A biometric analysis of infant mortality and temperature, northern Sweden 1895-1950

Göran Broström and Tommy Bengtsson

2022-01-15

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, 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 climate variations in the same geographical area 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 in a joint analysis, the classification into endogenous and exogenous factors will be discussed.

One important reason for studying neonatal and postneonatal mortality separately is the findings by Bourgeois-Pichat (Bourgeouis-Pichat, 1951a,b) about endogenous and exogenous mortality and the log-cube transform, so we start by recollecting his results and formulate them in a modern statistical setting.

2 The Bourgeois-Pichat model

Bourgeois-Pichat (*B-P*) suggests that mortality during the last eleven months (28 days) of infancy follows a *uniform* distribution given a specific transformation of time (age), the *log-cube transformation* g.

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

where t is age measured in *days*. Its graph is shown in Figure 1.

[Figure 1 about here.]

The *B-P* model specifies that, on the transformed age scale, the *exogeneous* mortality distribution, conditional on death before age 365 days, is *uniform*, that is, the survivor function of T is *linear*,

$$S(t) = 1 - \alpha t, \; 0 \le t \le g(365). \tag{1}$$

Furthermore, it is assumed that the *endogeneous* component is non-zero only during the first month of life.

Next we apply the model to our infant mortality data from the Umeå-Skellefteå region.

2.1 Infant mortality, a first look

From the Umeå-Skellefteå region, we follow all persons born in the region 1895–1950 until death, reaching the age of 1 year, or leaving follow-up for whatever reason, whichever comes first. The first few lines (infants) are shown in Table 1. Note that covariates are left out for the time being.

If `event == TRUE`, then the infant died at age (in days) given by `exit`, otherwise the infant survived (at least) 365 days and follow-up is lost at `exit == 365`. So, in Table 1, infants No. 1–4 survived, while infant No. 5 died at age 162 days. We assume here that all infants are followed from birth, that is, `enter == 0` in all cases (no left truncation).

Table 1: A few infants in the data set.

id	enter	exit	event
1	0	365	FALSE
2	0	365	FALSE
3	0	365	FALSE
4	0	365	FALSE
5	0	162	TRUE

2.1.1 The Bourgeois-Pichat procedure

The BP procedure consists of three steps:

1. Order the data set ascending by `exit`,
2. apply the log-cube transformation to the age data, and
3. plot the cumulative sum of the number of deaths vs. age.

The result is shown in Figure 2, together with the original data. The transform data is stretced stretched to 365 for easier comparison.

[Figure 2 about here.]

The curves are the empirical *cumulative distribution function (CDF)* expressed on the two scales, that is, one minus the *Kaplan-Meier (Kaplan and Meier, 1958) estimates* of the *survivor functions.*

2.1.2 Postneonatal mortality

We begin by studying mortality after age 28 days on the g-scale, *B-P* predicts that it is approximately *uniform.* The uniform distribution, however, is not very useful in a survival analysis context, so we will investigate a close alternative, the *exponential* distribution. Modifying the *B-P* plot by using successive divisions by current *risk set size* instead of original sample size should result in an approximately linear graph if the exponential assumption is valid.

[Figure 3 about here.]

As is evident from Figure 3, the difference between the the two curves is negligible, and both are well approximated by a straight line, so we will from now on work with the assumption that the transformed postneonatal data are *exponentially distributed.* There is also a mathematical argument motivating the similarity: The cdf for an exponentially distributed random variable T is

$$F_T(t) = 1 - \exp(-\lambda t), \quad t, \lambda > 0,$$

and the first order approximation of this expression is

$$F_T(t) \approx \lambda t, \quad t, \lambda > 0,$$

which is the *cumulative hazards* function of T, and the approximation is good for small values of λt.

So, the *hazard function* (which is the slope of the lines in Figure 3) is simply estimated by the ratio of the *number of deaths* to the *total time of exposure* (an *occurrence/exposure* rate):

```
posthaz <- with(ginfdat, sum(event) / sum(exit - enter))
```

which evaluates to 2.56e-04. A small number, but remember that the time unit is defined on the scale defined by the function g (close to *day*).

2.1.3 Neonatal mortality

Is the sum of two components, *external* and *internal* causes. According to the *B-P* theory, the external component is equal to the postneonatal mortality, which contains only an external component.

We create the neonatal data set by censor all lives at the age 28:

```
neo <- age.window(infdat, c(0, 28))
neo$enter <- g(neo$enter)
neo$exit <- g(neo$exit)
```

and then fit a piecewise constant hazards model with cutpoints at $g(0, 2, 4, 7, 11, 16, 22, 28)$, see Figure 4.

[Figure 4 about here.]

Let us take a closer look at what happens at the end of the neonatal interval, see Figure 5. It shows a smooth transition towards the postneonatal (external) mortality.

[Figure 5 about here.]

2.1.4 Putting all together

Looking at the full infant mortality interval on the g-scale, we have a piecewise constant hazards model with cutpoints at $g(0, 1, 2, 4, 7, 11, 16, 22, 28)$, see Figure 6 (where the value close to zero is truncated).

[Figure 6 about here.]

Now we would like to see the content of Figure 6 on the natural time scale. If $Y = g(T)$, we have the relation between their respective hazard functions

$$h_T(t) = \frac{d}{dt}g(t)h_Y\big(g^{-1}(t)\big), \quad 0 \le t \le 365,$$

and with h_Y shown in Figure 6, we get h_T in Figure 7 **(must be checked!)**.

[Figure 7 about here.]

3 Data

3.1 Infant mortality, summary graphs

Figure 8 shows the average monthly crude infant mortality, and a clear seasonal pattern is visible.

[Figure 8 about here.]

The average monthly neonatal mortality is shown in Figure 9.

[Figure 9 about here.]

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

The average monthly postneonatal mortality is shown in Figure 10.

[Figure 10 about here.]

The seasonal pattern is once again similar to the one we found for infant mortality. Next, the decline over the years in Figures 11 and 12.

[Figure 11 about here.]

[Figure 12 about here.]

3.2 Temperature

Temperature data are collected from two weather stations, *Umeå* and *Bjuröklubb* (used with population data from Skellefteå coastal area). Both stations deliver daily temperature data covering our time period, usually three measures per day, morning, noon, and evening. Weekly averages are calculated year by year, and deviations from the averages of the weekly averages are used as a time-varying *communal covariate*. See Figure 13 for the monthly distributions.

[Figure 13 about here.]

4 Infant mortality, climate and weather

The analyses are performed on the log-cube scale 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.

We start by performing separate analyses for neonatal and postneonatal mortality. Then we make comparisons.

4.1 Postneonatal mortality

This is the simplest part, because *B-P* predicts that the baseline distribution (given the log-cube transformation) is *exponential*, that is, the hazard function is *constant*.

The result in Table 2 shows that *climate* (`emintemp`) is more important than *weather* (`cold`). Moreover, no signs of interaction between weather or climate and the rest of covariates (not shown).

4.2 Neonatal mortality

It turns out that here we have some significant interactions, with `sex` involved. Maybe surprising. So we run separate analyses for the sexes.

4.2.1 Girls

Same here, *climate* is more important than *weather*.

4.2.2 Boys

Here, however, boys are sensitive to temperatures below average, in addition to the sensitivity for cold climate.

Table 2: Postneonatal mortality.

Covariate		Mean	Coef	H.R.	S.E.	L-R p
sex						0.000
	boy	0.511	0	1	(reference)	
	girl	0.489	−0.237	0.789	0.035	
socBranch						0.000
	official	0.085	0	1	(reference)	
	farming	0.424	0.615	1.850	0.097	
	business	0.039	0.385	1.469	0.130	
	worker	0.452	0.622	1.863	0.091	
socStatus						0.000
	low	0.414	0	1	(reference)	
	high	0.586	−0.210	0.810	0.048	
illeg		0.044	0.492	1.635	0.075	0.000
parity						0.000
	1	0.266	0	1	(reference)	
	2-4	0.464	0.151	1.163	0.049	
	5+	0.270	0.472	1.603	0.051	
period						0.000
	(1.9e+03,1.91e+0	0.342	0	1	(reference)	
	(1.91e+03,1.94e+	0.365	−0.187	0.829	0.038	
	(1.94e+03,1.95e+	0.293	−1.157	0.315	0.060	
cold		0.027	−0.008	0.992	0.004	0.052
emintemp		−2.253	−0.020	0.980	0.004	0.000
season						0.000
	winter	0.249	0	1	(reference)	
	spring	0.249	0.129	1.137	0.060	
	summer	0.251	0.207	1.230	0.110	
	fall	0.251	−0.238	0.788	0.071	
subreg						0.000
	ume	0.372	0	1	(reference)	
	ske	0.628	−0.333	0.717	0.036	
Events		3285	TTR	12780368		
Max. logLik.		−29861				

4.3 Postneonatal mortality, winter and spring

This is the simplest part, because *B-P* predicts that the baseline distribution (given the log-cube transformation) is *exponential*, that is, the hazard function is *constant*.

The result in Table 5 shows that *climate* (`emintemp`) is more important than *weather* (`cold`). Moreover, no signs of interaction between weather or climate and the rest of covariates (not shown).

4.4 Neonatal mortality, winter and spring

It turns out that here we have some significant interactions, with `sex` involved. Maybe surprising. So we run separate analyses for the sexes.

Table 3: Neonatal mortality, girls.

Covariate		Mean	Coef	H.R.	S.E.	L-R p
socStatus						0.965
	low	0.416	0	1	(reference)	
	high	0.584	−0.003	0.997	0.080	
illeg		0.043	0.491	1.633	0.122	0.000
parity						0.000
	1	0.261	0	1	(reference)	
	2-4	0.463	−0.274	0.760	0.076	
	5+	0.276	0.024	1.024	0.082	
period						0.000
(1.9e+03,1.91e+0		0.344	0	1	(reference)	
(1.91e+03,1.94e+		0.370	−0.156	0.856	0.066	
(1.94e+03,1.95e+		0.286	−0.619	0.538	0.088	
socBranch						0.322
	official	0.082	0	1	(reference)	
	farming	0.424	0.210	1.234	0.142	
	business	0.040	−0.042	0.958	0.214	
	worker	0.455	0.164	1.178	0.132	
cold		−0.006	0.009	1.009	0.007	0.189
emintemp		−2.124	−0.010	0.990	0.007	0.159
subreg						0.080
	ume	0.371	0	1	(reference)	
	ske	0.629	0.112	1.118	0.064	
season						0.207
	winter	0.246	0	1	(reference)	
	spring	0.254	−0.012	0.988	0.106	
	summer	0.252	0.088	1.092	0.188	
	fall	0.249	−0.109	0.897	0.124	
Events		1142	TTR	1461345		
Max. logLik.		−11993				

4.4.1 Girls

Same here, *climate* is more important than *weather*.

4.4.2 Boys

Here, however, boys are sensitive to temperatures below average, in addition to the sensitivity for cold climate.

5 Conclusion

Remains to be written. Feels like two papers, one about Bourgeouis-Pichat and one about climate and weather. Maybe discuss exogeneity/endogeneity?

Table 4: Neonatal mortality, boys.

Covariate	**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
socStatus					0.264
low	0.412	0	1	(reference)	
high	0.588	0.078	1.081	0.070	
illeg	0.046	0.256	1.292	0.113	0.028
parity					0.000
1	0.269	0	1	(reference)	
2-4	0.461	−0.261	0.770	0.066	
5+	0.270	−0.036	0.965	0.071	
period					0.000
(1.9e+03,1.91e+0	0.347	0	1	(reference)	
(1.91e+03,1.94e+	0.363	−0.094	0.910	0.058	
(1.94e+03,1.95e+	0.290	−0.566	0.568	0.076	
socBranch					0.390
official	0.085	0	1	(reference)	
farming	0.427	0.105	1.111	0.121	
business	0.038	0.270	1.310	0.168	
worker	0.450	0.146	1.158	0.111	
cold	0.017	−0.017	0.983	0.006	0.006
emintemp	−2.258	−0.005	0.995	0.006	0.364
subreg					0.355
ume	0.377	0	1	(reference)	
ske	0.623	0.051	1.052	0.055	
season					0.301
winter	0.247	0	1	(reference)	
spring	0.258	−0.044	0.957	0.092	
summer	0.245	−0.103	0.902	0.164	
fall	0.250	−0.165	0.848	0.107	
Events	1494	TTR	1539122		
Max. logLik.	−15789				

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.

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

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

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.

Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481.

Table 5: Postneonatal mortality, winter and spring.

Covariate	**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
sex					0.000
boy	0.512	0	1	(reference)	
girl	0.488	−0.266	0.766	0.046	
socBranch					0.000
official	0.085	0	1	(reference)	
farming	0.426	0.629	1.876	0.127	
business	0.039	0.344	1.411	0.173	
worker	0.451	0.622	1.862	0.119	
socStatus					0.003
low	0.413	0	1	(reference)	
high	0.587	−0.189	0.828	0.063	
illeg	0.044	0.505	1.656	0.098	0.000
parity					0.000
1	0.266	0	1	(reference)	
2-4	0.464	0.104	1.110	0.064	
5+	0.270	0.462	1.587	0.066	
period					0.000
(1.9e+03,1.91e+0	0.344	0	1	(reference)	
(1.91e+03,1.94e+	0.366	−0.177	0.838	0.049	
(1.94e+03,1.95e+	0.290	−1.060	0.346	0.077	
cold	0.021	−0.011	0.989	0.004	0.014
emintemp	−9.101	−0.012	0.988	0.003	0.000
subreg					0.000
ume	0.372	0	1	(reference)	
ske	0.628	−0.299	0.741	0.047	
Events	1928	TTR	6363114		
Max. logLik.	−17264				

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.

Table 6: Neonatal mortality winter and spring, girls.

Covariate		**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
socStatus						0.977
	low	0.415	0	1	(reference)	
	high	0.585	0.003	1.003	0.109	
illeg		0.045	0.598	1.818	0.159	0.000
parity						0.040
	1	0.253	0	1	(reference)	
	2-4	0.467	−0.197	0.821	0.105	
	5+	0.280	0.023	1.023	0.114	
period						0.000
(1.9e+03,1.91e+0		0.345	0	1	(reference)	
(1.91e+03,1.94e+		0.367	−0.213	0.808	0.090	
(1.94e+03,1.95e+		0.288	−0.749	0.473	0.122	
socBranch						0.536
	official	0.082	0	1	(reference)	
	farming	0.426	0.013	1.013	0.185	
	business	0.039	−0.359	0.699	0.304	
	worker	0.452	0.013	1.013	0.170	
cold		−0.031	0.010	1.010	0.008	0.222
emintemp		−9.033	−0.012	0.988	0.006	0.048
subreg						0.080
	ume	0.375	0	1	(reference)	
	ske	0.625	0.152	1.164	0.088	
Events		612	TTR	730108		
Max. logLik.		−5999				

Table 7: Neonatal mortality winter and spring, boys.

Covariate		**Mean**	**Coef**	**H.R.**	**S.E.**	**L-R p**
socStatus						0.212
	low	0.409	0	1	(reference)	
	high	0.591	0.117	1.124	0.093	
illeg		0.046	0.053	1.055	0.163	0.745
parity						0.001
	1	0.264	0	1	(reference)	
	2-4	0.462	−0.301	0.740	0.088	
	5+	0.274	−0.073	0.930	0.096	
period						0.000
(1.9e+03,1.91e+0		0.345	0	1	(reference)	
(1.91e+03,1.94e+		0.362	−0.050	0.951	0.078	
(1.94e+03,1.95e+		0.292	−0.650	0.522	0.104	
socBranch						0.720
	official	0.083	0	1	(reference)	
	farming	0.430	−0.010	0.990	0.159	
	business	0.037	0.081	1.084	0.233	
	worker	0.449	0.091	1.095	0.145	
cold		0.013	−0.017	0.983	0.007	0.014
emintemp		−9.014	−0.005	0.995	0.005	0.287
subreg						0.784
	ume	0.378	0	1	(reference)	
	ske	0.622	−0.020	0.980	0.073	
Events		826	TTR	776997		
Max. logLik.		−8165				

List of Figures

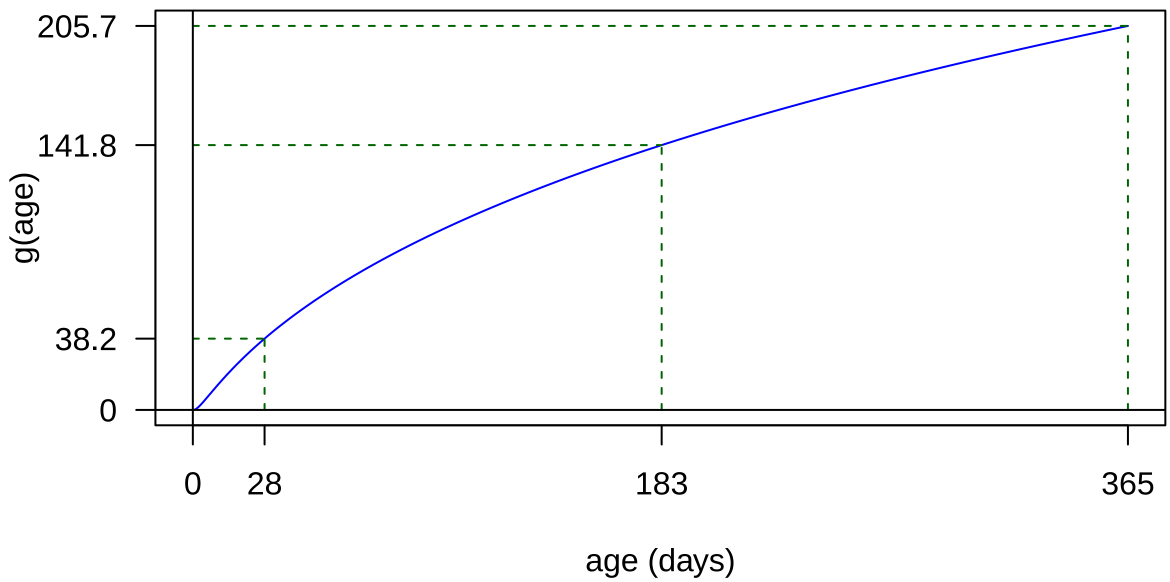

Figure 1: The log-cube transform g.

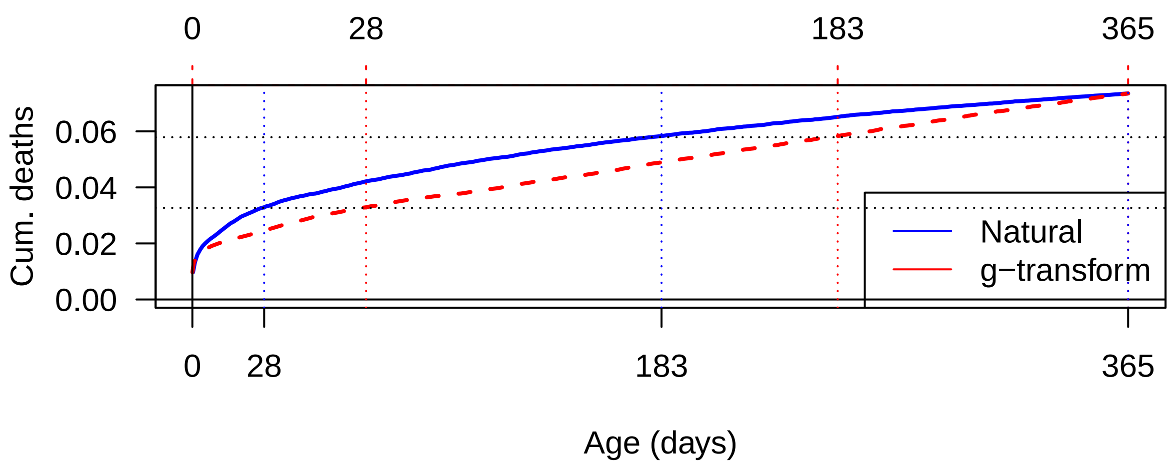

Figure 2: BB plots with and without time transform, Umeå-Skellefteå 1895-1950.

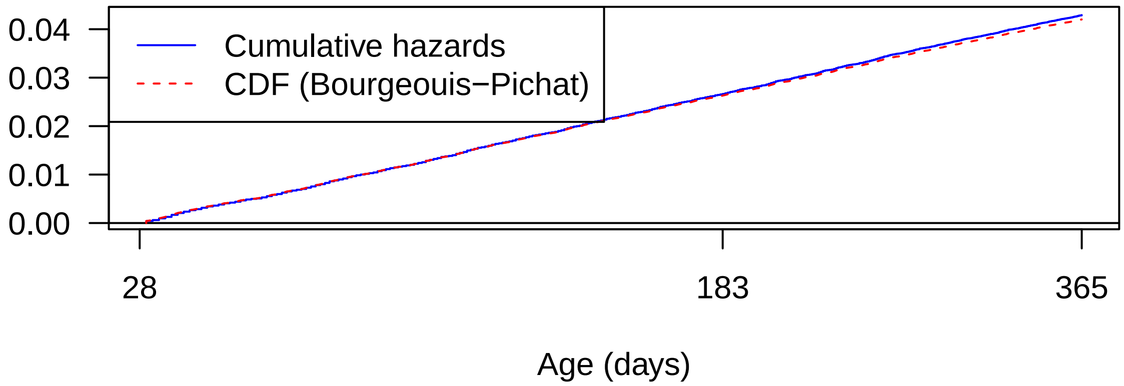

Figure 3: Relation between cumulative hazards and CDF for transformed postneonatal mortality data.

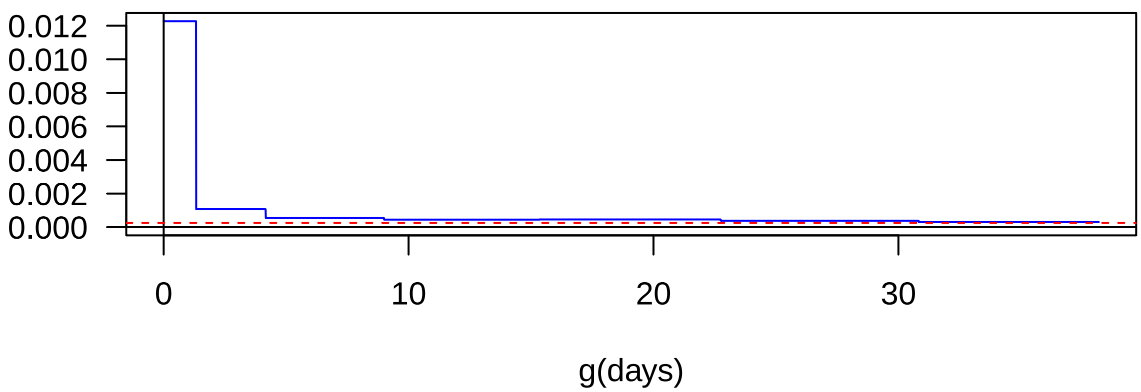

Figure 4: Neonatal hazard function with external component (dashed).

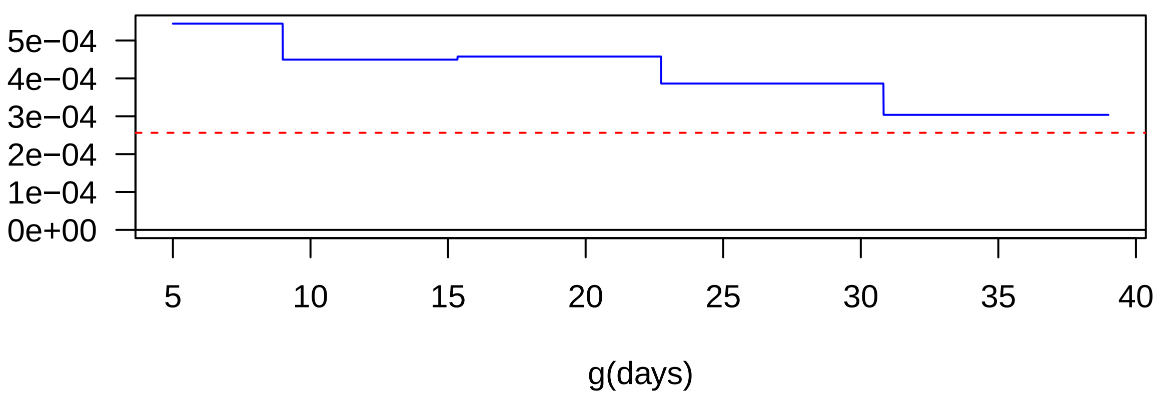

Figure 5: Neonatal hazard function with external component (dashed), right hand tail.

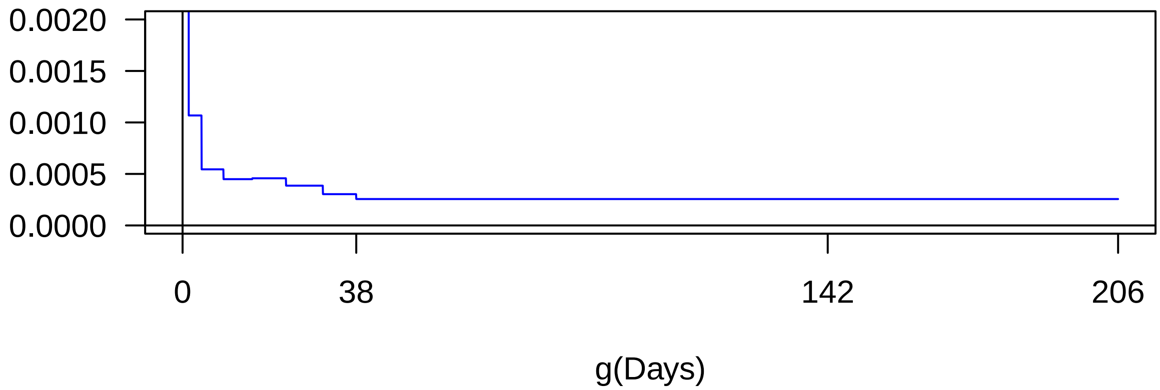

Figure 6: Hazard function, infant mortality, transformed time.

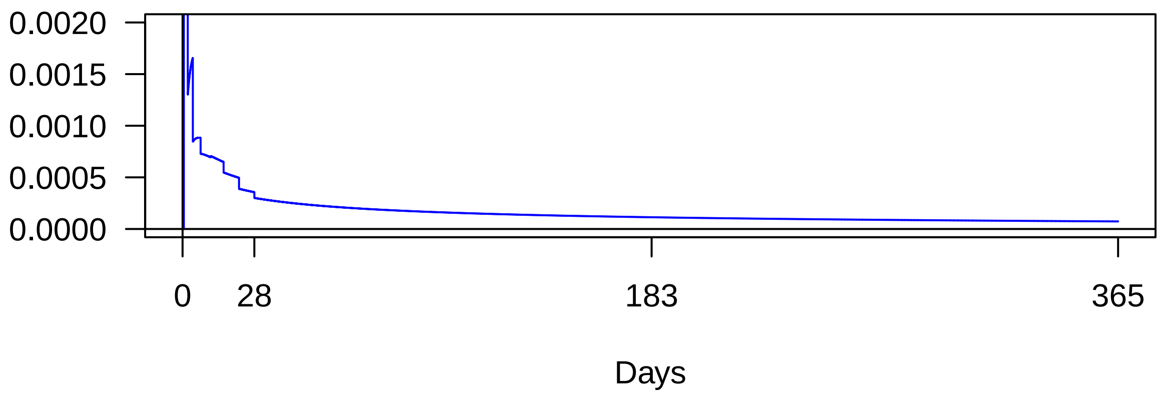

Figure 7: Hazard function, infant mortality, natural time scale.

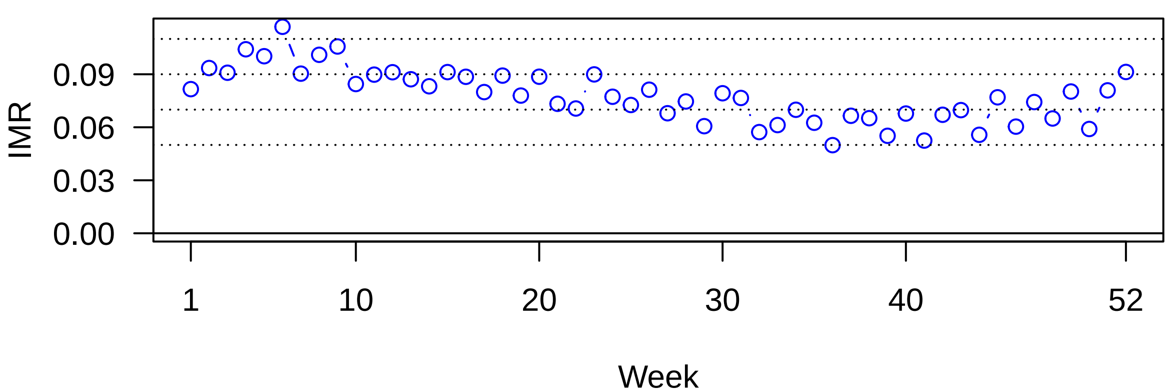

Figure 8: Crude infant mortality by week of year, Umeå/Skellefteå 1895–1950.

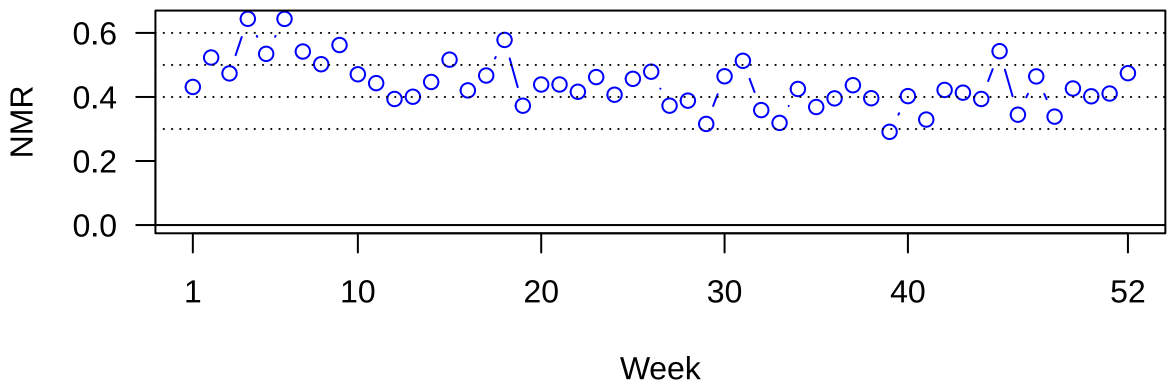

Figure 9: Crude neonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

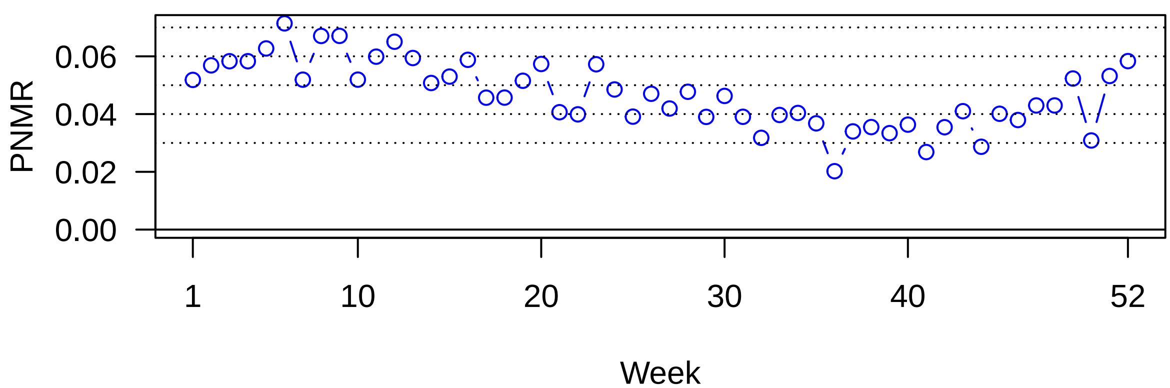

Figure 10: Crude postneonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

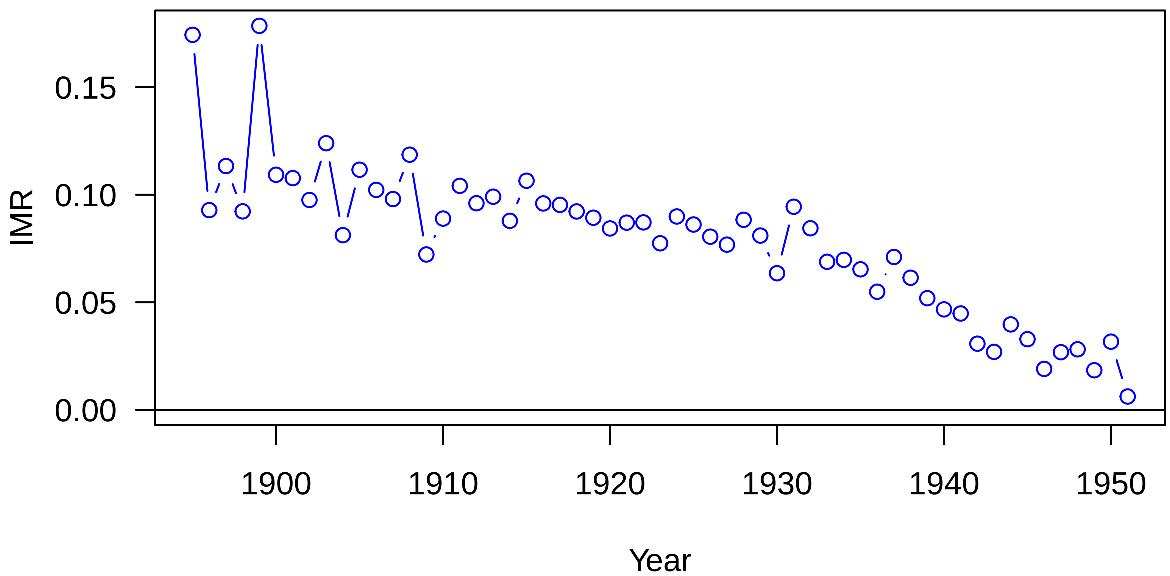

Figure 11: Crude IMR by year, Umeå-Skellefteå 1895–1950.

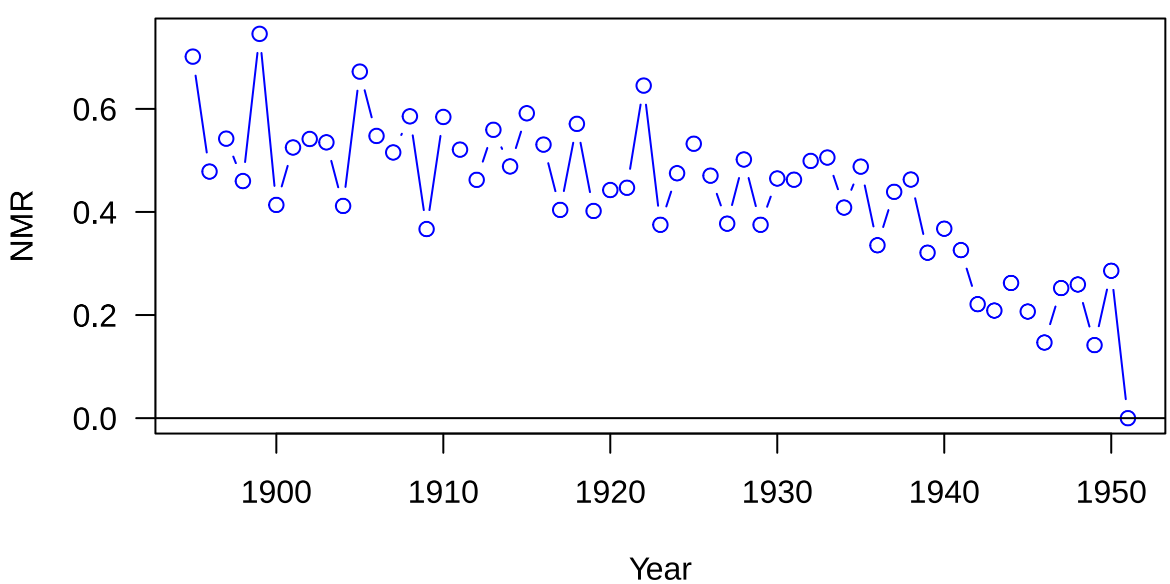

Figure 12: Crude NMR by year, Umeå-Skellefteå 1895–1950.

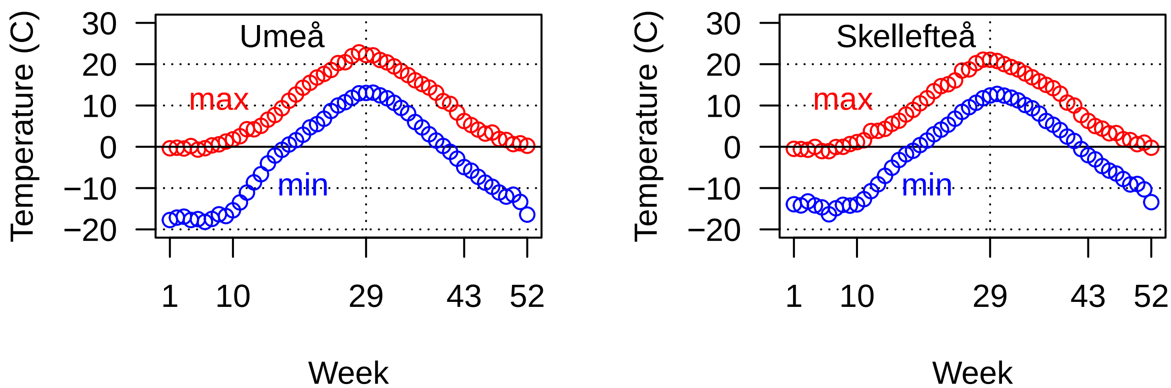

Figure 13: Weekly max and min temperature averages, 1895–1950.