Clustering and intergenerational transfers of infant mortality in 19th century northern Sweden

Elisabeth Engberg, Sören Edvinsson, and Göran Broström

January 16, 2017

# Abstract

This study is part of an international comparative initiative where the same analyses are performed on data sets from different regions, all available in IDS format. It is shown that the IDS data structure indeed can deliver qualified data for demographic analysis. Moreover, we show that the propensity of infant death for a mother is inherited from mother to daughter.

# Introduction

This article has a threefold aim, including two methodological objectives and one demographic query. The first methodological aim is to assess the power of the IDS data format 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. The second methodological objective is to evaluate and compare different models for statistical analysis, in this case of inter-generational transfers. The third purpose is to apply the IDS structure and the statistical modelling in a demographic study of a possible inter-generational transmission of mortality in infancy.

For more than 25 years it has been observed in numerous studies of historical and contemporary populations that infant deaths seem to be clustered into high-risk families. In a study of infant mortality in two regions in 19th century Sweden, Skellefteå and Sundsvall, Sören Edvinsson et al. (2005) found that infant deaths were not uniformly distributed; more families than expected had no infant death or two or more infant deaths. Despite the fact that infant mortality rates shifted over time, from about 200 per thousand in the beginning of the nineteenth century to approximately 100 per thousand in the 1890s, most families never experienced an infant death. Today there seems to be consensus that the non-uniform distribution of infant deaths among families is a complex interplay of different factors, difficult to identify and separate. The observed family dependency has led to an increased attention to a possible inter-generational transmission of demographic patterns (S. Edvinsson and Janssens 2012), sometimes even considering epigenetic effects (Pembrey, Saffery, and Bygren 2014). One should however be aware of the fact that clustering emerges naturally from pure randomness, often very counter-intuitive. This is discussed by Holmberg and Broström (2012).

Although the clustering of infant mortality appears to be a salient phenomenon throughout history, there are also major regional differences in the strength of the inter-generational transmission of infant mortality. 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 also might be related to strong local variants in gene defects. In practice, regional differences could also be related to the fact that most studies focus on a limited number of rather small regions, and that different studies thus are hard to compare due to differences in methodology, both in terms of database management and in terms of statistical analysis.

# Area

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).

Skellefteå was during the studied period 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. The majority of the farmers in the region were smallholders and there were no large estates. Some small sawmills were established early in the century, but before 1900, industrialization 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).

Infant mortality was comparatively low. 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 district (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. The illegitimacy rate fluctuated between three and six per cent during the nineteenth century (Alm Stenflo 1994).

# Data sources

The analysis is based on information found in the Swedish church book records, that is, birth, death, migration, and catechetical registers (Nilsdotter Jeub 1993). 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 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. From the 1970s, these records have been digitized by the Demographic Data Base 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](http://www.cedar.umu.se/).

In the databases, all entries in the catechetical registers and in the birth, death, marriage and migration registers have been linked, enabling life course analyses of the population as long as individuals are present in the digitized regions. During recent years, the database has been extended with parish registers until the 1950s for the Skellefteå and Umeå regions. Data until 1950 are available in the anonymized data base POPLINK with a generational depth of up to 12 (Westberg, Edvinsson, and Engberg 2016). The registration in POPLINK of the Skellefteå data starts in 1699 with birth records. The catechetical registers are available from 1720. Unfortunately there are flaws in the 18th century data about infant deaths. Children that died young are under-reported in the 18th century catechetical registers, and the death registers are missing before 1815. Hence the analysis in this paper is restricted to the 19th century. It includes the reproductive history of women born 1826--50 and the association with the reproductive history of the previous generation. The public IDS database is extracted from the POPLINK database, but only until 1900 due to legal requirements of protection of privacy.

# Implementation of IDS

It was easy to utilize the IDS data base with the aid of Quaranta's (2016) Stata script. The only drawback was that the script required , which is a quite expensive piece of software. Luckily, Stata provides the possibility to try the software for free during one month, and that was more than enough time to get the script running. However, this is of course not a sustainable state of the matter: Efforts should be made to utilize less costly approaches.

Once we had the data, the analyses and report (this one) writing was performed in *RStudio* (RStudio Team 2016) with the aid of the **R** package *knitr* (Xie 2016; Xie 2015). An environment that supports truly reproducible statistical research and is free in all aspects of the word.

# Data

We use the data set that is created from the IDS data base with a standard extraction script (Quaranta 2016). In the analyses presented here we are using **R**, a free software environment for statistical computing and graphics (R Development Core Team 2016). The study is limited to mothers born between Sunday, January 1, 1826 and Tuesday, December 31, 1850.

## Variables

Some variables need to be redefined, centered and categorized. Here is a description.

* **mDeaths** The number of infant deaths a mother experiences. This is the basic *response* variable in all models. In some cases it is categorized.
* **gmDeaths** The number of infant deaths a grandmother (mother's mother) experiences. It is the primary . It is categorized in the categories "0", "1", and "2+".
* **gmBirths** The number of grandmother's births. Is always (by design) two or larger. In the analyses it is categorized: "2", "3", "4-6", "7+".
* **mBirthdate** Mother's birthdate expressed as years between 1 January 1 and her day of birth, minus one, *minus 1840*. An example: A woman born 3 April 1842 gets the value 1842.253 to begin with. Then subtract 1840, and the value is 2.253. The reason for this around 1840 is *good statistical practice*: The reference point (zero) of a covariate should lie within the range of the variable in regression analyses.
* **mAge** Mother's age at the birth of a child. Categorized into the intervals "17-19", "20-22", ..., "41-43", "44-49".
* **parity** The birth order of a mother's child.
* **childBirthdate** The birthdate of a mother's child, centered in the same way as *mBirthdate*.
* **gmIMR** Grandmother's infant mortality rate, that is, the number of deaths divided by the number of births. Used as an alternative to *gmDeaths* in some analyses. It is also multiplied by 100 and thus measured on a scale ranging from 0 to 100.
* **sex** The sex of a mother's child.

# Descriptive statistics of the IDS extraction

The yearly numbers of births and deaths for the *mothers in the data set* and, as a comparison, for the data in *Poplink*, are shown in Figure 1.

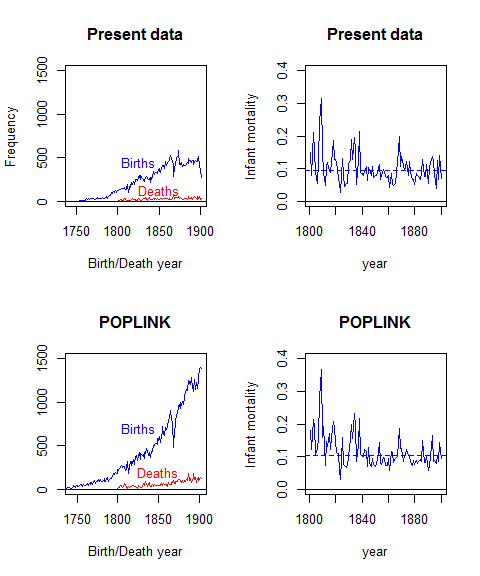


Figure 1 Number of births and infant deaths and infant mortality by year.

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 data frame is then aggregated to one record per mother for the Poisson regression approach, where the influence of grandmother's %infant deaths on the number of her daughter's' infant deaths is studied.

The average number of births to a mother in the data set is . The corresponding number for a grandmother is , but keep in mind that grandmothers are restricted to have at least two births, while mothers are restricted to have at least one birth.

## The covered time period

Our study sample consists of all mothers born 1826--1950. The distribution of their birth years and their infant mortality (by year of child death) are shown in Figure 2. The distribution of the grandmothers' birthdates is unknown, because it is not part of the IDS retrieval. The first births ought to have appeared in the early 1800s.

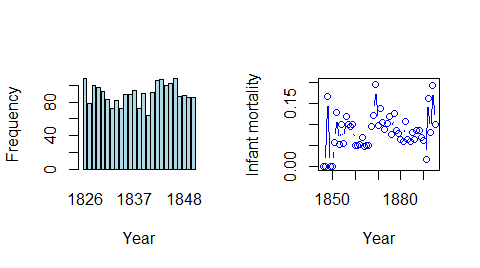


Figure 2 Distribution of mother birth years and mothers' infant mortality rate by infant death year.

## Grandmothers, mothers and mother--sisters

There are mothers and 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 inter-generational 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.

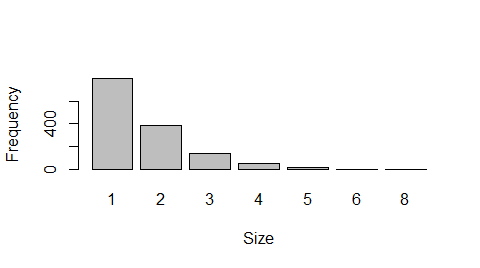


Figure 3 Distribution of sister group sizes.

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

# Results

The first section of the results, with the standard models, reports the analyses that are the same for all the different participants. The data are first analyzed with Poisson regression, and thereafter with survival analysis. In the second section, with the extended models, alternative approaches and models are investigated.

## Standard models

### Poisson regression

The expected value of the number of infant deaths for mother No. , is modeled by a Poisson distribution as

where is total risk time for mother No. , a vector of her explanatory variables, and is the vector of regression coefficients. (For a mother with no infant deaths, the risk time is equal to her number of births.) Formally, is entered into the model as an after taking logs.

The results are presented in two steps: First, *the statistical significance* is calculated and shown in Figure @(fig:poisp). The solid horizontal red line at 5% is our (conventionally) chosen nominal limit for statistical significance. The dashed line is the limit that should be respected in honor of the *multiple comparisons* situation (Holm 1979). The first (leftmost) covariate, *gmDeaths*, is clearly statistically significant, while *gmBirths* is just barely significant. *mBirthdate* is clearly out. Second, the *effect sizes* are graphically evaluated in Figure 5.

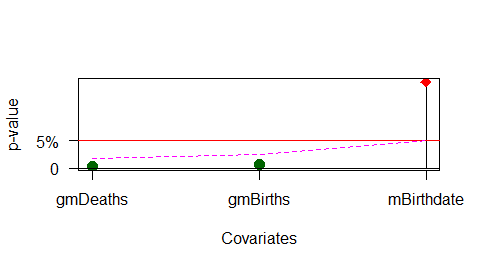


Figure 4 LRT p-values for the covariates, Poisson regression. The dashed line is the significance limit with "multiple comparisons" correction.

So, the likelihood ratio test (LRT) shows that *gmDeaths* and *gmBirths* are highly *statistically* significant in the model. The important result is the significance, the effect sizes. Figure 5 shows the of infant death by the number of grandmother's deaths and the number of her births.

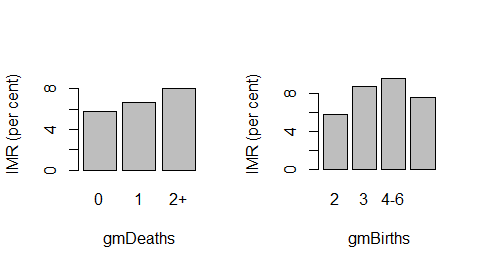


Figure 5 Infant mortality by grandmother's number of infant deaths (left) and grandmother's number of births(right). Comparisons made at reference value of other covariates.

### Survival analysis

The **R** package *eha* (G. Broström 2015; G. Broström 2012) is used, and the explanatory variables are almost the same as in the Poisson regression analysis. The difference is that instead of *mBirthdate*, the *childBirthdate* is used (no big difference), and a new variable, *mAge*, is introduced. It is categorized into three-year intervals. Other covariates, *parity*, etc. are omitted, since they do not contribute much to the model.

The main features of the fit are shown graphically: First, the statistical significance of involved covariates is found in Figure 6. The *p*-values are marked as small filled (green) circles for *p*-values falling below the dashed line showing significance at the 5% level even in the presence of a multiple comparisons correction. The *p*-values are ordered increasingly from left to right in the figure. All included covariates have a significant influence on the model fit.

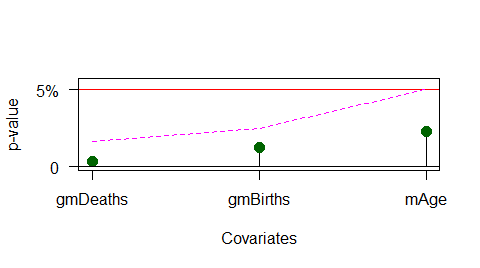


Figure 6 LRT p-values for the covariates, Cox regression. Dashed line is significance limit with "multiple comparisons" correction.

Second, the effects are shown in Figure 7. For 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. *gmBirths* is also influential, as is *mAge* with its characteristic *U*-shaped effect on infant mortality.

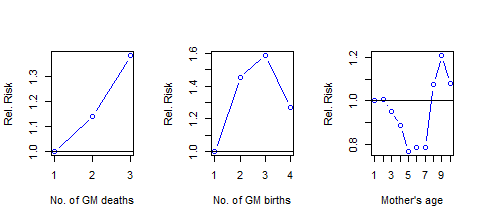


Figure 7 Effects of included covariates, Cox regression. The leftmost value is the reference in all panels.

The estimated cumulative hazards functions for the strata of *gmDeaths* are shown in Figure 8.

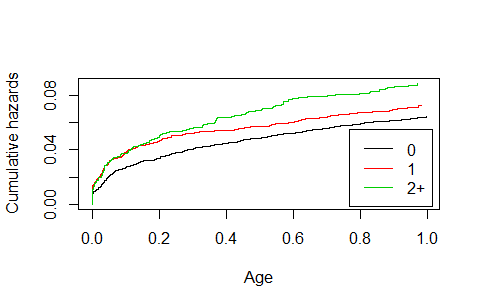


Figure 8 Cumulative hazards by the number of grandmother's infant deaths.

There is an evident deviation from the assumption of *proportional hazards*: It is cases with exactly one grandmother infant death that deviates. However, this does not disturb the main conclusion: Two or more grandmother infant deaths is harmful, quite in line with the results from the Poisson regression.

## Extended models

Findings that are specific for the Skellefteå project are presented here. The emphasis is that it generally is a good idea to look at a problem from different angles.

### The Poisson model

It turns out that the model fit to data is somewhat better with *number of births* as the offset (log scale). Also showing a slight improvement is to use *gmIMR* as explanatory variable rather than her *absolute number of deaths*. The *IMR* is defined as the number of infant deaths divided by the number of births. These changes also implies that the comparison *mother vs. grandmother* happens on a probability scale rather than on the intensity one.

Table 1 Poisson regression, grandmother's IMR. Analysis of deviance.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Deviance | AIC | LRT | Pr(>Chi) |
|  | NA | 2584.055 | 4060.971 | NA | NA |
| gmIMR | 1 | 2592.209 | 4067.125 | 8.15383 | 0.00430 |
| mBirthdate | 1 | 2585.703 | 4060.619 | 1.64816 | 0.19921 |
| parity | 4 | 2588.951 | 4057.867 | 4.89645 | 0.29809 |

The general conclusion is not changed: Grandmother's IMR has a strong influence on her daughter's IMR.

The size of the effect is shown in Figure 9.

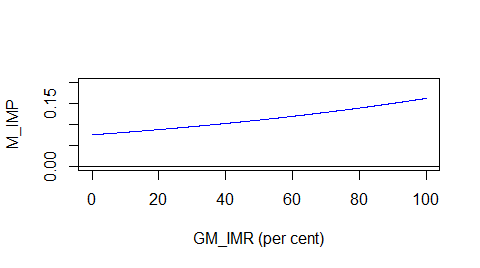


Figure 9 Probability of infant death for mother (M IMP) by IMR for grandmother (GM IMR). Poisson regression.

### Survival analysis

The same modification as in the Poisson case is introduced here.

Table 2 Cox regression, extended model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Df | AIC | LRT | Pr(>Chi) |
|  | NA | 16551.94 | NA | NA |
| gmIMR | 1 | 16557.83 | 7.890257 | 0.0049702 |
| childBirthdate | 1 | 16551.81 | 1.867376 | 0.1717760 |
| parity | 4 | 16549.65 | 5.710213 | 0.2218602 |
| sex | 1 | 16557.69 | 7.747161 | 0.0053797 |

The regression coefficient for *gmIMR* is 0.007, highly statistically significant, as seen in Table 2. The other covariates are of less importance.

### 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, Borgan, and Gjessing (2008)} and the implementation of *robust variances* (Therneau and Grambsch 2000). We have tried both, but neither do change the results in any noticeable way.

A radical way to eliminate the sibling effect among infants is to include the firstborn for each mother. We get the results shown in Table 3.

Table 3 Cox regression, extended model with only one birth per mother.

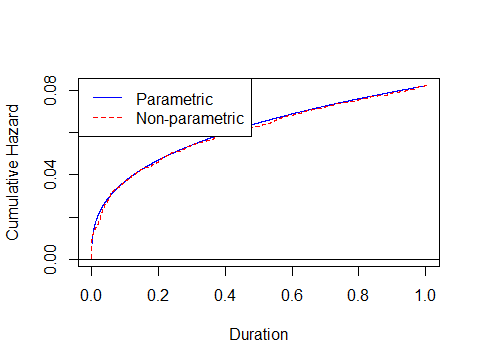
|  |  |  |  |
| --- | --- | --- | --- |
|  | coef | exp(coef) | se(coef) |
| gmIMR | 0.0089737 | 1.0090140 | 0.0056830 |
| childBirthdate | 0.0200612 | 1.0202638 | 0.0085286 |
| sexMale | -0.0348960 | 0.9657058 | 0.1443492 |

The effect of *gmIMR* is even stronger for firstborn than generally (estimated to 0.009), however, the statistical significance is weak, a logical consequence of the much lower number of infant deaths (compare Tables 2 and 3).

### Parametric proportional hazards

The Weibull model usually fits infant mortality data well.

Call:  
phreg(formula = Surv(exit - enter, event) ~ gmIMR + childBirthdate +   
 parity + sex, data = ef)  
  
Covariate W.mean Coef Exp(Coef) se(Coef) Wald p  
gmIMR 7.955 0.007 1.008 0.003 0.004   
childBirthdate 30.594 0.005 1.005 0.004 0.171   
parity   
 1 0.207 0 1 (reference)  
 2 0.174 -0.060 0.942 0.108 0.580   
 3 0.148 -0.123 0.884 0.116 0.289   
 4-6 0.320 -0.153 0.858 0.099 0.123   
 7+ 0.151 0.059 1.061 0.119 0.619   
sex   
 Female 0.486 0 1 (reference)  
 Male 0.514 0.187 1.205 0.067 0.006   
  
log(scale) 7.841 0.465 0.000   
log(shape) -1.057 0.033 0.000   
  
Events 895   
Total time at risk 10180   
Max. log. likelihood -2303.8   
LR test statistic 24.38   
Degrees of freedom 7   
Overall p-value 0.000976948

 A graphical comparison of the Weibull and nonparametric baseline cumulative hazards (Figure 10) shows an exceptionally good fit.

The advantage of the parametric (Weibull) model vs. the Cox regression model is that we can effortlessly estimate the baseline hazard function (no kernel estimation with ad hoc bandwidth selection). It is further possible to *formally test* the hypothesis of proportional hazards for a categorical covariate through stratification and the LRT test. If proportionality is rejected it is still possible to test for non-proportional effects. (Logically, if proportionality is rejected, then there cannot be equality, but statistical hypothesis testing is not always logical in a common sense: Best to make sure.)

### Test of proportionality

The model with *stratifying* on *gmDeaths* (with the categories "0", "1", "2+") gives the result shown in Figure 11.

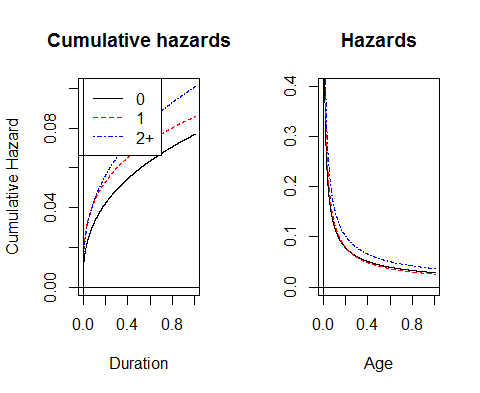


Figure 11 Baseline hazards from the stratified Weibull regression.

A formal test rejects the hypothesis of proportional hazards (no surprise), and a further test of equality of the curves also rejects (ditto). This is in exact agreement with our conclusion regarding the Cox regression model (Figure @(fig:relcumhaz)). For the stratified Weibull model, the shape and scale parameters both vary freely over the strata, while in the non-stratified model (but with *gmDeaths* as a covariate instead of defining strata), the shape parameter is the same in all strata. So the stratified model requires six baseline parameters to estimate, while the non-stratified only requires four.

The LR test now takes two times the difference of the two maximized log likelihoods, which under the null hypothesis (no stratification necessary) is -distributed with degrees of freedom. The maximized log-likelihood values are and , so the test statistic is observed to be 16.77, which with two degrees of freedom gives a -value of 210^{-4}. So the null hypothesis is rejected.

### A 2-by-2 table

In order to make it really simple, let us just record whether a mother and a grandmother experienced an infant death or not. The result is, in tabular form,

|  |  |  |
| --- | --- | --- |
|  | FALSE | TRUE |
| FALSE | 971 | 605 |
| TRUE | 373 | 298 |

The *odds ratio* in this table is 1.28, and *Fisher's exact test* (Fisher 1922) gives a -value of 0.0085 and a 95% confidence interval (1.06, 1.55). Expressed in probabilities: If no grandmother death, then the probability of a mother death is *605 / 1576 = 0.38*, while if grandmother experienced a death the corresponding probability is *298 / 671 = 0.44*, an increase by *16%*.

An even simpler table is obtained if we do not allow siblings among mothers, that is, to each grandmother only one mother (her first-born daughter) is connected:

|  |  |  |
| --- | --- | --- |
|  | FALSE | TRUE |
| FALSE | 587 | 357 |
| TRUE | 241 | 199 |

Now, the *odds ratio* in this table is 1.36, and *Fisher's exact test* (Fisher 1922) gives a *p*-value of 0.0096 and a 95% confidence interval (1.07, 1.72). Expressed in probabilities: If no grandmother death, then the probability of a mother death is 0.38, while if grandmother experienced a death the corresponding probability is 0.45, an increase by 16%.

An astonishing similarity!

### The 2-by-2 table with covariates

We can of course analyze these table data by *binomial regression*, including the same covariates as earlier. Since the results from this exercise do not differ from earlier results, we refrain from showing them.

# A sensitivity analysis

The death books for the years 1822--1830 are lost in a fire, so the only information about deaths for these years is what is found in the husförhörslängder. This means that very early deaths may not be reported. See Figure 1.

# Concluding remarks

Three main aims were set out for this article. The first was about assessing the power 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 (2016), has fulfilled its purposess 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 oldest daughter an infant death or not: A contingency table analysis). Complicated'' refers tostrong 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. You may call it *sensitivity analysis* (Cox and Oakes 1984, chap. 6) if you wish.

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 38 per cent, otherwise it is 45 per cent.

# Discussion

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 possible pathways.

One obvious possible factor refers to genetics. People are differently frail and it is reasonable that this frailty is genetically transferred. This might affect childrens' capacity of survival as well as factors related to the mother associated with delivery and her the ability to take care of the newborn. There are also other possible biological reasons. The Rh negative blood group is common in our 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. Still it can still be a component in the observed association (Häggström Lundevaller and Edvinsson 2012). Furthermore, a mother and her sibling group can 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 can 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 organization 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 (S. Edvinsson 2004). 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 analyzed in models where residence and/or social position are considered.

Regarding the results, the main hypothesis was confirmed: The risk for a woman to experience infant deaths is increased if her mother also had that experience. While statistically significant, the effect size is moderate. We applied different models, all plausible, and got almost identical results in all cases.

# References

Aalen, O.O., Ø. Borgan, and H.K. Gjessing. 2008. *Survival and Event History Analysis, a Process Point of View*. New York: Springer.

Alm Stenflo, G. 1994. *Demographic Description of the Skellefteå and Sundsvall Regions During the 19th Century*. Umeå, Sweden: Demographic Data Base, Umeå University.

Brändström, A., S. Edvinsson, M. Lindkvist, and J. Rogers. 2008. “Clustering Across Generations: A Comparative Analysis of Infant Mortality in 19th Century Sweden.”

Broström, Göran. 2012. *Event History Analysis with R*. Boca Raton: Chapman & Hall/CRC.

———. 2015. *eha: Event History Analysis*. <https://CRAN.R-project.org/package=eha>.

Coale, Ansley J., and Susan Cotts Watkins. 1986. *The Decline of Fertility in Europe*. Princeton University Press.

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

Edvinsson, S. 2004. “Att Ha Omsorg Om Sina Barn. Traditionens Och Miljöns Roll För Spädbarnsdödlighet.” In *Befolkningshistoriska Perspektiv. Festskrift till Lars-Göran Tedebrand*, edited by A. Brändström, S. Edvinsson, T. Ericsson, and P. Sköld. Demographic Data Base.

Edvinsson, S., and A. Janssens. 2012. “Clustering of Deaths in Families—infant and Child Mortality in Historical Perspective.” *Biodemography and Social Biology* 58 (2): 75–86. doi:[10.1080/19485565.2012.738575](https://doi.org/10.1080/19485565.2012.738575).

Edvinsson, Sören, Anders Brändström, John Rogers, and Göran Broström. 2005. “High-Risk Families: The Unequal Distribution of Infant Mortality in Nineteenth-Century Sweden.” *Population Studies* 59: 321–37.

Engberg, E. 2005. “I Fattiga Omständigheter. Fattigvårdens Former Och Understödstagare I Skellefteå Socken Under 1800-Talet.” Umeå: Demographic Data Base.

Fisher, R. A. 1922. “On the Interpretation of Chi-Squared from Contingency Tables, and the Calculation of P.” *Journal of the Statistical Society* 85: 87–94.

Häggström Lundevaller, E., and S. Edvinsson. 2012. “The Effect of the Rh Negative Disease on Perinatal Mortality. Evidence from Skellefteå 1840–1900.” *Biodemography and Social Biology* 58:2: 116–32. doi:[0.1080/19485565.2012.720450](https://doi.org/0.1080/19485565.2012.720450).

Holm, S. 1979. “A Simple Sequentially Rejective Test Procedure.” *Scandinavian Journal of Statistics* 6: 65–70.

Holmberg, H., and G. Broström. 2012. “On Statistical Methods for Clustering: A Case Study on Infant Mortality, Northern Sweden, 1831–1890.” *Biodemography and Social Biology* 58: 173–84.

Meloni, M. 2014. “How Biology Became Social, and What It Means for Social Theory.” *The Sociological Review* 62: 593–614. doi:[10.1111/1467-954X.12151](https://doi.org/10.1111/1467-954X.12151).

Nilsdotter Jeub, U. 1993. “Parish Records. 19th Century Ecclesiastical Registers.” Umeå: Demographic Data Base.

Pembrey, M., R. Saffery, and L.O. Bygren. 2014. “Network in Epigenetic Epidemiology; Network in Epigenetic Epidemiology 2014. ‘Human Transgenerational Responses to Early-Life Experience: Potential Impact on Development, Health and Biomedical Research’.” *Journal of Medical Genetics* 51 (9). doi:[10.1136/jmedgenet-2014-102577](https://doi.org/10.1136/jmedgenet-2014-102577).

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.

———. 2016. *Program for Studying Intergenerational Transmissions in Infant Mortality Using Ids*. Centre for Economic Demography, Lund University.

R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

RStudio Team. 2016. *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>.

Therneau, Terry M., and Patricia M. Grambsch. 2000. *Modeling Survival Data: Extending the Cox Model*. New York: Springer.

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., S. Edvinsson, and E. Engberg. 2016. “A Unique Source for Innovative Longitudinal Research: The POPLINK Database.” *Historical Life Courses* 3: 20–31.

Xie, Y. 2015. *Dynamic Documents with R and Knitr*. Boca Raton: Chapman & Hall/CRC Press.

———. 2016. *Knitr: A General-Purpose Package for Dynamic Report Generation*. <https://CRAN.R-project.org/package=knitr>.