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

Elisabeth Engberg, Sören Edvinsson, and Göran Broström

September 1, 2016

1 Introduction

In this article, we have a threefold aim, two of methodological character and one demographic. The first methodological one concerns the application of the IDS structure for demographic analysis. Experiences from this is useful for further development of the structure and for future uses of it. The second methodological question relates to the statistical analysis of intergenerational transfers, evaluating different models. These methodological issues are used for the study of possible intergenerational 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, 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 intergenerational transmission of demographic patterns (Edvinsson and Janssens, 2012), sometimes even considering epigenetic effects (Pembrey et al., 2014).

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.

Finally, one should also be aware of the fact that clustering emerges naturally from pure randomness, often very counterintuitive. This is discussed by Holmberg and Broström (2012).

2 Area

The area under study is the 19th century Skellefteå region in the province of Västerbotten in the northern part of Sweden. The region was vast, and consisted at the outset of the study of one large rural parish, Skellefteå parish. In 1875, the northern part, Byske was detached into a separate administrative unit, but the population is nevertheless 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's 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 was around 6900, and it increased rapidly during the first half of the century. By 1850 it had reached to about 14000 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 farm-

ers 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 an international 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 prenuptial conceptions and illegitimate births were common. The illegitimacy rate fluctuated between three and six per cent during the nineteenth century (Alm Stenflo, 1994).

3 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 a continuous record of all demographic events for all individuals residing in a parish, making it possible to follow individuals over time and to identify their relatives. Furthermore, a rich variety of additional information, for example on occupation, was recorded. A selected number of parishes are digitized by the Demographic Data Base at Umeå University (`http://www.cedar.umu.se/`). The entries in the catechetical registers and the birth, death, marriage and migration registers have been linked, making it possible to analyse life courses of individuals while present in the digitalized regions. During recent years, the database has been prolonged with parish registers until the 1950s for the Skellefteå and Umeå regions. This is available in the anonymized database POPLINK with a generational depth of up to 15 generations (Westberg et al., 2016).

The registration in POPLINK of the Skellefteå region starts for the year

1699 with birth registers and from 1720 with catechetical registers. Since the 18th century catechetical registers underreport children that died young and the death registers are missing before 1815, we restrict our analysis to the 19th century. We include the reproductive history of women born 1826–50.

The public IDS database is extracted from the POPLINK database, but only until 1900 due to requirements of personal integrity.

4 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 *Stata*, 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, 2015) with the aid of the **R** package `knitr` (Xie, 2016, 2015). An environment that supports truly reproducible statistical research and is free in all aspects of the word.

5 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.

5.1 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 *explanatory variable*. 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 *centering* 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.

6 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.

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 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, not part of the IDS retrieval.

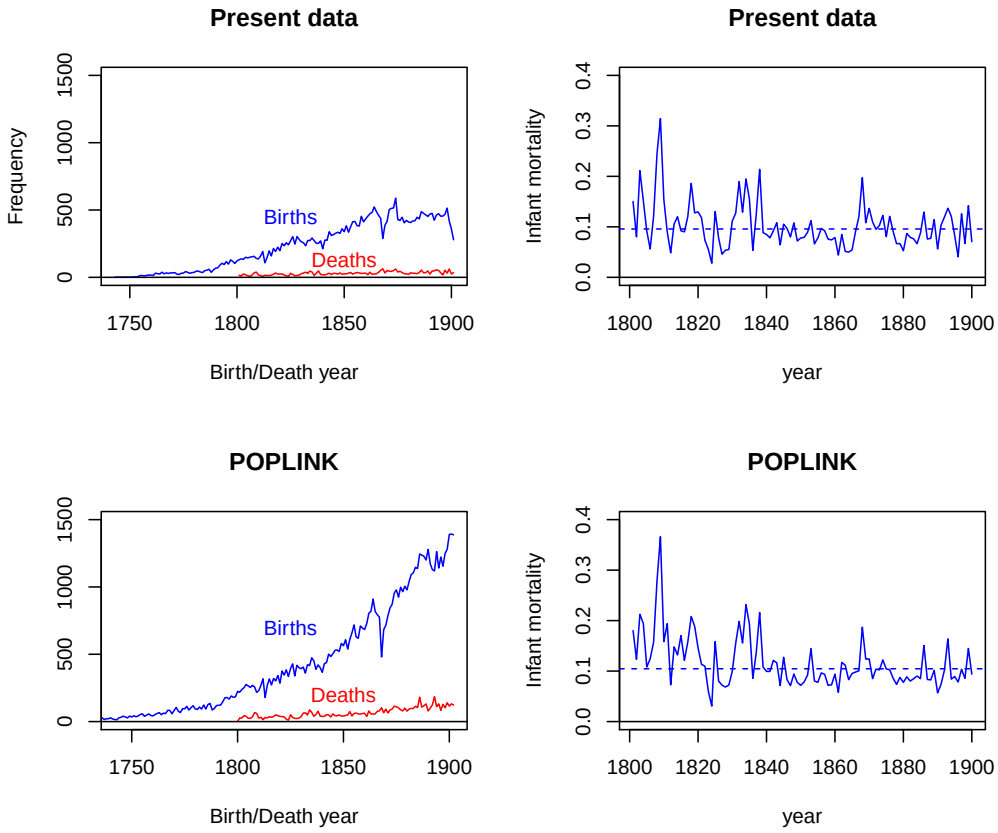

Figure 1: Number of births and infant deaths and infant mortality by year.

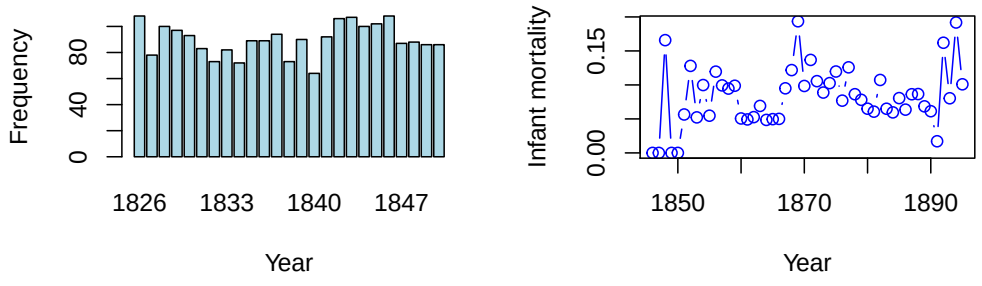

Figure 2: Distribution of mother birth years and mothers' infant mortality rate by infant death year.

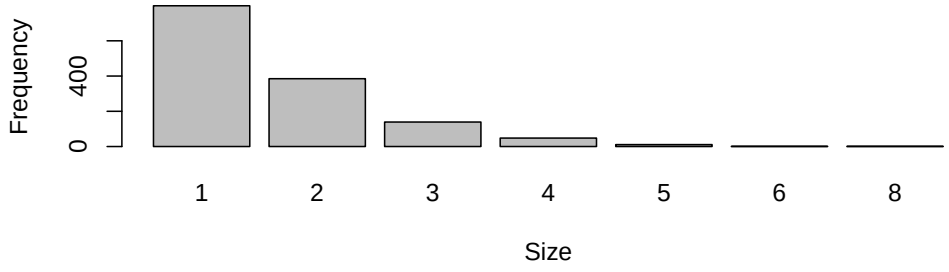

Figure 3: Distribution of sister group sizes.

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 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.

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.

7 Results

7.1 Standard results

The standard models are common to the subgroups of this research project.

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

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

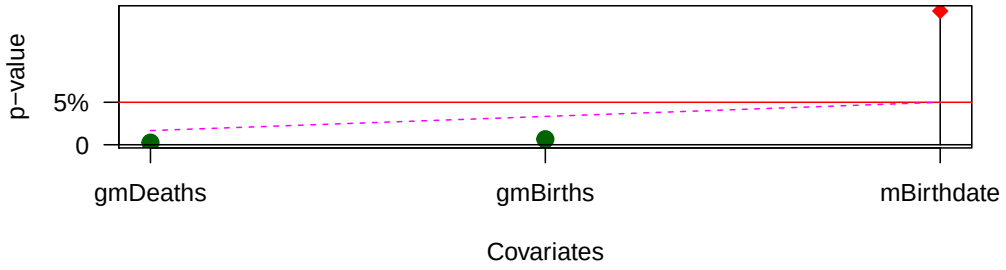

Figure 4: LRT p-values for the covariates, Poisson regression. The dashed line is the significance limit with "multiple comparisons" correction.

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 Figure 4. 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.

So, the likelihood ratio test (LRT) shows that `gmDeaths` and `gmBirths` are highly *statistically* significant in the model.

So, the important result is the *practical* significance, the effect sizes. Figure 5 shows the *expected risks* of infant death by the number of grandmother's deaths and the number of her births.

Survival analysis

The **R** package `eha` (Broström, 2015, 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.

[Table 1 about here.]

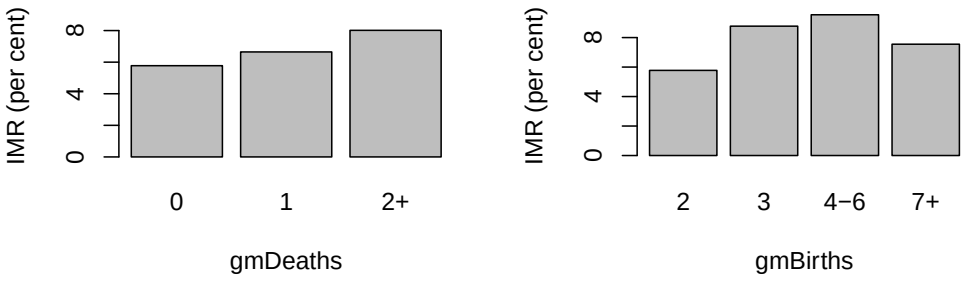

Figure 5: Infant mortality by grandmother's number of infant deaths (left) and grandmother's number of births(right). Comparisons made at reference value of other covariates.

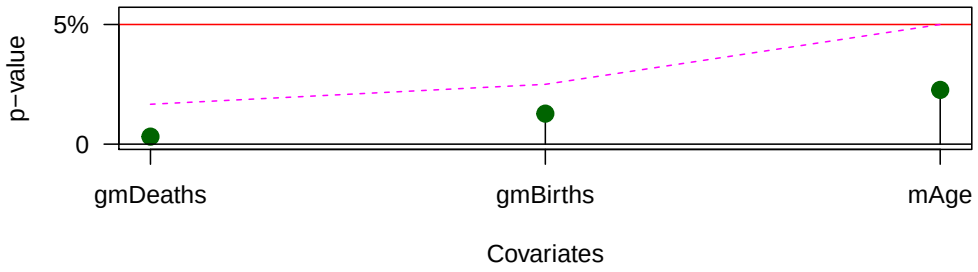

Figure 6: LRT p-values for the covariates, Cox regression. Dashed line is significance limit with "multiple comparisons" correction.

The standard tabular form of results is shown in Table 1, and 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.

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,

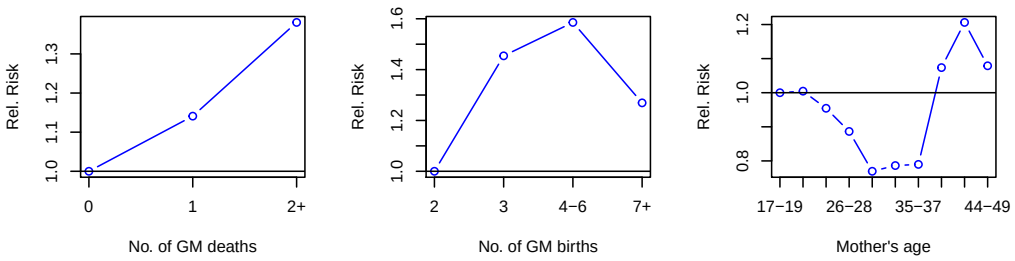

Figure 7: Effects of included covariates, Cox regression. The leftmost value is the reference in all panels.

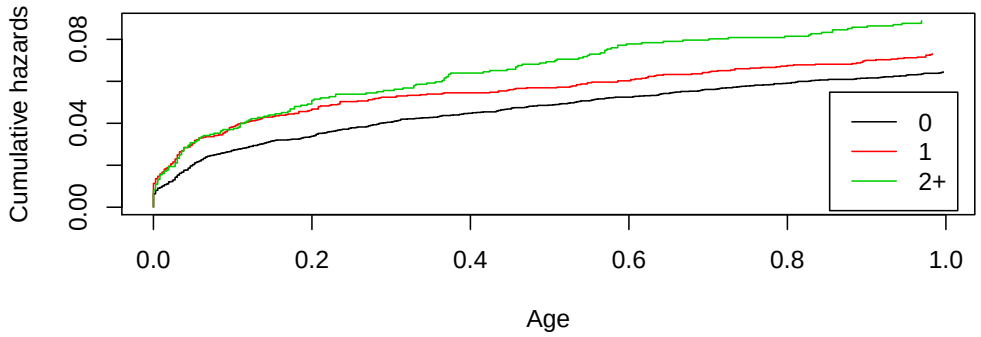

Figure 8: Cumulative hazards by the number of grandmother's infant deaths.

as is `mAge` with its characteristic U-shaped effect on infant mortality.

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

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.

7.2 Extended models

This section is specific for the Skellefteå project.

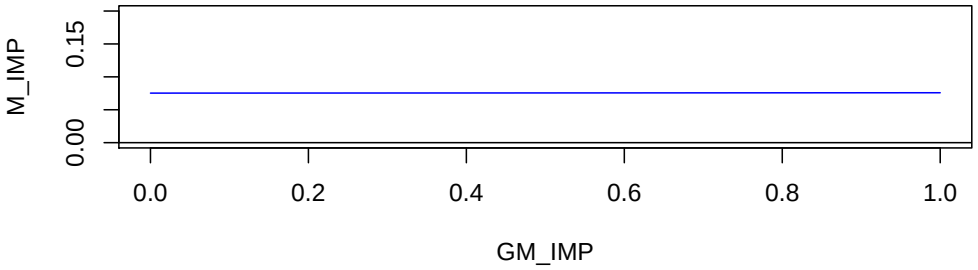

Figure 9: Probability of infant death for mother (M IMP) by probability of infant death for grandmother (GM IMP). Poisson regression.

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 2 about here.]

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.

Survival analysis

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

[Table 3 about here.]

The regression coefficient for `gmIMR` is 0.007, highly statistically significant, as seen in Table 3. The other covariates are of less importance.

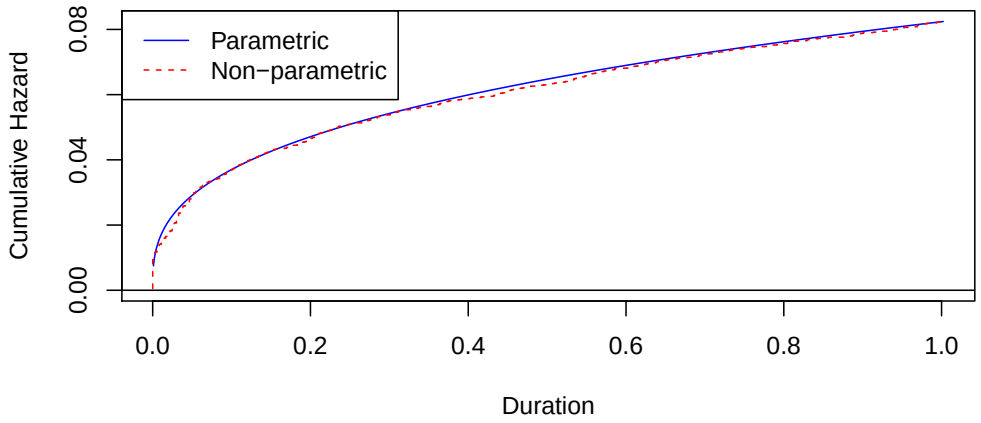

Figure 10: Weibull cumulative hazard vs. Nelson-Aalen estimate.

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 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 4.

[Table 4 about here.]

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 3 and 4).

Parametric proportional hazards The Weibull model usually fits infant mortality data well. Let us check this (Table 5).

[Table 5 about here.]

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

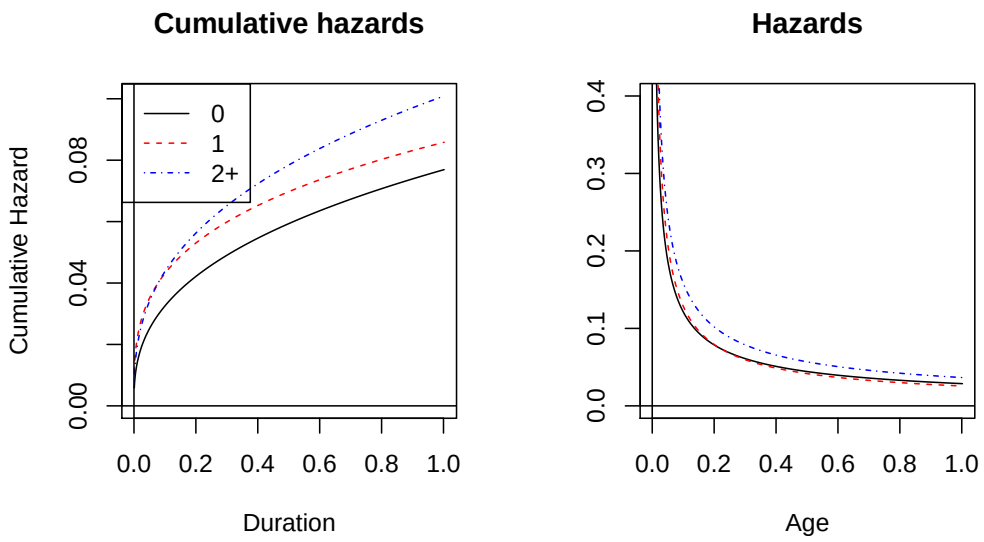

Figure 11: Baseline hazards from the stratified Weibull regression.

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 We modify the model by *stratifying* on `gmDeaths` (with the categories "0", "1", "2+") and plot the result, see Figure 11.

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 8). 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 χ^2-distributed with $6 - 4 = 2$ degrees of freedom. The maximized log-likelihood values are -2298.8 and -2307.2, so the test statistic is observed to be 16.767, which with two degrees of freedom gives a p-value of 2×10^{-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,

	Mother death		
Grandmother death	No	Yes	**Sum**
No	971	605	1576
Yes	373	298	671
Sum	1344	903	2247

The *odds ratio* in this table is 1.28, and *Fisher's exact test* (Fisher, 1922) gives a p-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:

	Mother death		
Grandmother death	No	Yes	**Sum**
No	587	357	944
Yes	241	199	440
Sum	828	556	1384

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.

8 Conclusion

We set out three main aims 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 has fulfilled its services well. The use of the script has facilitated comparisons between countries.

What is important to remember is that the IDS format do not relieve the researcher from 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, infant deaths were severely underestimated. There are ways taking care of this problem, but hat require clear definitions of rules on how to identify infant deaths, something that must be performed in a preparatory state.

Second, we have tried different analytical 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 to "strong assunptions 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 simpe-to-understand probabilities.

As is often the case, simple is at least as good as complicated: The different methods give similar results, and it is easy to feel confident in the ones emerging from models with the weakest assumptions. On the other hand, it is generally good (statistical) practise to try different approaches to the analysis of the same problem. You may call it *sensitivity analysis* (Cox and Oakes, 1984, chapter 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 probablility for mother to have one is 38 per cent, otherwise it is 45 per cent.

9 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 nurture that is important, a division that now is becoming muddled and partly obsolete due to recent scientific development (Meloni, 2014). In the analytical approach taken here, we cannot really distinguish between purely biological versus social and cultural factors, but we briefly discuss possible pathways.

One obvious possible factors refer to genetic factors and the inheritance of genes between generations. People are differently frail and it is reasonable that this is genetically transferred, both making children less capable of surviving and for conditions related to the mother when it comes to the delivery or 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 had 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 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 régime. This can relate to local patterns of child care, for example on the practice of breast-feeding, but it can also be practices that are transferred within families. Our closest family is the closest social organization for learning. In the 19th century Sundsvall region, mothers moving to the district kept their child care practices in the new environment, leading to higher infant mortality of when the mother came from a high mortality region vice versa (Edvinsson, 2004).

Another explanation for the association between generations is that they only 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 of a woman to experience infant deaths is strongly increased if her mother has that experience. We applied different models, all plausible, and got almost identical results in all cases.

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.

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

Broström, G. (2015). *eha: Event History Analysis*. R package version 2.4-3.

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.

Edvinsson, S. (2004). 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., 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. 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.

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 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.

Holm, S. (1979). A simple sequentially rejective test procedure. *Scandinavian Journal of Statistics*, 6:65–70.

Holmberg, H. and Broström, G. (2012). On statistical methods for clustering: A case study on infant mortality, northern Sweden, 1831–1890. *Biodemography and Social Biology*, 58:173–184.

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

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

Pembrey, M., Saffery, R., and Bygren, L. (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).

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. (2016). *Program for studying intergenerational transmissions in infant mortality using IDS*. Centre for Economic Demography, Lund University. Version 2, 2016-06-20.

R Development Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RStudio Team (2015). *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA.

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

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.

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

Xie, Y. (2016). *knitr: A general-purpose package for dynamic report generation*. R package version 1.13.

List of Tables

Covariate		Mean	Coef	Rel.Risk	S.E.	L-R p
gmDeaths						0.0031
	0	0.5913	0	1	(reference)	
	1	0.2689	0.1319	1.1410	0.0785	
	2+	0.1398	0.3224	1.3804	0.0951	
gmBirths						0.0128
	2	0.0283	0	1	(reference)	
	3	0.0513	0.3744	1.4542	0.2832	
	4-6	0.2810	0.4610	1.5856	0.2501	
	7+	0.6394	0.2384	1.2692	0.2480	
mAge						0.0227
	17-19	0.0096	0	1	(reference)	
	20-22	0.0545	0.0046	1.0046	0.3439	
	23-25	0.1182	-0.0467	0.9544	0.3303	
	26-28	0.1547	-0.1206	0.8864	0.3278	
	29-31	0.1680	-0.2619	0.7696	0.3286	
	32-34	0.1555	-0.2403	0.7864	0.3293	
	35-37	0.1409	-0.2360	0.7898	0.3305	
	38-40	0.1044	0.0711	1.0737	0.3303	
	41-43	0.0689	0.1874	1.2061	0.3350	
	44-49	0.0253	0.0760	1.0790	0.3706	
Events		895	TTR	10180		
Max. Log Likelihood		-8262				

Table 1: Cox regression, standard model.

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		2584.05485	4060.97070		
gmIMR	1.00000	2592.20868	4067.12453	8.15383	0.00430
mBirthdate	1.00000	2585.70301	4060.61886	1.64816	0.19921
parity	4.00000	2588.95130	4057.86716	4.89645	0.29809

Table 2: Poisson regression, grandmother's IMR. Analysis of deviance.

Covariate		Mean	Coef	Rel.Risk	S.E.	L-R p
gmIMR		7.9548	0.0075	1.0075	0.0026	0.0050
childBirthdate		30.5945	0.0054	1.0055	0.0040	0.1718
parity						0.2219
	1	0.2070	0	1	(reference)	
	2	0.1744	-0.0601	0.9416	0.1085	
	3	0.1478	-0.1235	0.8838	0.1162	
	4-6	0.3200	-0.1532	0.8579	0.0993	
	7+	0.1508	0.0593	1.0611	0.1190	
sex						0.0054
	Female	0.4865	0	1	(reference)	
	Male	0.5135	0.1870	1.2057	0.0674	
Events		895	TTR	10180		
Max. Log Likelihood		-8269				

Table 3: Cox regression, extended model.

Covariate		Mean	Coef	Rel.Risk	S.E.	L-R p
gmIMR		7.7433	0.0090	1.0090	0.0057	0.1249
childBirthdate		25.2885	0.0201	1.0203	0.0085	0.0187
sex						0.8090
	Female	0.4906	0	1	(reference)	
	Male	0.5094	-0.0349	0.9657	0.1443	
Events		192	TTR	2107		
Max. Log Likelihood		-1471				

Table 4: Cox regression, extended model with only one birth per mother.

Covariate		Mean	Coef	Risk Ratio	S.E.	L-R p
gmIMR		7.9548	0.0075	1.0075	0.0026	0.0050
childBirthdate		30.5945	0.0055	1.0055	0.0040	0.1708
parity						0.2229
	1	0.2070	0	1	(reference)	
	2	0.1744	-0.0601	0.9417	0.1085	
	3	0.1478	-0.1233	0.8840	0.1162	
	4-6	0.3200	-0.1531	0.8581	0.0993	
	7+	0.1508	0.0592	1.0610	0.1190	
sex						0.0054
	Female	0.4865	0	1	(reference)	
	Male	0.5135	0.1869	1.2055	0.0674	
Baseline parameters						
log(scale)			7.8407	2541.9366	0.4651	0.0000
log(shape)			-1.0574	0.3474	0.0330	0.0000
Events		895	TTR	10180		
Max. Log Likelihood		-2304				

Table 5: Weibull regression, extended model.