A biometric analysis of infant mortality and temperature, northern Sweden 1895-1950, including the Stensele station and the inland

Göran Broström and Tommy Bengtsson

2022-02-05 17:51:19

Abstract

The effect of extreme temperatures on infant mortality in the Umeå region 1895-1950 is studied in a biometric analysis setting. More precisely, the effect of climate and weather, measured by temperature, on infant mortality is investigated. It turns out that climate is more important than weather, low average temperatures are more important than temporary dips in temperature. The investigated geographical region is situated too far north for bad effects of high temperatures to show.

Contents

1 Introduction

The impact of ambient temperature variations on infant mortality is studied for a northern Sweden coastal area, the Umeå and Skellefteå regions, during the first half of the twentieth century. Two recent papers (Junkka et al., 2021; Karlsson et al., 2021) studied neonatal mortality and temperature variations in a larger geographical area containing the present one during the years 1880–1950. Climate and mortality in general is a research area that has generated great interest over the last years, see Bengtsson and Broström (2010).

The effect of seasonal variation and the occurrence of extreme monthly temperatures is studied and interacted with sex, social class, and legitimacy. Studies are performed separately for neonatal and postneonatal mortality, and for winter and summer seasons, and the classification into endogenous and exogenous factors will be discussed.

One important reason for studying neonatal and postneonatal mortality separately is the empirical findings by Bourgeois-Pichat (Bourgeois-Pichat, 1951a,b) about endogenous and exogenous mortality and the log-cube transform.

Figure 1 shows the study area within Sweden, with the weather stations marked. The map is taken (with permission) from the paper by Junkka et al. (2021).

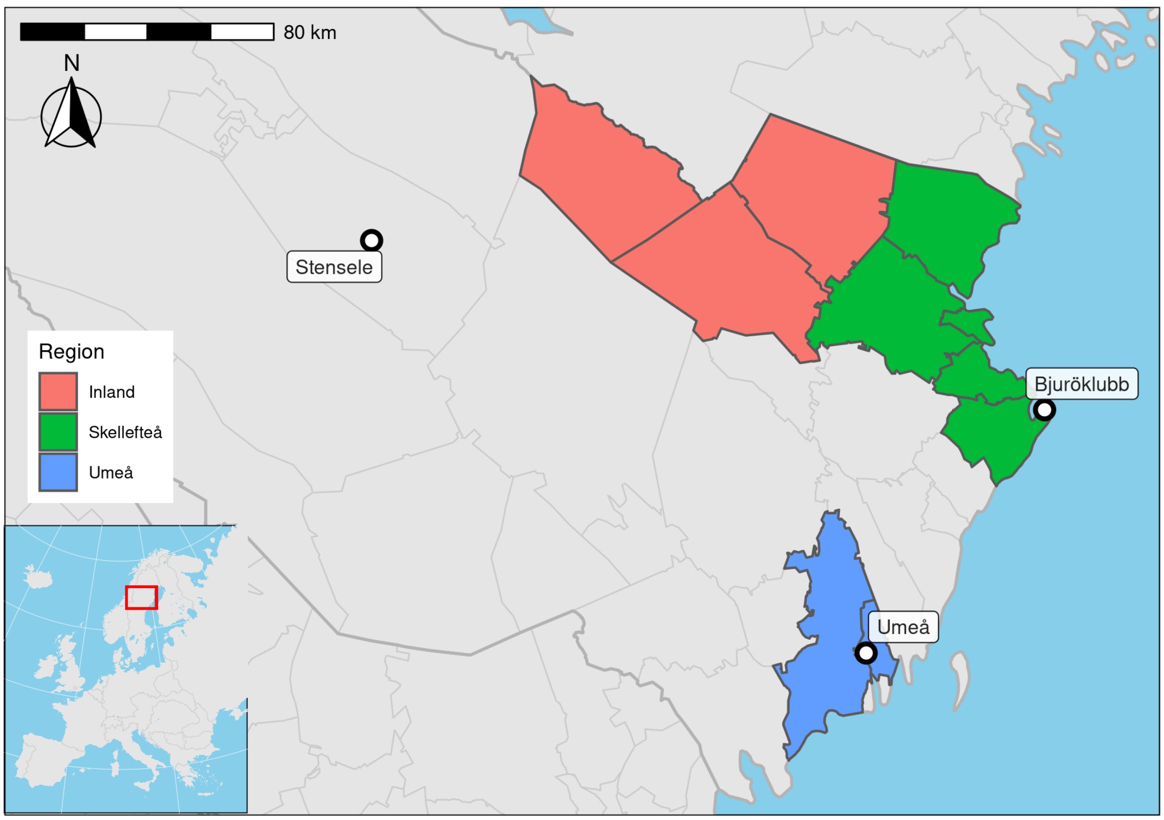

Figure 1: Umeå, Skellefteå (orange) and its inland (green).

2 Data

We have two sources of data which we combine into one data set suitable for our purpose. The first is demographic data obtained from the *Centre for Demographic and Ageing Research* (CEDAR, https://cedar.umu.se), the second is daily temperature measurements obtained from the *Swedish Meteorological and Hydrological Institute* (SMHI, https://www.smhi.se).

2.1 Infant mortality

Individual data with all births between 1 January 1895 and 31 December 1950 in two coastal areas of north Sweden, Skellefteå (51560 births) and Umeå (31213 births). They were followed until death or age 365 days, whichever came first. The following *static* characteristics were observed on each child:

birthdate Date of birth.

sex Girl or boy.

exit Number of days under observation.

event Logical, *TRUE* if a death is observed.

socBranch Working branch of father (if any).

socStatus Social status of family, based on HISCLASS.

illeg Mother unmarried?

parity Order among siblings.

Some crude statistics about infant, neonatal, and postneonatal mortality are shown in Figures.

Figure 2 shows the average monthly crude infant mortality, and a clear seasonal pattern is visible.

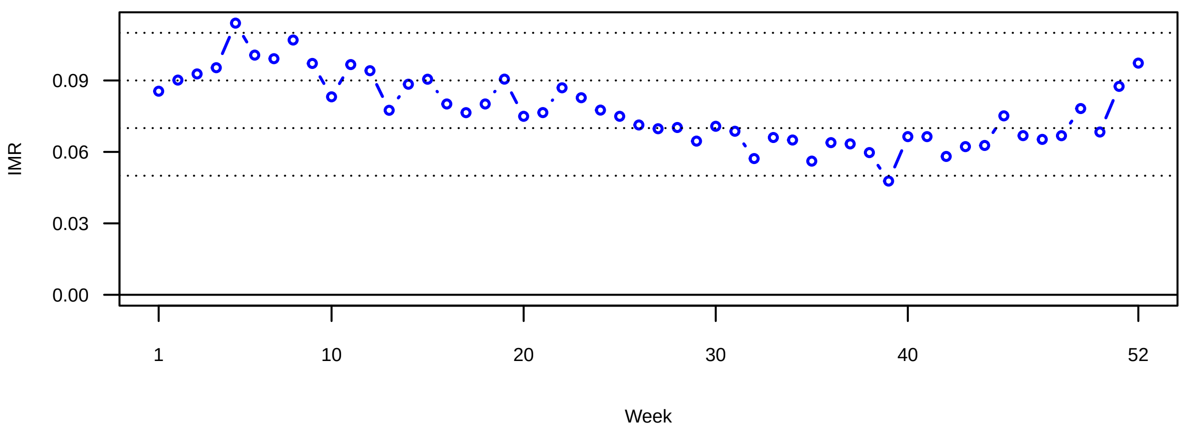

Figure 2: Crude infant mortality by week of year, Umeå/Skellefteå 1895–1950.

The average monthly neonatal mortality is shown in Figure 3.

The seasonal pattern is similar to the one we found above for infant mortality.

The average monthly postneonatal mortality is shown in Figure 4.

The seasonal pattern is once again similar to the one we found for infant mortality. Next, the decline over the years in Figures 5 and 6.

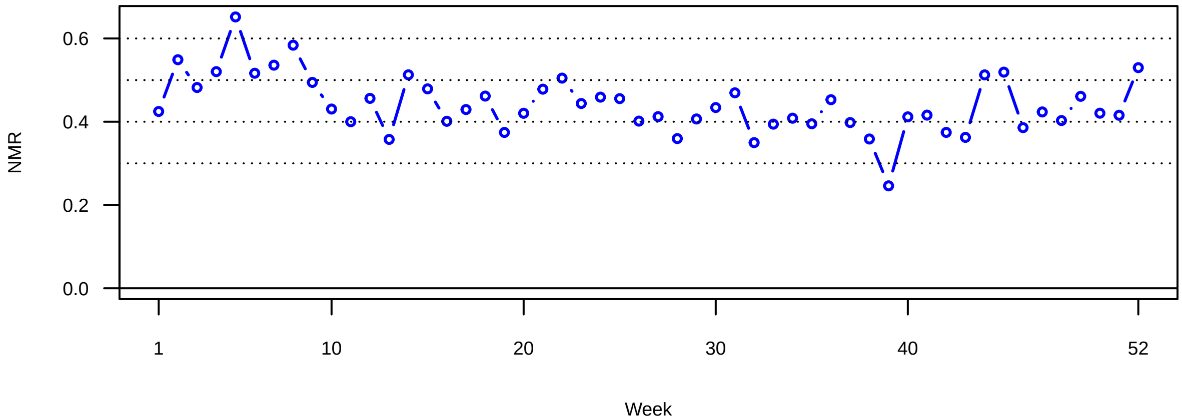

Figure 3: Crude neonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

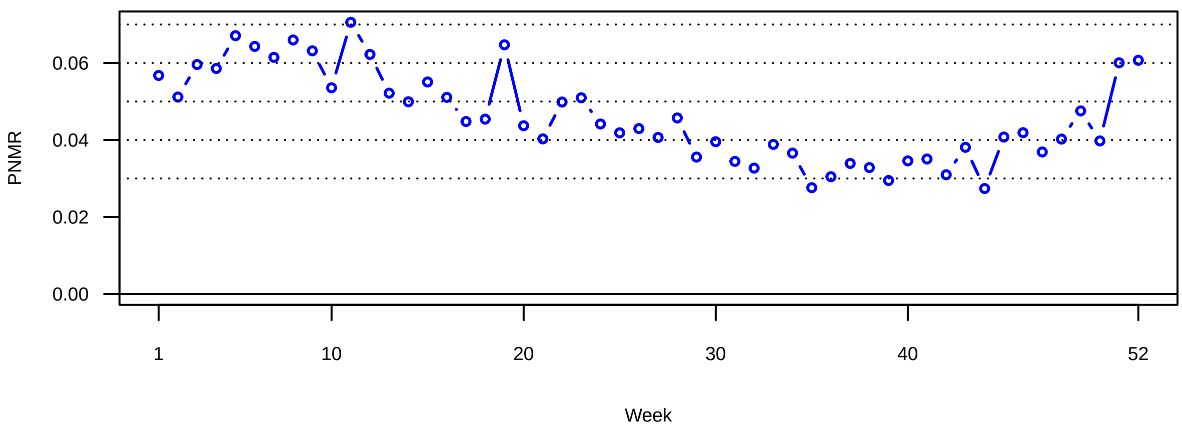

Figure 4: Crude postneonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

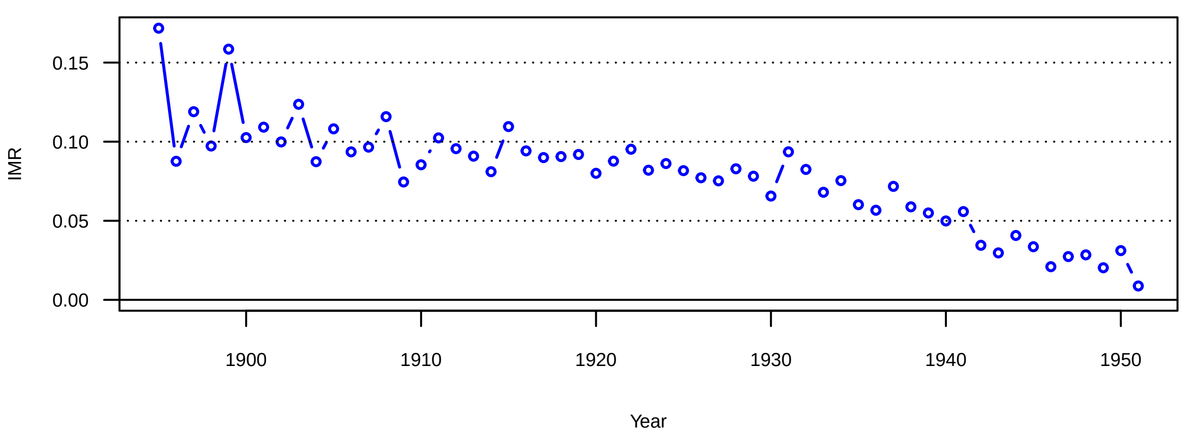

Figure 5: Crude IMR by year, Umeå-Skellefteå 1895–1950.

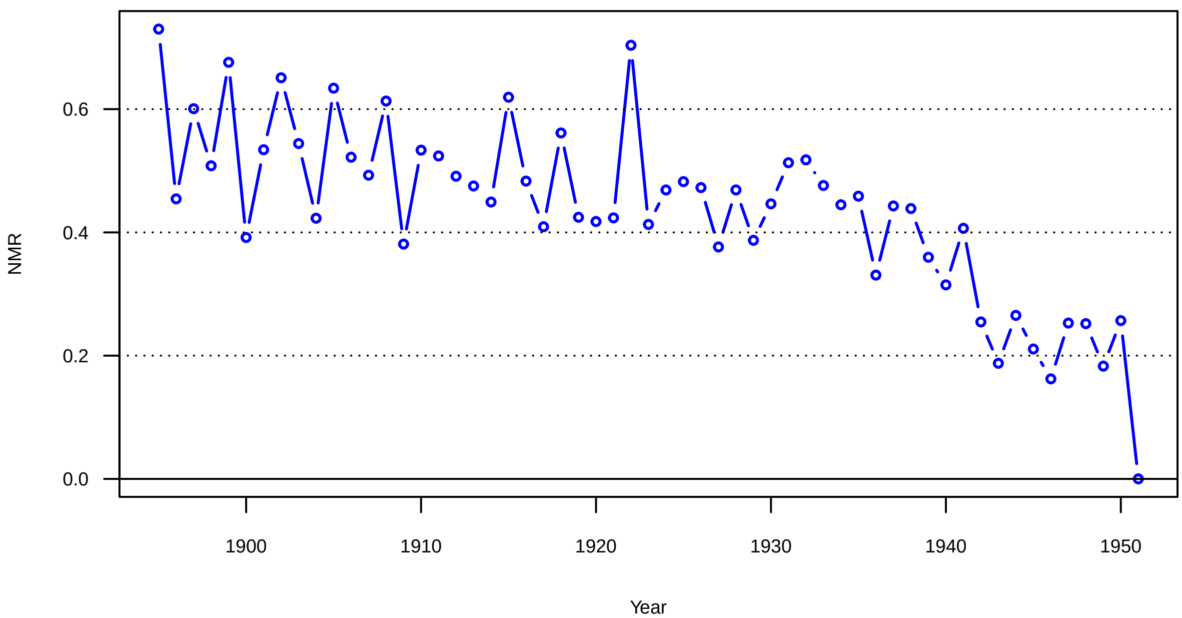

Figure 6: Crude NMR by year, Umeå-Skellefteå 1895–1950.

2.2 Temperature

Temperature data are collected from three weather stations, *Umeå*, *Bjuröklubb* (used with population data from Skellefteå coastal area), and *Stensele* (Inland). All stations deliver daily temperature data covering our time period, usually three measures per day, morning, noon, and evening. In Table 1, the Umeå data from the week 1–7 January, 1923 is shown.

There are three measurements per day, or 21 per week. In the forthcoming analyses, the weekly data are summarized in a few measurements, see Table 2.

Weekly averages (`mintemp`, `maxtemp`, `meantemp`) are calculated by week and year, and deviations from the averages (`emintemp`, `emaxtemp`, `emeantemp`) of the weekly averages are used as time-varying *communal covariates*. As an example, see Figure 7, where the variation around the average minimum temperature (`emintemp`) week 1 is shown.

Curiously, our randomly selected year 1923 turns out contain the warmest first week of all years, see Figure 7.

Figure 8 shows the average monthly distribution over all years. The subregional patterns and levels are very similar.

Time trends of yearly average temperatures, see Figure 9.

2.3 Temperature as communal covariates

The two data sets, mortality and weather, are combined into one by treating temperature data as a communal covariate and incorporate it as such in the mortality data set. The

Table 1: Raw temperature data from first week of 1923, Umeå.

Date	Time	Temperature	Quality
1923-01-01	07:00:00	0.4	G
1923-01-01	13:00:00	0.6	G
1923-01-01	20:00:00	0.0	G
1923-01-02	07:00:00	-1.4	G
1923-01-02	13:00:00	-1.4	G
1923-01-02	20:00:00	-1.2	G
1923-01-03	07:00:00	0.4	G
1923-01-03	13:00:00	0.8	G
1923-01-03	20:00:00	1.2	G
1923-01-04	07:00:00	1.4	G
1923-01-04	13:00:00	1.2	G
1923-01-04	20:00:00	1.0	G
1923-01-05	07:00:00	-1.4	G
1923-01-05	13:00:00	-3.2	G
1923-01-05	20:00:00	-3.4	G
1923-01-06	07:00:00	1.0	G
1923-01-06	13:00:00	0.4	G
1923-01-06	20:00:00	0.4	G
1923-01-07	07:00:00	0.6	G
1923-01-07	13:00:00	0.4	G
1923-01-07	20:00:00	0.4	G

Table 2: Weekly summarized temperature data: Umeå 1923, first week.

week	year	mintemp	maxtemp	meantemp	emintemp	emaxtemp	emeantemp
1	1923	-3.4	1.4	-0.1	-17.73	-0.36	-7.54

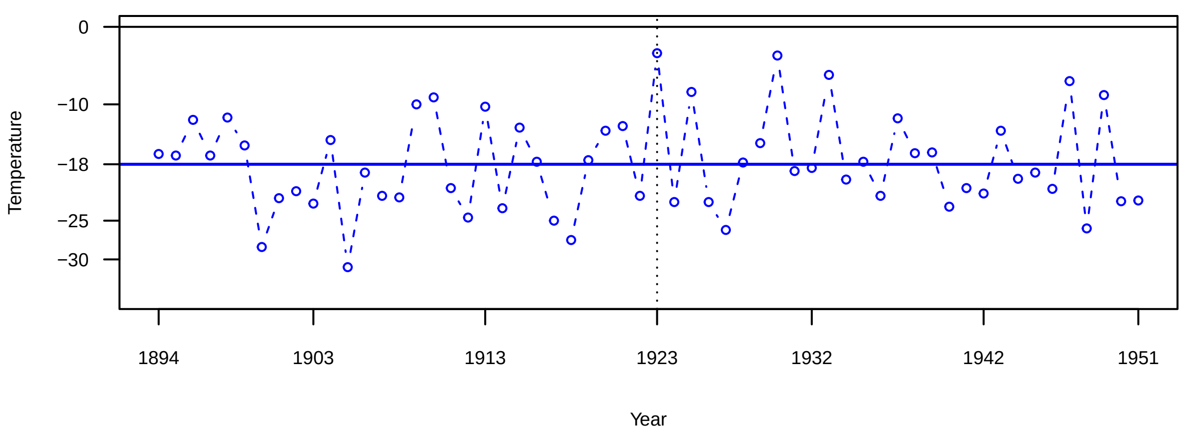

Figure 7: Minimum temperature the first week of each year.

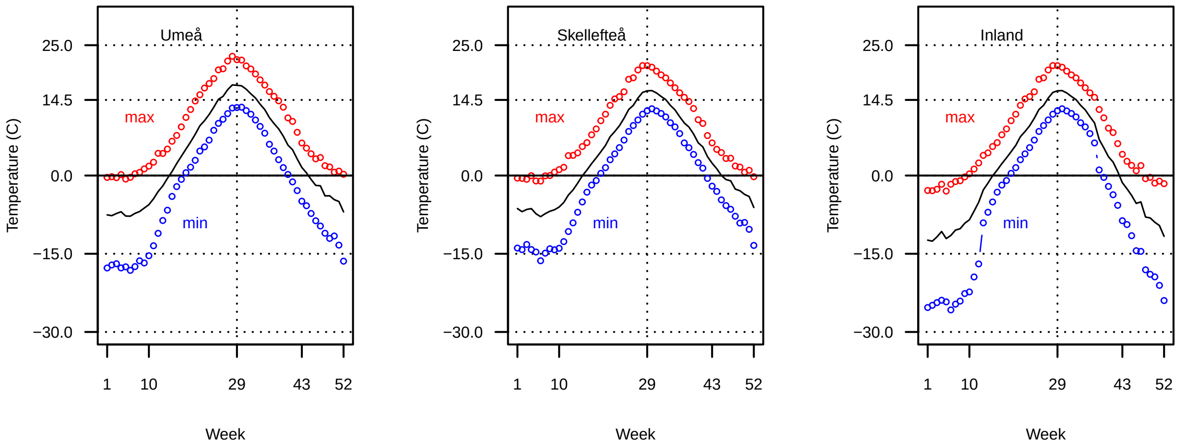

Figure 8: Weekly max, mean, and min temperature averages, 1895–1950.

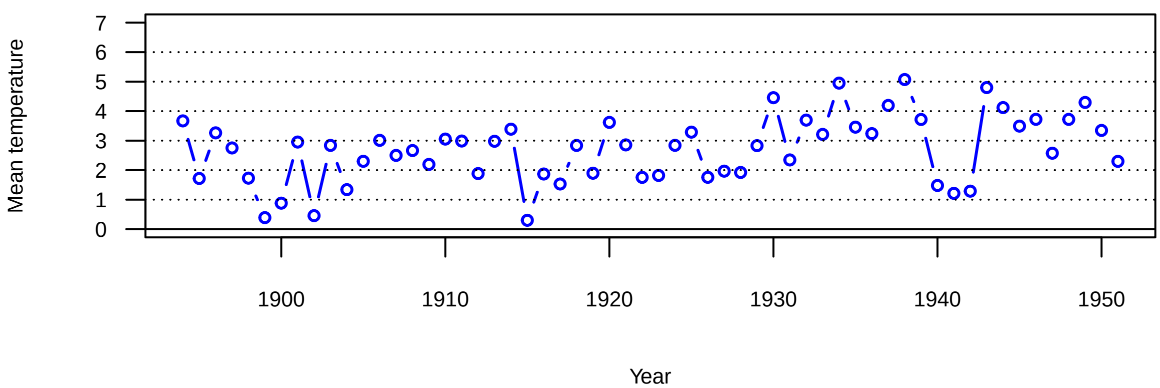

Figure 9: Yearly average temperatures, Umeå and Skellefteå.

Table 3: Data with communal covariates.

enter	exit	event	lowTemp	highTemp	aver	emeantemp	week	year
0.0000000	0.0180327	0	FALSE	FALSE	0	12.584483	35	1900
0.0180327	0.0372634	0	FALSE	FALSE	-4	11.286207	36	1900
0.0372634	0.0564942	0	FALSE	FALSE	-1	10.086207	37	1900
0.0564942	0.0757250	0	FALSE	FALSE	1	6.910345	38	1900
0.0757250	0.0949557	0	FALSE	FALSE	-2	5.315517	39	1900
0.0949557	0.1141865	0	FALSE	FALSE	0	3.610345	40	1900

function *make.communal* in the **R** (R Core Team, 2021) package *eha* (Broström, 2021a,b) is used for that purpose. Resulting data drame is partly shown in Table 3.

3 Statistical modelling

The analyses are performed on the *log-cube* scale, following the hints of Bourgeois-Pichat (Bourgeois-Pichat, 1951a,b), with proportional hazards modelling. Note that the property of proportional hazards are preserved under a strictly monotone increasing time transform. However, the estimates of baseline distribution characteristics will change, of course.

It turns out that extremely low temperature (`lowTemp`) is bad during all seasons except summer, and extremely high temperature (`highTemp`) is bad during summer, but good otherwise. So we group season into two categories, *summer* and *not summer*. In each case separate analyses for neonatal and postneonatal mortality are performed.

4 Results

The results regarding neonatal mortality is much in accordance with the results found by Junkka et al. (2021). However, they used temperature in a "hockey-stick" regression with a breakpoint at 14.5 degrees Celsius and a negative slope (decreasing risk) to the left and a positive slope (increasing risk) to the right. Instead, we are using the average weekly temperature for the 52 weeks of a year, for each week averaging over all the years in the study, as our "reference points" ("climate"), adding deviances up and down ("weather") as "short-term temperature stress". This is similar to the way prices and mortality were related in for instance Bengtsson and Broström (2011), that is, a time series split into long time trend and short term variation.

In accordance with the results of Bourgeois-Pichat (1951a) and Bourgeois-Pichat (1951b), we separate the investigation into two parts, *neonatal* and *postneonatal* mortality.

4.1 Neonatal mortality

The analyses are split into two parts by season, *winter + spring* is one, and *summer + fall* the other.

4.1.1 Winter and Spring

This period refers to the months *December to May.* A Cox regression involves as interesting variables *highTemp*, an indicator of temperature at least four degrees above the expected for at least two weeks in a row, *emeantemp* the expected temperature the actual weeek, and *aver* the *excess temperature* the actual week.

Table 4: Neonatal mortality, winter and spring. Adjusted for social branch, sex, illegitimacy, parity, time period, and subregion.

Covariate	**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
highTemp					0.031
FALSE	0.916	0	1	(reference)	
TRUE	0.084	−0.213	0.808	0.100	
aver	−0.012	−0.004	0.996	0.006	0.558
emeantemp	−2.856	−0.002	0.998	0.006	0.783
sex					0.000
boy	0.514	0	1	(reference)	
girl	0.486	−0.260	0.771	0.047	
season					0.047
winter	0.489	0	1	(reference)	
spring	0.511	−0.142	0.868	0.072	
subreg					0.665
ume	0.295	0	1	(reference)	
ske	0.481	0.024	1.025	0.055	
inland	0.224	−0.030	0.970	0.069	
illeg	0.055	0.352	1.422	0.091	0.000
parity					0.000
1	0.257	0	1	(reference)	
2-4	0.454	−0.231	0.794	0.059	
5+	0.288	0.100	1.105	0.063	
socBranch					0.064
office	0.104	0	1	(reference)	
farming	0.470	0.195	1.216	0.087	
worker	0.426	0.134	1.143	0.087	
Events	1896	TTR	4013		
Max. logLik.	−20565				

4.1.2 Summer and Fall

This period refers to the months *April to October.* A Cox regression involves as interesting variables *highTemp*, an indicator of temperature at least four degrees above the expected for at least two weeks in a row, *emeantemp* the expected temperature the actual weeek, and *aver* the *excess temperature* the actual week.

4.2 Postneonatal mortality

This is the simplest part, because Bourgeois-Pichat (1951a) predicts that the baseline distribution (given the log-cube transformation **g**) is *exponential*, that is, the hazard function is

Table 5: Neonatal mortality, summer and fall. Adjusted for social branch, sex, illegitimacy, parity, time period, and subregion.

Covariate	**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
aver	0.011	-0.023	0.977	0.010	0.017
emeantemp	8.909	-0.011	0.989	0.007	0.111
sex					0.000
boy	0.512	0	1	(reference)	
girl	0.488	-0.175	0.840	0.050	
season					0.108
summer	0.502	0	1	(reference)	
fall	0.498	-0.143	0.867	0.089	
subreg					0.224
ume	0.295	0	1	(reference)	
ske	0.492	0.102	1.107	0.060	
inland	0.213	0.048	1.049	0.074	
illeg	0.054	0.537	1.711	0.094	0.000
parity					0.000
1	0.272	0	1	(reference)	
2-4	0.452	-0.170	0.844	0.064	
5+	0.276	0.100	1.105	0.068	
socBranch					0.003
office	0.108	0	1	(reference)	
farming	0.460	0.317	1.373	0.099	
worker	0.432	0.232	1.261	0.098	
Events	1599	TTR	3911		
Max. logLik.	-17292				

constant. To give substance to this claim, a short introduction to the ideas of Bourgeois-Pichat is given.

4.2.1 Bourgeois-Pichat and the log-cube transform

The *log-cube transform* g, suggested by Bourgeois-Pichat (1951a), is defined as

$$g(t) = \log^3(1 + t), \quad 0 \le t \le 365, \tag{1}$$

where t is age in days. A graph of the transform is shown in Figure 10.

It was used as a tool for dividing infant mortality into *endogenous* and *exogenous* causes, and a an important part in that venue was the observations (i) postneonatal mortality is purely exogenous, and (ii) the distribution of exogenous mortality is *truncated uniform* (right truncated at $t = g(365)$) on the g-transformed time scale.

Today, assumption (ii) is slightly outdated, for three reasons: (a) The uniform distribution is not a very practical tool in survival analysis, (b) replacing the uniform with an *exponential* not only fits better (usually), it also is a survival distribution that is extremely easy to work with in survival analysis, and (c) “all models are wrong, but some are useful (Box, 1976)”.

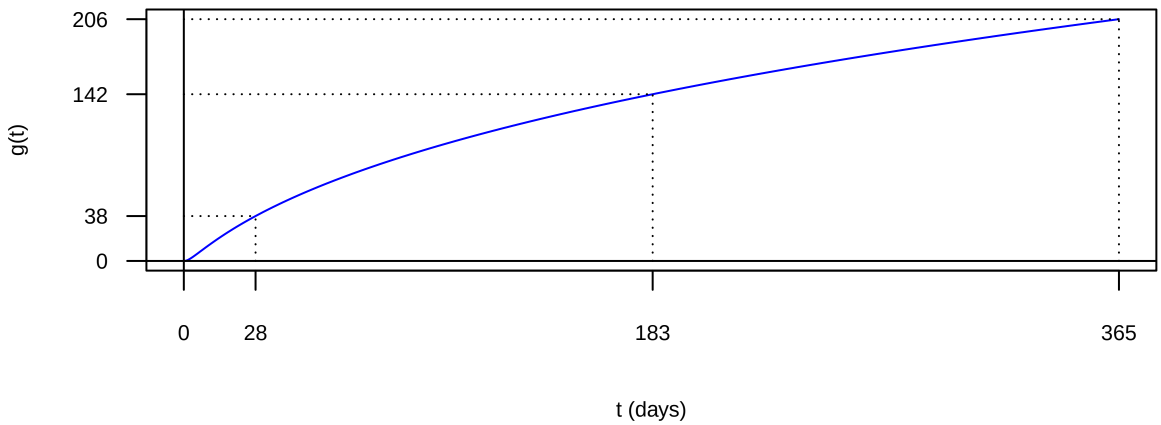

Figure 10: The g transform.

For now, let us look at the basic distribution of postneonatal life on the *g* time scale. We do it nonparametrically by calculating the *Nelson-Aalen* estimator of the cumulative hazards function. See Figure 11, where the two curves are indistinguishable.

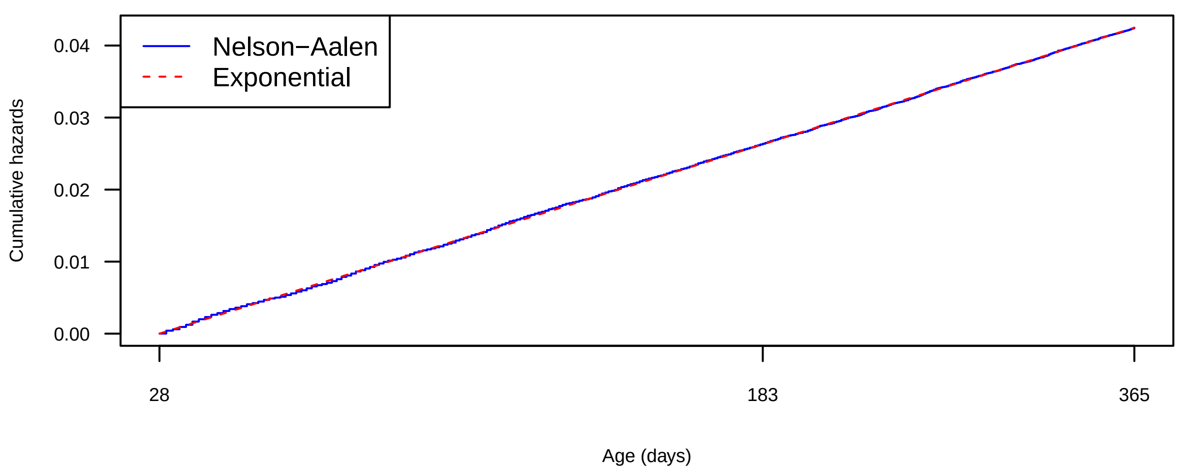

Figure 11: Nelson-Aalen estimator and exponential cumulative hazards of postneonatal survival on the g scale.

The survival data set we work with here contains 102 thousand infants and 4.9 million records, an average of 48 records per infant, one record per calendar week in life. In total there are 4260 deaths, and a typical Cox regression execution time is slightly below one minute. A “naive” application of an exponential regression model to the same data is typically using around six minutes.

However, with the exponential model (or generally, the *piecewise constant hazard* (pch)) in mind, data can be heavily reduced by *tabulation*, and the statistical *sufficiency principle* will guarantee that we get identical results. The tabulation itself takes about 20 seconds, and the analysis after that about 15 seconds, a total of 35 seconds. But the tabulation usually only needs to be done once. More about this later.

4.2.2 Winter and Spring

Table 6: Postneonatal mortality, winter and spring. Adjusted for social status and branch, sex, illegitimacy, parity, time period, and subregion.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
lowTemp					0.030
FALSE	0.875	0	1	(reference)	
TRUE	0.125	0.143	1.153	0.065	
highTemp					0.036
FALSE	0.914	0	1	(reference)	
TRUE	0.086	−0.181	0.835	0.087	
aver	0.005	−0.010	0.990	0.006	0.092
emeantemp	−2.949	−0.016	0.984	0.004	0.000
sex					0.000
boy	0.512	0	1	(reference)	
girl	0.488	−0.228	0.796	0.040	
socBranch					0.000
office	0.107	0	1	(reference)	
farming	0.465	0.512	1.669	0.085	
worker	0.428	0.506	1.658	0.084	
illeg	0.053	0.549	1.731	0.080	0.000
parity					0.000
1	0.265	0	1	(reference)	
2-4	0.455	0.199	1.221	0.055	
5+	0.280	0.615	1.850	0.057	
subreg					0.000
ume	0.294	0	1	(reference)	
ske	0.488	−0.301	0.740	0.046	
inland	0.218	−0.250	0.779	0.056	
Events	2549	TTR	8351030		
Max. logLik.	−23013				

The result in Table 6 shows that *climate* (`emintemp`) is more important than *weather* (`excessTemp`). Moreover, no signs of interaction between weather or climate and the rest of covariates (not shown).

4.2.3 Summer and Fall

The result in Table 7 shows that *climate* (`emintemp`) is more important than *weather* (`excessTemp`). Moreover, no signs of interaction between weather or climate and the rest of covariates (not shown).

5 The biometric analysis of infant mortality

A modern version of the Bourgeois-Pichat biometric modelling (Bourgeois-Pichat, 1951a,b) is given here, *without* the "log-cube transform". A first try is fit a *Weibull* distribution to the time to death for an infant alive at age 28 /365 years (28 days), that is, time is measured

Table 7: Postneonatal mortality, winter and spring. Adjusted for social status and branch, sex, illegitimacy, parity, time period, and subregion.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
lowTemp					0.008
FALSE	0.975	0	1	(reference)	
TRUE	0.025	0.398	1.489	0.144	
highTemp					0.102
FALSE	0.983	0	1	(reference)	
TRUE	0.017	0.296	1.344	0.174	
aver	0.014	−0.009	0.991	0.010	0.415
emeantemp	8.764	0.005	1.005	0.004	0.215
sex					0.000
boy	0.511	0	1	(reference)	
girl	0.489	−0.176	0.839	0.049	
socBranch					0.000
office	0.107	0	1	(reference)	
farming	0.463	0.488	1.629	0.102	
worker	0.430	0.559	1.748	0.101	
illeg	0.054	0.618	1.855	0.094	0.000
parity					0.000
1	0.266	0	1	(reference)	
2-4	0.454	0.276	1.318	0.068	
5+	0.280	0.670	1.955	0.071	
subreg					0.000
ume	0.294	0	1	(reference)	
ske	0.486	−0.392	0.676	0.055	
inland	0.220	−0.468	0.626	0.070	
Events	1711	TTR	8445678		
Max. logLik.	−16141				

from age 28 days until death or time reaches 365 - 28 days, when follow-up is censored. In **R** language, with the package *eha* (Broström, 2021a), we try

```
fit.w <- phreg(Surv(enter - 28/365, exit - 28/365, event) ~ 1, data = pneo)
fit <- coxreg(Surv(enter - 28/365, exit - 28/365, event) ~ 1, data = pneo)
```

where `phreg` performs the Weibull regression (no covariates) and `coxreg` the Cox regression. In Figure 12 the result is shown in the form of the two cumulative hazards estimates.

Not a very good fit. However, we need a *left-truncated Weibull distribution*, that is, assuming that *exogeneous infant mortality* is Weibull distributed.

```
fit.w <- phreg(Surv(enter, exit, event) ~ 1, data = pneo)
fit <- coxreg(Surv(enter, exit, event) ~ 1, data = pneo)
```

Note the difference: No subtraction of the 28 days from the time variables, which implies that the time origin is *birth* and *not* the 28th day of life. Thus our data are *left truncated* at 28 days, and the Weibull attempt estimates neonatal exogeneous mortality through *extrapolation*.

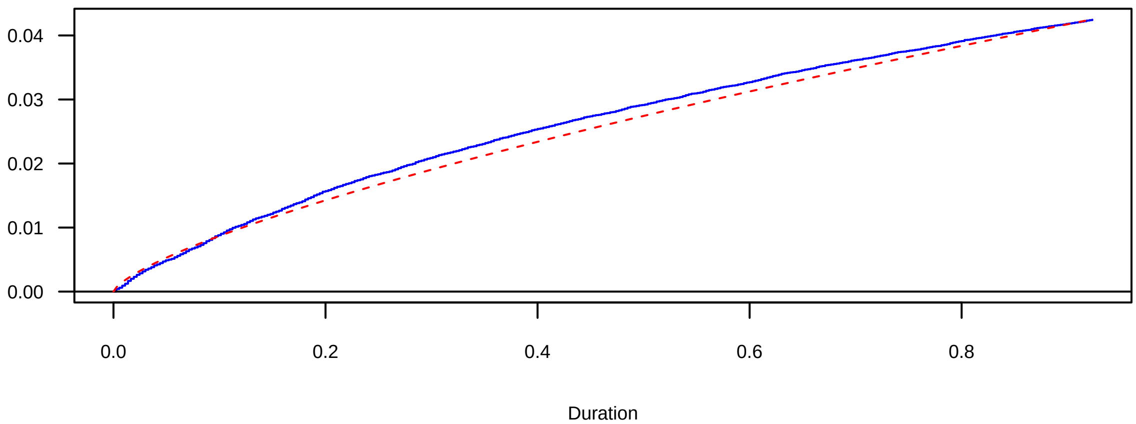

Figure 12: Postneonatal survival, Weibull vs. Nelson-Aalen.

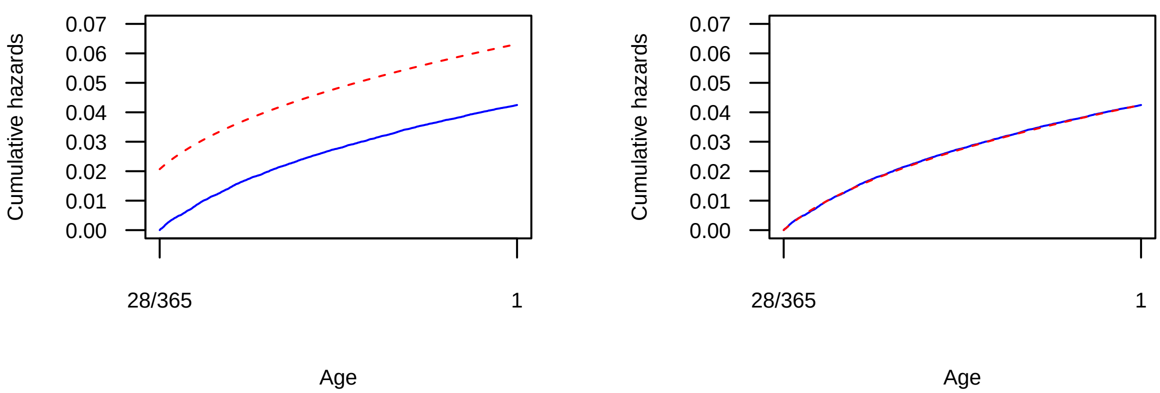

Figure 13: Postneonatal survival, truncated Weibull vs. Nelson-Aalen.

The Cox regression approach is a non-parametric procedure, and thus not capable of any extrapolated estimation, so neonatal mortality will simply be taken as zero.

Look here (Figure 13)! A perfect fit, just like the one under the g- transform. But the Weibull curve does not start at zero (left hand panel), however it is parallell to the Nelson-Aalen estimator, and the right hand panel shows what happens if we subtract the start value from the Weibull curve and redraw.

5.1 Separate exogeneous and endogeneous Infant mortality

So, finally, we show the full procedure how to perform the "biometric analysis" *without* the log-cube transform. It simply consists of two steps:

1. Run a Cox regression on the full data set.
2. Run a *left truncated* (at 28 days) Weibull regression.
3. Plot the two cumulative hazards functions in the same graph and read their crossings of the vertical line at age = 1 year (365 days). Figure 14.

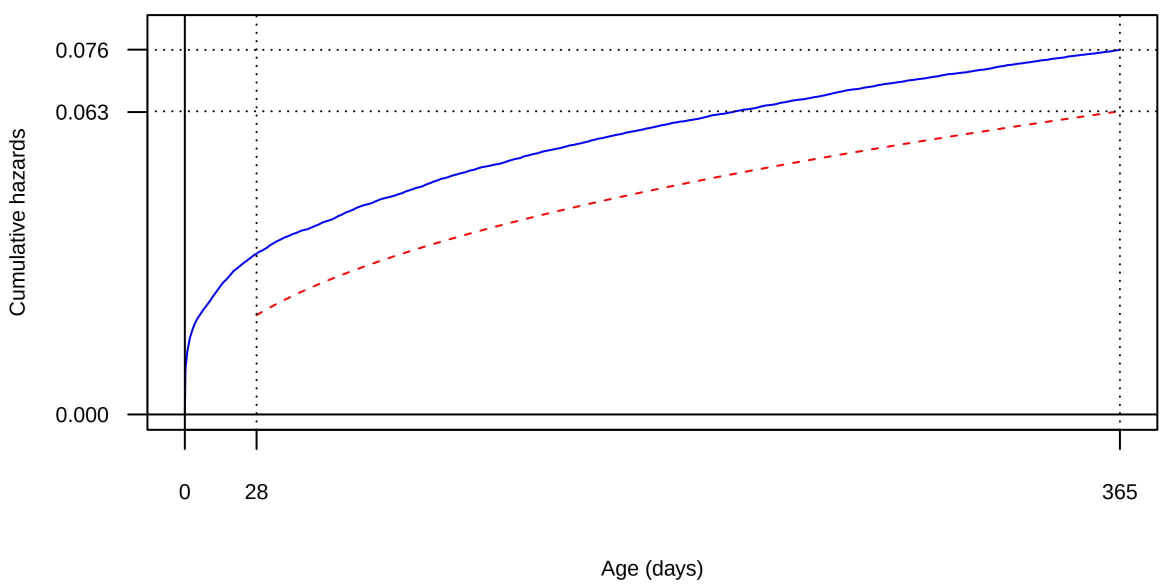

Figure 14: Biometric analysis.

So, total mortality is 0.076, exogeneous mortality is 0.063, and endogeneous mortality is the difference, 0.013, or 17 per cent of the total mortality.

Now, try the original "biometric" approach and compare *(Exercise)*.

6 Conclusion

Remains to be written, especially the discussion about temperature and mortality.

References

Bengtsson, T. and Broström, G. (2010). Mortality crisis in rural southern Sweden 1766–1860. In Kurosu, T., Bengtsson, T., and Campbell, C., editors, *Demographic Response to Economic and Environmental Crisis*, pages 1–16. Reitaku University Press, Kashiwa.

Bengtsson, T. and Broström, G. (2011). Famines and mortality crises in 18th to 19th century southern Sweden. *Genus*, 67:119–139.

Bourgeois-Pichat, J. (1951a). La mesure de la mortalité infantile. I. Principes et métodes. *Population (French Edition)*, 6:233–248.

Bourgeois-Pichat, J. (1951b). La mesure de la mortalité infantile. II. Les causes de décès. *Population (French Edition)*, 6:459–480.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, pages 791–799.

Broström, G. (2021a). *eha: Event History Analysis*. R package version 2.9.0. https://CRAN.R-project.org/package=eha.

Broström, G. (2021b). *Event History Analysis with R, Second Edition*. Chapman & Hall/CRC, Boca Raton.

Junkka, J., Karlsson, L., Lundevaller, E., and Schumann, B. (2021). Climate vulnerability of Swedish newborns: Gender differences and time trends of temperature-related neonatal mortality, 1880–1950. *Environmental Research*, 192. article id 110400.

Karlsson, L., Junkka, J., Schumann, B., and Häggström Lundevaller, E. (2021). Socioeconomic disparities in climate vulnerability: neonatal mortality in northern Sweden, 1880–1950. *Population and Environment*, 10.1007.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.