the analysis.

Hence the analyses in this paper are restricted to the 19th century, when less problems of this sort are expected. It includes the reproductive history of women born 1826–1854, the information on the reproductive history of their mothers, and the deaths of all infants involved. The start of the mother cohort is motivated by the ambition to avoid the under-registration caused by missing death registers as much as possible. There are no problems with information on children in this cohort. Furthermore, by setting the end year set to 1854 we can follow the full reproductive history of the mother cohort. However, there is a possible under-estimation of infant deaths in the grandmother generation. The children of that generation can potentially have been born in periods with missing death registers, something we need to evaluate.

We have made some checks and found that there certainly are an under-registration of deaths before 1815 and the period 1823-1830. For the period before 1815, this is no real problem since only a very small proportions of siblings in the mother cohort were born at that time. There are however more babies born during the second period. A check of ages with a concentration of under-registration shows that it took place primarily during the first three months, that is, for a much more restricted age window compared to what we find for the 18th century. We have checked if the results would change if we instead consider all infants with only a birth record as having died during infancy, but that approach does not change the main conclusions. The unrecorded infant deaths represent a fairly small part of all births.

The rather short time window used here have some methodological advantages: Restricting it to the chosen mother cohort, we situate our study to a fairly consistent context, and the risk of finding some women appearing as grand-mothers, mothers as well as children is small.

4 Implementation of IDS on POPUM

The implementation of the IDS format at the DDB has been performed on the relational database POPUM. POPUM consists of a large amount of variables, whereof some represent the literal information from the sources, while others present updated information. The structure of POPUM differs from the one of IDS, so implementing IDS has consisted of a major restructuring of the data. DDB was one of the pioneers in the work of testing and implementing IDS, the first version was created in 2009. Thereafter, new and improved

versions have been developed. In this study we use version 3 of IDS, which is based on a sub-section of available parishes in POPUM: Skellefteå parish from the early 18th century until 1900. Quite a lot of the variables required in the IDS have already been constructed in POPUM and much of the problems with inconsistent and fuzzy data, as well as duplicate information, have been taken care of. These variables are constructed from a set of well-defined and documented rules. Furthermore, record linkage as well as relational linkages have already been performed, and tests have proven the linkage process to be of high quality (Wisselgren et al., 2014).

The transformation of the data from POPUM to IDS has been quite straight-forward. This does not mean that the available information is without problems. The sources sometimes contain missing data, which can be a problem when it comes to analysis. For this study, the missing data due to death records not being preserved, as mentioned above, is of course problematic.

The context information on linking individuals to their households causes some problems. Families and households are presented together in the catechetical registers. We find the family on the same page in the registers with an empty line separating it from the next family. Households are however difficult to identify. Servants and others are usually presented at the bottom of the page, and if there are several families/households on the same page it is not always obvious to which family these individuals belong. Furthermore, the separation of families/households can also be difficult, particularly in urban environments where migrations were frequent making the registers more difficult to interpret. The information therefore primarily represent the nuclear families as households. This is a complicating circumstance that DDB has identified when implementing IDS, but it does not interfere or cause any problems with the present analysis.

5 Statistical modelling

The main objective of this project is to apply some common models to data sets from different data bases, all stored in the IDS format. For this purpose, it was desired to use some complicated models to test the strength of the IDS format. The scientific objective, on the other hand, is best served by much simpler statistical models. Therefore, we split the analyses into two categories, (i) those performed by all teams, and (ii) those specific to the Skellefteå data.

5.1 Common models

For the common part, a proportional hazards model is applied for estimating the effect of the number of grandmother's infant deaths on the age at death of mother's children.

The Cox regression approach suffers from a severe dependent-data problem: Each infant to a mother is treated as an independent observation, neglecting the fact that siblings

tend to be very similar. By summing over mothers, using her number of infant deaths as response, this dependency is eliminated, as in the Poisson regression approach. However, on the next level, groups of mothers in the data set have a common mother, that is, they are sisters, creating dependency problems in both approaches. Ideally we should have unique grandmother–mother pairs, in the sense that each woman appears only once, either as a grandmother or a mother. We test the possible effects of dependency in the data by running shared frailty models and implementing robust variances. Frailty models (Hougaard, 2000) are often used in attempts to handle dependencies of this kind in survival analysis, but they come at a price of complicated validity problems. Aggregation, when applicable as here, is preferred.

In addition, a problem is that of an asymmetry between mother and grandmother: For the grandmother only, we know that at least one female child survived and got own children. An attempt to overcome all those problems is made by the "family tree" approach, presented next.

5.2 Simple and sound models

A simple model makes as few assumptions as possible about the data-generating process in the real-world problem. For instance, a non-parametric model makes fewer assumptions than a parametric one. A sound model focuses on the target in an unbiased (as much as possible, unbiasedness is not necessarily an important property) and efficient (for instance, small variance) way. A "simple and sound" model has both these properties. In our study, the outcome target is death before age one, and a simple model focuses on that target, that is, a binary (logical) response variable equal to *TRUE* if the subject dies before her/his first birth-date and *FALSE* otherwise. The total number of deaths aggregated for a mother can then be approximated by a Binomial or Poisson distribution. In contrast, a not-so-simple model, for instance Cox regression, aims only indirectly at the target: Instead, it compares hazard rates of mortality under the assumption of proportional hazards, which is unnecessary under the simple model. On the other hand, Cox regression can deal with the problem of right censoring, that is, if follow-up of an infant's life trajectory during its first year of life is interrupted. If this is a common state of affairs, it is necessary to adopt methods allowing for right censoring and paying the price of stricter assumptions. At least in our parishes, moving out of the region with infant(s) was very rare.

A compromise is the more direct approach of estimating the effect of the number of grandmother's infant deaths on the number of infant deaths for mother is taken by applying *Poisson regression*.

The negative effects of the dependent-data problem in the proportional hazards and Poisson regression approaches may be reduced by (1) building family trees (based on women), and (2) only study *male* infant mortality.

In this approach, the first mission is to operationalize and interpret the concept of "intergenerational transfer". There are essentially two ways to think about it.

1. There is a causal link from the IMR of a woman to her daughter. That is, if a woman experience a fatality, the probability that her daughter will experience one increases (or decreases). For such a relation to be reasonable, the woman's fatality must come before (in calendar time) her daughter experience one.
2. There is no direct causal link as described above, but a mother and her daughter share some common properties, they are more or less prone to have fatalities or not: Either no one or both.

In statistical terms, 1. implies a regression model, while 2. would best be described by a correlation structure. We think that 2. is the most appropriate way to think about it, but for the sake of testing the extraction of data via the IDS format, both implied models will be considered.

When we compare a woman's experience of death of her infants to her daughter's we must take into consideration the fact that by design we *know* that the woman has at least one surviving child, the daughter we are comparing her to. This induces an asymmetry that must be resolved. We can do it by considering only male infant deaths.

A second issue is to avoid dependencies in the data. Each woman, whether mother or daughter, should be present in the data set at most once. And since data are inherently paired (grandmother and mother), siblings among daughters should not be allowed.

6 Data

We use the data set that is created from the IDS database with a standard extraction script (Quaranta, ressb). The possible response variables are **Age at death** (proportional hazards regression modelling) and **Death before age one** (Poisson regression modelling). In both cases, the explanatory variables of interest are **gmBirths** (grandmother's number of births) and **gmDeaths** (grandmother's number of infant deaths). In all cases, the following confounders were included in the analysis:

- **mBirthdate** Mother's birth date expressed as years between 1 January 1 and her day of birth, minus one.
- **mAge** Mother's age at the birth of a child. Categorised into the intervals "15-24", "25-34", "35-49".
- **parity** The birth order of a mother's child.
- **childBirthdate** The birth-date of a mother's child.
- **sex** The sex of a mother's child.

Table 1: Generation 0.

id	m_id	generation	birthdate	births	deaths
10000250	NA	0	1785-12-06	2	0
10002535	NA	0	NA	6	0
10005200	NA	0	1792-10-15	7	1
10018244	NA	0	1811-03-15	9	2
10020340	NA	0	1774-06-22	12	1
10021499	NA	0	1796-04-21	9	1

These variables are never explicitly reported in the analyses, but they are always present as confounders.

6.1 The common approach

The above-mentioned data set was used as is, with the exception that mothers were limited to have been born in the time interval 1826–1854.

6.2 The simple-and-sound approach

Our data extraction is not well suited for a strict analysis with a minimum of dependency structures. What would be needed are *family trees* (or rather trees of *women*), and for each node in the tree we want the number of male and female births and male and female infant deaths. However, in order to simplify the tables below, only the total numbers of births and deaths are presented.

We must first find the "atom mothers", that is, women without mothers (in our data set). Since each "MotherID" also has a "GrandmotherID", we find the "atoms" as those grandmothers who are not present as mothers in the common data set.

So we have 1591 atom mothers, and now we want to find their children. Unfortunately, this information is not present in this IDS extraction, we only have their daughters who survive and have children. However, we have information on the number of male and female births and male and female infant deaths for each atom, and we can choose the first female birth (a "Mother") as a continuation of the tree.

So, first we build generation 0 (first six rows shown, Table 1):

Then we search for the daughters (with births) of generation 0, generation 1 (first six rows shown, Table 2):

For each woman in *gen1* we put on information from her mother about number of male births and deaths, see Table 3.

Table 2: Generation 1.

id	m_id	generation	birthdate	births	deaths
10006071	16819628	1	1818-03-26	8	0
10006997	14038034	1	1792-10-16	7	1
10016829	11433628	1	1815-01-30	8	0
10019731	17254223	1	1858-04-20	3	0
10026879	16917342	1	1829-05-19	6	0
10028884	16989942	1	1864-02-22	5	0

Table 3: Mother/Grandmother pairs, final table. First six rows shown.

birthdate	births	deaths	m_births	m_deaths
1818-03-26	8	0	13	4
1792-10-16	7	1	6	0
1815-01-30	8	0	6	0
1858-04-20	3	0	8	0
1829-05-19	6	0	6	0
1864-02-22	5	0	5	0

Note that in this approach we do not include any of the confounders described above, and there is no restriction of mother's birth date, which is the case in the common analyses. However, for comparative reasons, the restriction is imposed later, see the results.

We can now test the effect of mother's infant deaths on the probability that her daughter will experience an infant death.

7 Descriptive statistics of the IDS extraction

The yearly numbers of births and infant deaths for the *mothers in the data set* and, as a comparison, for the data in POPUM, are shown in Figure 1. The number of births in the data set is 36 232, and the number of infant deaths is 3011, constituting an overall infant mortality rate of 8.3 per cent.

The difference between the two data sets is of course explained by the fact that in the present data file there are restrictions on which births to include: Mother and grandmother present, grandmother must have at least two children, etc.

The average number of births to a mother in the data set is 4.87. The corresponding number for a grandmother is 7.44, but keep in mind that grandmothers are restricted to have at least two births, while mothers are restricted to have at least one birth. Our study sample consists of all mothers born 1826–1954. The distribution of their birth

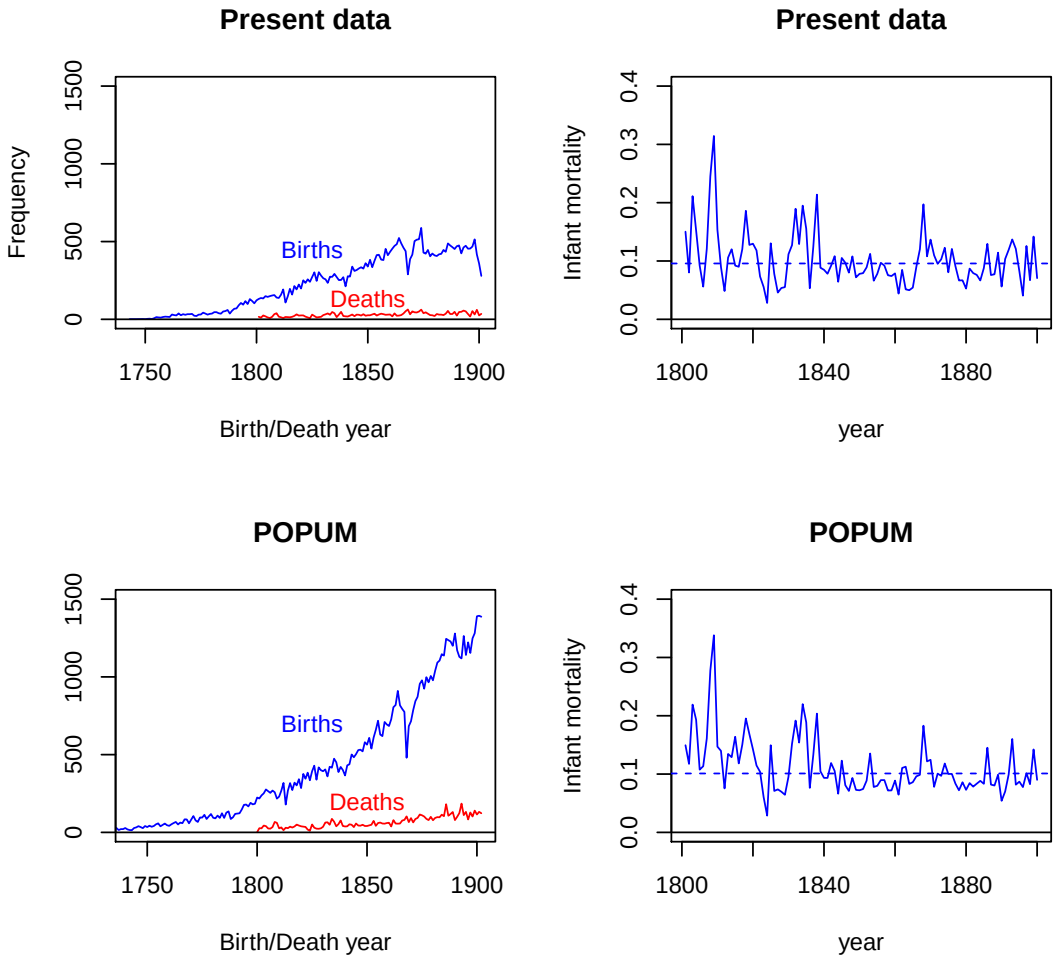

Figure 1: Number of births and infant deaths and infant mortality by year.

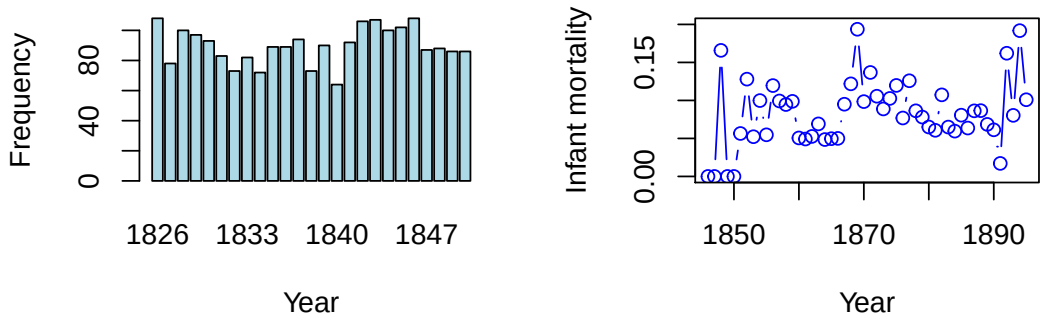

Figure 2: Distribution of mother birth years and mothers' infant mortality rate by infant death year.

years and their infant mortality (by year of child death) are shown in Figure 2.

7.1 Grandmothers, mothers and mother–sisters

There are 2247 mothers and 1384 grandmothers in the data, so obviously there are many sister groups among mothers in the data. This fact induces dependencies in the data set, which may either be a problem (using methods assuming independence), but it may also be possible to turn this fact into an advantage (using mixed effects models and think of intergenerational transfer as similarity between siblings). In the latter case the explanatory variable *gmIMR* is replaced by clustering on grandmother.

The distribution of the sizes of sibling groups is shown in Figure 3.

How many grandmothers are also mothers (and vice versa)? The answer is 23, or 1.66 per cent of the grandmothers. This small amount is of no practical importance for the results.

8 Results, common models

The first result section, with the standard models, reports the analyses that are the same for all the different teams, that is, with survival analysis. In the second section, with the extended models, alternative approaches and models are investigated.

In the analyses presented here we are using **R**, a free software environment for statistical computing and graphics (R Development Core Team, 2017). The analyses and report

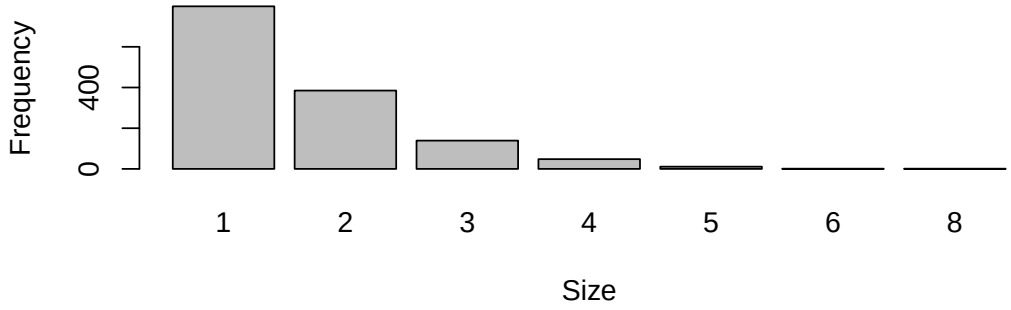

Figure 3: Distribution of sister group sizes.

Table 4: Analysis of deviance: (1) Null model, (2) gmBirths added, and (3) gmDeaths added. Cox regression.

	loglik	Chisq	Df	P(>|Chi|)
1	-6172.601	NA	NA	NA
2	-6168.435	8.3322	3	0.0396
3	-6162.499	11.8721	2	0.0026

writing were performed in *RStudio* (RStudio Team, 2018) with the aid of the **R** package *knitr* (Xie, 2017, 2015).

8.1 Survival analysis

The **R** (R Development Core Team, 2017) packages *eha* (Broström, 2017, 2012), *survival* (Therneau, 2017; Therneau and Grambsch, 2000), and *coxme* (Therneau, 2018) are used, and the explanatory and potentially confounder variables are the same as in the Poisson regression analysis. The statistical significance of a Cox regression is shown in Table 4: The first row is the null model without *gmBirths* and *gmDeaths*, the second with the variable *gmBirths* added, and the third with *gmDeaths* added as well. The small p-values show that both variables are needed in the model, especially that *gmDeaths* is badly needed even after *gmBirths* is included.

Second, the sizes of the effects are shown in Figure 4. For our prime target, the covariate *gmDeaths*, the real effect on her daughter's risk of experiencing infant deaths appears when the number of deaths are two or larger. The variable *gmBirths* is also influential.

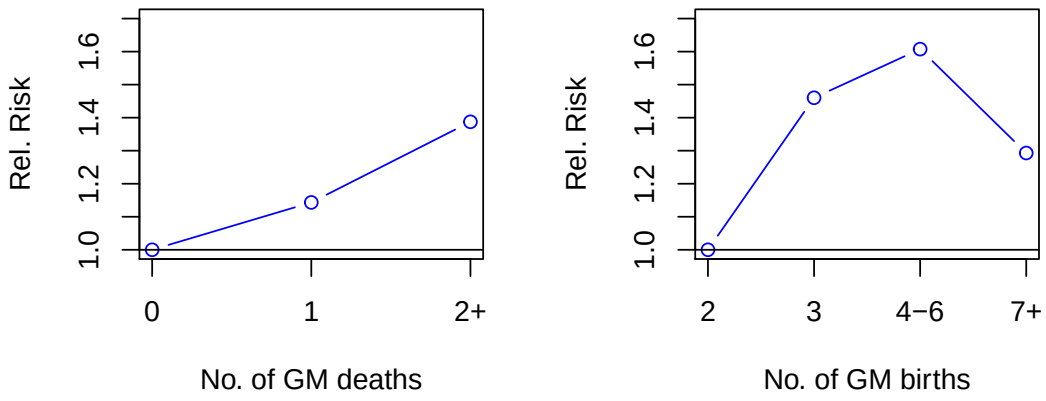

Figure 4: Effects of included covariates, Cox regression. The leftmost value is the reference in both panels.

8.1.1 The proportional hazards assumption

The estimated cumulative hazards functions for the strata of *gmDeaths* are shown in Figure 5.

There is an evident deviation from the assumption of *proportional hazards*: However, this does not disturb the main conclusion too much: Two or more grandmother infant deaths are harmful. The reported p-values are cast in doubt, however.

On the other hand, mortality before two months of age seems to be equally affected by any number of grandmother's infant deaths, and after two months, there is no increased risk if the number of grandmother's infant deaths is exactly one. This may indicate that there are two competing risk processes connecting mother and daughter.

Among confounders, both **parity** and **mother's age** show clear non-proportional effects. They are therefore included as stratum variables in the Cox regressions.

8.1.2 Dependency structures

There are a couple circumstances that introduce dependence structures in the data: Infants being siblings share genetic and environmental (unmeasurable) properties, some mothers have sisters that themselves are present as mothers, thus sharing grandmother. Possible ways of handling the situation are *shared frailty models* (Aalen et al., 2008) and the implementation of *robust variances* (Therneau and Grambsch, 2000). We have tried

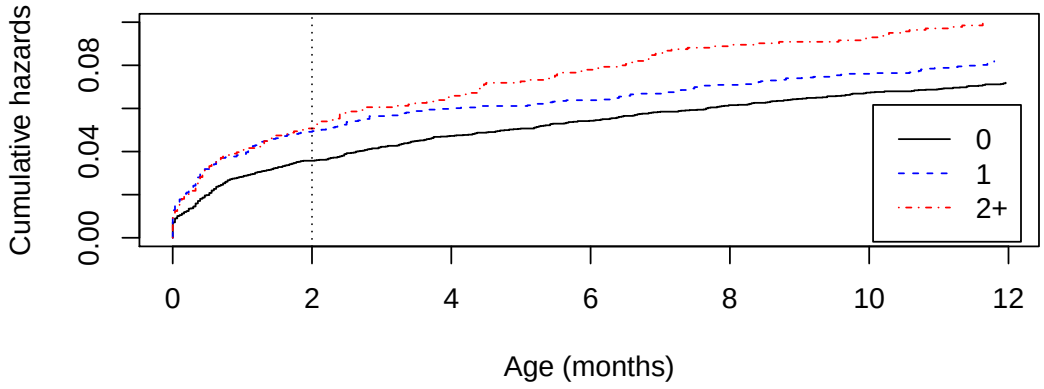

Figure 5: Cumulative hazards by the number of grandmother's infant deaths.

Table 5: Analysis of deviance: (1) fixed effects model, (2) random effects model. Cox regression.

	loglik	Chisq	Df	P(>|Chi|)
1	-6852.790	NA	NA	NA
2	-6833.647	38.2843	1	6.12e-10

both, but neither do change the results in any noticeable way. For the shared frailty model with Gaussian random effects we get a very strong effect of the grouping of infants by mother, see Table 5.

The estimated fixed effects coefficients are shown in Figure 4.

Our main hypothesis, there is no intergenerational transfer, is still rejected, and the estimated effect sizes are essentially the same.

9 Simple-and-sound results

Findings that are specific for the Skellefteå project are presented here. First we look at the simple-and-sound model based on the family trees with both female and male infant mortality.

The simple data can be summarised in a (collapsed) contingency table (Table 6).

A standard χ^2 test of independence in Table 6 results in a p-value of 0.06, not very

Table 6: Contingency table of mother's (mD) and grandmother's (gmD) infant deaths, collapsed.

	gmD = 0	gmD = 1	gmD = 2+	Sum
mD = 0	1415	320	154	1889
mD = 1	372	106	48	526
mD = 2+	153	32	27	212
Sum	1940	458	229	2627

Table 7: Excess number of cases compared to independence.

	gmD = 0	gmD = 1	gmD = 2+	Sum
mD = 0	20	-9	-11	0
mD = 1	-16	14	2	0
mD = 2+	-4	-5	9	0
Sum	0	0	0	0

spectacular. In order to get a hint of the direction of the association, a look at the *raw residuals* (observed minus expected number of cases) is illuminating, see Table 7. The main diagonal is "heavy positive", implying that mothers and grandmother share infant mortality experiences in a higher degree than expected, were there no association at all.

Second, the idea of only looking at male infant mortality (for "balance" reasons discussed earlier) was also executed, but the result was not very illuminating: Essentially the same effects were found, but with weaker statistical significance, which is a natural consequence of the fact that half the sample of infants was lost.

Third, restricting to mothers born 1826–1854 gives Table 8. In order to calculate risks of experience *any* infant death for a mother, we calculate proportions from Table 8, shown in Table 9.

From Table 9 we see that the risk of experiencing *any* infant death increases from 27 to 33 per cent if grandmother has experienced an infant death.

Table 8: Contingency table of mother's (mD, 1826–1854) and grandmother's (gmD) infant deaths, collapsed.

	gmD = 0	gmD = 1+	Sum
mD = 0	261	174	435
mD = 1+	95	84	179
Sum	356	258	614

Table 9: Contingency table of mother's (mD, 1826–1854) and grandmother's (gmD) infant deaths, per cent.

	gmD = 0	gmD = 1+
mD = 0	73	67
mD = 1+	27	33
Sum	100	100

Table 10: Analysis of deviance with (1) and without (2) gmDeaths. Poisson regression.

	Resid. Df	Resid. Dev	Df	Deviance
1	2242	2587.79	NA	NA
2	2240	2575.77	2	12.02

9.1 Poisson regression

The expected value of the number of infant deaths D_i for mother No. i, $i = 1, \ldots, n$, is modeled by a Poisson distribution as

$$E(D_i) = R_i e^{\boldsymbol{\beta}\mathbf{x}_i},$$

where R_i is total risk time for mother No. i, $\mathbf{x}_i$ a vector of her explanatory variables, and $\boldsymbol{\beta}_i$ is the vector of regression coefficients. (For a mother with no infant deaths, the risk time is equal to her number of births.) Formally, R_i is entered into the model as an *offset* after taking logs.

The results are presented in two steps: First, *the statistical significance* is calculated and shown in Table 10: The analysis-of-deviance shows a deviance of 12.02 with two degrees of freedom, corresponding to a p-value of 0.0025, implying a strong statistical significance.

Regarding the practical significance, see Figure 6.

The size of the effect of grandmother's infant mortality on mother's is not large, but, maybe surprisingly, the effect of her number of births is slightly larger.

10 Discussion

In this article we have analysed possible intergenerational transfers of infant mortality. An important part of our work is to discuss how to best approach the question. We discuss some principal choices when modelling the analysis. To be strict, we ought to take care of the asymmetry between the mother and grand-mother generation. The potential problem motivating the strategy is that, given everything else is equal, a grandmother

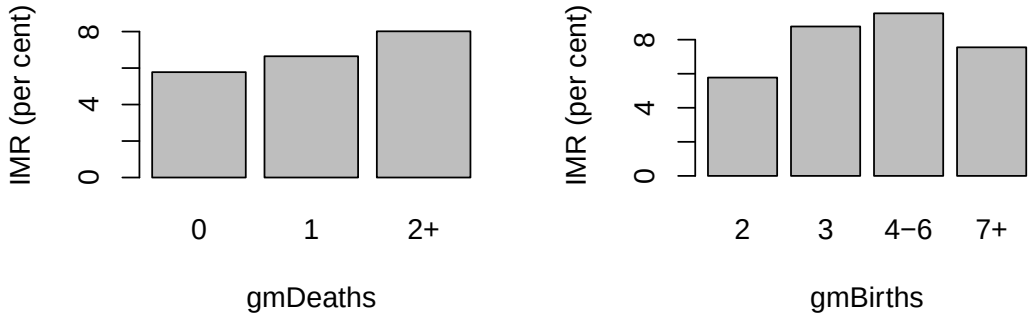

Figure 6: Infant mortality by grandmother's number of infant deaths (left) and grandmother's number of births(right). Comparisons made at reference value of other covariates. Poisson regression.

will on average have fewer births than her daughter, since we must subtract one from her number of births. This can be solved by restricting the analysis to male infant mortality (thereby not having to subtract a birth from the total number of births of grand-mothers). The disadvantage is that we have a much smaller population to investigate, requiring a substantial large population and number of births from the beginning. Our analysis showed however that the less accurate approach gives similar results as the "correct" one, thus we can feel confident when choosing the less optimal solution. We have also tried different models, all plausible, reaching similar results. Therefore we feel comfortable when arguing for "simple and sound" models, based on as few assumptions as possible.

The main result is that the null hypothesis of no association was rejected: The risk for a woman to experience infant deaths is increased if her mother also had that experience. While statistically significant, the effect size (18 per cent) is moderate. We applied different models, all plausible, and got very similar results.

Our study confirms the existence of family dependency in infant mortality risks. How can the association be explained? This relates to the old debate of whether humans are primarily formed by nature or if nurture is important, a division that now is becoming muddled and partly obsolete due to recent scientific developments (Meloni, 2014). In the analytic approach taken here, we cannot really distinguish between purely biological versus social and cultural factors, but there are some plausible alternatives.

Obvious possible factors refer to genetics. People are differently frail and it is reasonable that this frailty is genetically transferred. This, as well as conditions related to the mother during and after delivery, might affect children's capacity of survival. There are also other possible biological reasons. The Rh negative blood group is common in the Skellefteå

region, thus increasing the risk for Rh disease. It has been shown that this condition did have some impact on perinatal health in the region, although on a fairly restricted level, but it can still be a component in the observed association (Häggström Lundevaller and Edvinsson, 2012). Furthermore, a mother and her sibling group may have experienced problematic conditions in early childhood that may have scarred her for life, making her biologically less fit for child birth and thereby increasing the mortality risks for her children (Quaranta, 2013).

Can the survival of children have to do with transferred behaviours between generations? We know that the ways children are taken care of is crucial for their health and survival, particularly in a high mortality regime. This may relate to local patterns of childcare, for example to the practice of breast-feeding, but also to practices that are transferred within families. Our immediate family is the closest social organisation for learning. In 19th century Sundsvall, mothers moving to the region kept their childcare practices from their home region in the new environment. Mothers from a high mortality region experienced a higher infant mortality and vice versa (Edvinsson, 2004a). Another explanation for the association between generations is that they merely reflect similar living conditions, for example the physical environment they lived in or the belonging to the same social class. This could be more thoroughly analysed in models where residence and/or social position are considered.

11 Concluding remarks

Three main aims were set out. The first was about assessing the usefulness of the IDS data format for demographic analysis, in this case in relation to transnational comparative research. We can confirm that applying the IDS format, and the general script made by Quaranta (ressb) has fulfilled its purposes well. The use of the script has facilitated comparisons between countries. However, it is important to remember that the IDS format do not relieve the researcher from the duty of carefully evaluating the sources and their usefulness for the research questions raised. In the database for the Skellefteå region, we had to restrict our studied period substantially, due to deficient data for the 18th century, where infant deaths were severely underestimated. This problem can eventually be taken care of by applying clear definitions of rules on how to identify infant deaths, but that must be performed in a preparatory state.

Second, we have tried different analytic models, ranging from complicated (proportional hazards models for survival times of infants) to simple (had a woman and/or her daughter an infant death or not: A contingency table analysis). "Complicated" refers to "strong assumptions and strict model checking": The hard part is dealing with the complicated dependency structures arising from sisters having the same mother and the dependency between life lengths of sibling infants. On the other hand, "simple" means almost no restrictive assumptions, no complicated dependency structures, and results that are directly interpretable in terms of simple-to-understand probabilities. As is often the case,

simple is at least as good as complicated: The different models give similar results, and it is easy to feel confident in the ones with the weakest assumptions. On the other hand, it is generally good (statistical) practice to try different approaches to the analysis of the same problem. It may be called *sensitivity analysis* (Cox and Oakes, 1984, chapter 6).

Third, when it comes to the results, we find a clear association between infant mortality among the mothers and the grandmothers. Another question is if this is of practical significance. The simplest model is the easiest to interpret: If grandmother has no infant deaths, then the probability for mother to have one is 27 per cent, otherwise it is 33 per cent.

References

Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and Event History Analysis, A Process Point of View*. Springer, New York.

Alm Stenflo, G. (1994). *Demographic description of the Skellefteå and Sundsvall regions during the 19th century*. Demographic Data Base, Umeå University, Umeå, Sweden.

Alter, G. and Mandemakers, K. (2014). The intermediate data structure (IDS) for longitudinal historical microdata. *Historical Life Course Studies*, 1:1–26.

Brändström, A., Edvinsson, S., and Rogers, J. (2002). Illegitimacy, infant feeding practices and infant survival in Sweden, 1750–1950: A regional analysis. *Hygiea Internationalis*, 3(1).

Broström, G. (2012). *Event History Analysis with R*. Chapman & Hall/CRC, Boca Raton. ISBN 9781439831649.

Broström, G. (2017). *eha: Event History Analysis*. R package version 2.5.1.

Brändström, A., Edvinsson, S., Lindkvist, M., and Rogers, J. (2008). Clustering across generations: A comparative analysis of infant mortality in 19th century sweden. ESSHC Conference in Lisbon, Portugal, 26 February – 1 March 2008.

Coale, A. J. and Watkins, S. C. (1986). *The decline of fertility in Europe*. Princeton University Press.

Cox, D. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman & Hall/CRC, London.

Donrovich, R., Puschmann, P., and Matthijs, K. (in press). Mortality clustering in the family: Fast life history trajectories and the intergenerational transfer of infant death in late 19th- and early 20th-century Antwerp, Belgium. *Historical Life Course Studies*.

Edvinsson, S., , Brändström, A., Rogers, J., and Broström, G. (2005). High-risk families: The unequal distribution of infant mortality in nineteenth-century Sweden. *Population Studies*, 59:321–337.

Edvinsson, S. (1992). *Den osunda staden: sociala skillnader i dödlighet i 1800-talets Sundsvall*. PhD thesis, Umeå universitet.

Edvinsson, S. (2004a). Att ha omsorg om sina barn. Traditionens och miljöns roll för spädbarnsdödlighet. In Brändström, A., Edvinsson, S., Ericsson, T., and Sköld, P., editors, *Befolkningshistoriska perspektiv. Festskrift till Lars-Göran Tedebrand*. Demographic Data Base. Swedish.

Edvinsson, S. (2004b). Social differences in infant and child mortality in nineteenth century Sweden. In Breschi, M. and Pozzi, L., editors, *The Determinants of Infant and Child Mortality in Past European Populations*. Forum, Udine.

Edvinsson, S. and Janssens, A. (2012). Clustering of deaths in families—infant and child mortality in historical perspective. *Biodemography and Social Biology*, 58(2):75–86.

Engberg, E. (2005). I fattiga omständigheter. Fattigvårdens former och understödstagare i Skellefteå socken under 1800-talet. Umeå: Demographic Data Base. Swedish.

Häggström Lundevaller, E. and Edvinsson, S. (2012). The effect of the Rh negative disease on perinatal mortality. Evidence from Skellefteå 1840–1900. *Biodemography and Social Biology*, 58:2:116–132.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag, New York.

Kohler, H.-P., Rodgers, J. L., and Christensen, K. (1999). Is fertility behavior in our genes? findings from a danish twin study. *Population and development review*, 25(2):253–288.

Kolk, M. (2014). Understanding transmission of fertility across multiple generations–socialization or socioeconomics? *Research in Social Stratification and Mobility*, 35:89–103.

Meloni, M. (2014). How biology became social, and what it means for social theory. *The Sociological Review*, 62:593–614.

Murphy, M. (1999). Is the relationship between fertility of parents and children really weak? *Social Biology*, 46(1-2):122–145.

Murphy, M. and Knudsen, L. B. (2002). The intergenerational transmission of fertility in contemporary denmark: The effects of number of siblings (full and half), birth order, and whether male or female. *Population Studies*, 56(3):235–248.

Nilsdotter Jeub, U. (1993). Parish records. 19th century ecclesiastical registers. Umeå: Demographic Data Base.

Pembrey, M., Saffery, R., Bygren, L. O., et al. (2014). Human transgenerational responses to early-life experience: Potential impact on development, health and biomedical research. *Journal of Medical Genetics*, 51(9).

Pembrey, M. E., Bygren, L. O., Kaati, G., Edvinsson, S., Northstone, K., Sjöström, M., and Golding, J. (2006). Sex-specific, male-line transgenerational responses in humans. *European Journal of Human Genetics*, 14(2):159–166.

Quaranta, L. (2013). *Scarred for life. How conditions in early life affect socioeconomic status, reproduction and mortality in Southern Sweden, 1813–1968*. PhD thesis, Lund University. Lund Studies in Economic History 59.

Quaranta, L. (in press:a). Intergenerational transfers in infant mortality in southern Sweden, 1740–1968. *Historical Life Course Studies*.

Quaranta, L. (in press:b). Program for studying intergenerational transmissions of infant mortality using the intermediate data structure (IDS). *Historical Life Course Studies*.

R Development Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2018). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

Sommerseth, H. L. (in press). Intergenerational transfer in infant mortality in northern Norway during the 19th and early 20th century. *Historical Life Course Studies*.

Therneau, T. (2017). *survival: A Package for Survival Analysis in S*. R package version 2.41-3.

Therneau, T. M. (2018). *coxme: Mixed Effects Cox Models*. R package version 2.2-7.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.

Van Dijk, I. and Mandemakers, K. (in press). Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833–1912. *Historical Life Course Studies*.

Vandezande, M. (2012). *Born to die. Death clustering and the intergenerational transfer of infant mortality, the Antwerp district, 1846–1905*. PhD thesis, Faculteit Sociale Wetenschappen, Katholieke Universiteit, Leuven.

Westberg, A., Edvinsson, S., and Engberg, E. (2016). A unique source for innovative longitudinal research: The POPLINK database. *Historical Life Courses*, 3:20–31.

Wisselgren, M., Edvinsson, S., Berggren, M., and Larsson, M. (2014). Testing methods of record linkage on swedish censuses. *Historical Methods*, 47:138–151.

Xie, Y. (2015). *Dynamic Documents with R and knitr*. Chapman & Hall/CRC Press, Boca Raton. ISBN 978-1466561595.

Xie, Y. (2017). *knitr: A general-purpose package for dynamic report generation*. R package version 1.17.