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

Göran Broström and Tommy Bengtsson

2022-01-23 13:45:23

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 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 (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 of infancy follows a *uniform* distribution given a specific transformation of time (age), the *log-cube transformation* g, see equation (2),

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

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 28 days 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 B-P 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 stretched 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

Neonatal mortality is, according to Bourgeois-Pichat, 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 days:

```
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, 5, 8, 12, 17, 22, 25, 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 it all together

Looking at the full infant mortality interval on the g-scale, we have a piecewise constant hazards model with cutpoints at $g(0, 2, 5, 8, 12, 17, 22, 25, 28, 365)$, see Figure 6.

[Figure 6 about here.]

The way to disentangle the endogenous and exogenous components is to start with a plot of the *cumulative* hazard function and add the exogenous part from the postneonatal estimate, see Figure 7.

[Figure 7 about here.]

So we can see from Figure 7 (vertical axis on the right hand side) that of the total mortality, 69 per cent is exogenous and the rest is endogenous. Similarly, of the neonatal mortality, 28 per cent is exogenous and the rest is endogenous (vertical axis on the left hand side). As always, this requires that we trust the Bourgeois-Pichat model.

3 Infant mortality, climate and weather

3.1 Data

3.1.1 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.]

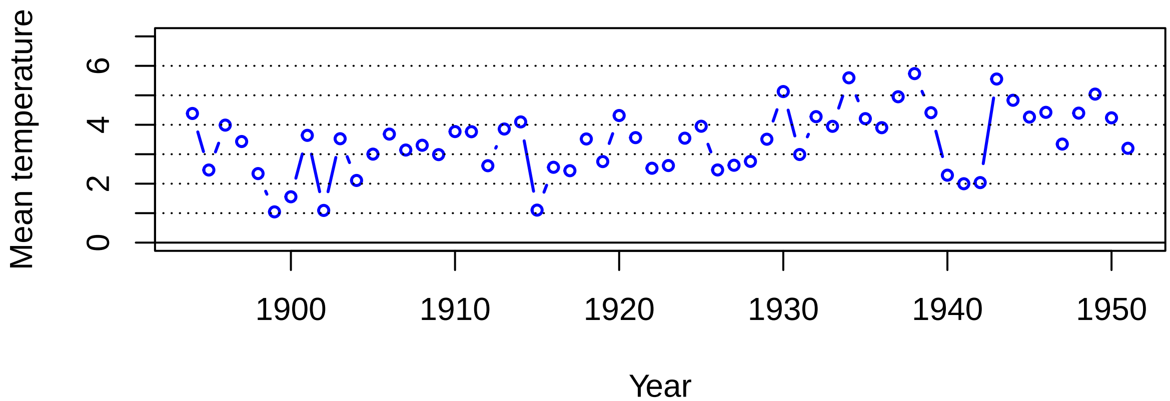
[Figure 12 about here.]

3.1.2 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.]

Time trends of yearly average temperatures, see Figure



3.2 Statistical modelling

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.

3.3 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 **??** 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).

3.4 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.

Table 2: Postneonatal mortality, Skellefteå.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
icold	−1.443	−0.037	0.964	0.008	0.000
emintemp	−1.887	−0.019	0.981	0.006	0.001
season					0.000
winter	0.249	0	1	(reference)	
spring	0.249	0.103	1.108	0.076	
summer	0.250	0.114	1.121	0.147	
fall	0.252	−0.260	0.771	0.097	
period					0.000
(1895,1914]	0.332	0	1	(reference)	
(1914,1935]	0.386	−0.121	0.886	0.051	
(1935,1951]	0.283	−0.941	0.390	0.078	
sex					0.000
boy	0.509	0	1	(reference)	
girl	0.491	−0.199	0.819	0.047	
socBranch					0.009
official	0.054	0	1	(reference)	
farming	0.460	0.469	1.598	0.150	
business	0.030	0.469	1.598	0.193	
worker	0.456	0.417	1.518	0.142	
socStatus					0.000
low	0.421	0	1	(reference)	
high	0.579	−0.257	0.773	0.072	
parity					0.000
1	0.250	0	1	(reference)	
2-4	0.456	0.137	1.146	0.066	
5+	0.293	0.449	1.566	0.068	
Events	1835	TTR	8020541		
Max. logLik.	−16974				

3.4.1 Girls

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

3.4.2 Boys

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

3.5 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 6 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).

Table 3: Postneonatal mortality, Umeå.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
icold	−1.628	−0.011	0.989	0.009	0.231
period					0.000
(1895,1914]	0.358	0	1	(reference)	
(1914,1935]	0.331	−0.235	0.791	0.057	
(1935,1951]	0.311	−1.449	0.235	0.097	
sex					0.000
boy	0.515	0	1	(reference)	
girl	0.485	−0.291	0.747	0.053	
socBranch					0.000
official	0.136	0	1	(reference)	
farming	0.364	0.729	2.073	0.127	
business	0.055	0.250	1.284	0.180	
worker	0.445	0.763	2.144	0.118	
socStatus					0.000
low	0.402	0	1	(reference)	
high	0.598	−0.234	0.791	0.066	
parity					0.000
1	0.293	0	1	(reference)	
2-4	0.476	0.025	1.026	0.069	
5+	0.231	0.329	1.389	0.073	
season					0.000
winter	0.249	0	1	(reference)	
spring	0.249	−0.021	0.980	0.070	
summer	0.251	−0.241	0.786	0.075	
fall	0.251	−0.375	0.687	0.077	
Events	1449	TTR	4759837		
Max. logLik.	−12879				

3.6 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.

3.6.1 Girls

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

3.6.2 Boys

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

Table 4: Neonatal mortality, girls.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
socStatus					0.964
low	0.416	0	1	(reference)	
high	0.584	−0.004	0.996	0.080	
illeg	0.043	0.491	1.635	0.122	0.000
parity					0.000
1	0.261	0	1	(reference)	
2-4	0.463	−0.272	0.762	0.076	
5+	0.276	0.028	1.029	0.081	
period					0.000
(1895,1914]	0.344	0	1	(reference)	
(1914,1935]	0.370	−0.147	0.863	0.066	
(1935,1951]	0.286	−0.606	0.546	0.088	
socBranch					0.314
official	0.082	0	1	(reference)	
farming	0.424	0.213	1.237	0.142	
business	0.040	−0.042	0.959	0.214	
worker	0.455	0.164	1.178	0.132	
cold	−0.002	0.003	1.003	0.007	0.715
emintemp	−2.129	−0.007	0.993	0.003	0.015
subreg					0.093
ume	0.371	0	1	(reference)	
ske	0.629	0.107	1.113	0.064	
Events	1143	TTR	1461345		
Max. logLik.	−12007				

4 Conclusion

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

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.

Table 5: Neonatal mortality, boys.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
socStatus					0.266
low	0.412	0	1	(reference)	
high	0.588	0.078	1.081	0.070	
illeg	0.046	0.258	1.294	0.113	0.027
parity					0.000
1	0.269	0	1	(reference)	
2-4	0.461	−0.260	0.771	0.066	
5+	0.270	−0.034	0.967	0.071	
period					0.000
(1895,1914]	0.347	0	1	(reference)	
(1914,1935]	0.363	−0.098	0.907	0.058	
(1935,1951]	0.290	−0.573	0.564	0.076	
socBranch					0.384
official	0.085	0	1	(reference)	
farming	0.427	0.108	1.114	0.121	
business	0.038	0.272	1.312	0.168	
worker	0.450	0.148	1.160	0.111	
cold	0.022	−0.010	0.990	0.006	0.091
emintemp	−2.263	−0.010	0.990	0.003	0.000
subreg					0.312
ume	0.377	0	1	(reference)	
ske	0.623	0.056	1.057	0.055	
Events	1494	TTR	1539122		
Max. logLik.	−15792				

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

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: 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	−0.263	0.769	0.046	
socBranch					0.000
official	0.085	0	1	(reference)	
farming	0.426	0.634	1.885	0.127	
business	0.039	0.327	1.387	0.174	
worker	0.451	0.619	1.857	0.119	
socStatus					0.003
low	0.413	0	1	(reference)	
high	0.587	−0.184	0.832	0.063	
illeg	0.044	0.502	1.652	0.098	0.000
parity					0.000
1	0.266	0	1	(reference)	
2-4	0.464	0.096	1.101	0.064	
5+	0.270	0.462	1.586	0.066	
period					0.000
(1895,1914]	0.344	0	1	(reference)	
(1914,1935]	0.366	−0.173	0.841	0.049	
(1935,1951]	0.290	−1.057	0.348	0.077	
cold	0.020	−0.015	0.985	0.004	0.001
emintemp	−9.099	−0.013	0.987	0.003	0.000
subreg					0.000
ume	0.372	0	1	(reference)	
ske	0.628	−0.295	0.745	0.047	
Events	1936	TTR	6362768		
Max. logLik.	−17324				

Table 7: Neonatal mortality winter and spring, girls.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
socStatus					0.737
low	0.415	0	1	(reference)	
high	0.585	−0.037	0.964	0.109	
illeg	0.045	0.543	1.722	0.160	0.001
parity					0.013
1	0.253	0	1	(reference)	
2-4	0.467	−0.243	0.784	0.105	
5+	0.280	0.004	1.004	0.113	
period					0.000
(1895,1914]	0.344	0	1	(reference)	
(1914,1935]	0.367	−0.180	0.835	0.090	
(1935,1951]	0.288	−0.712	0.491	0.121	
socBranch					0.467
official	0.082	0	1	(reference)	
farming	0.426	0.010	1.010	0.182	
business	0.039	−0.408	0.665	0.302	
worker	0.453	−0.024	0.977	0.166	
cold	−0.021	0.005	1.005	0.008	0.505
emintemp	−9.036	−0.011	0.989	0.006	0.073
subreg					0.228
ume	0.375	0	1	(reference)	
ske	0.625	0.104	1.110	0.087	
Events	613	TTR	731148		
Max. logLik.	−6012				

Table 8: Neonatal mortality winter and spring, boys.

Covariate	Mean	Coef	H.R.	S.E.	L-R p
socStatus					0.167
low	0.409	0	1	(reference)	
high	0.591	0.130	1.138	0.094	
illeg	0.046	0.031	1.032	0.164	0.849
parity					0.000
1	0.264	0	1	(reference)	
2-4	0.462	−0.315	0.730	0.088	
5+	0.274	−0.075	0.927	0.096	
period					0.000
(1895,1914]	0.345	0	1	(reference)	
(1914,1935]	0.363	−0.061	0.940	0.078	
(1935,1951]	0.293	−0.649	0.523	0.104	
socBranch					0.635
official	0.083	0	1	(reference)	
farming	0.430	0.002	1.002	0.160	
business	0.038	0.071	1.073	0.237	
worker	0.449	0.113	1.119	0.147	
cold	0.016	−0.009	0.991	0.007	0.194
emintemp	−9.021	−0.007	0.993	0.005	0.165
subreg					0.884
ume	0.378	0	1	(reference)	
ske	0.622	−0.011	0.989	0.074	
Events	822	TTR	777881		
Max. logLik.	−8128				

List of Figures

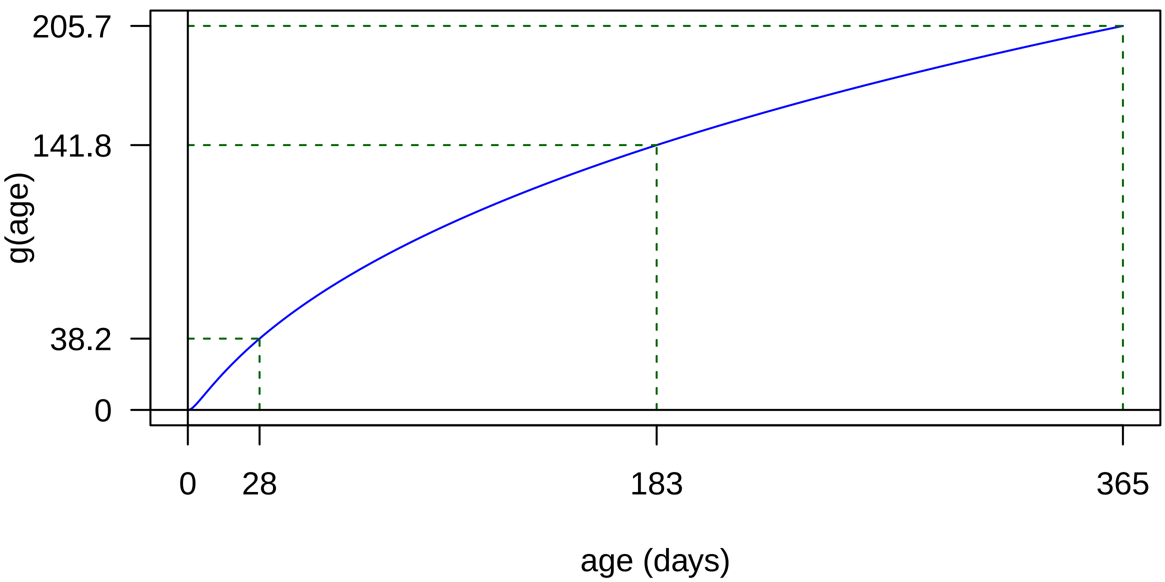


Figure 1: The log-cube transform g.

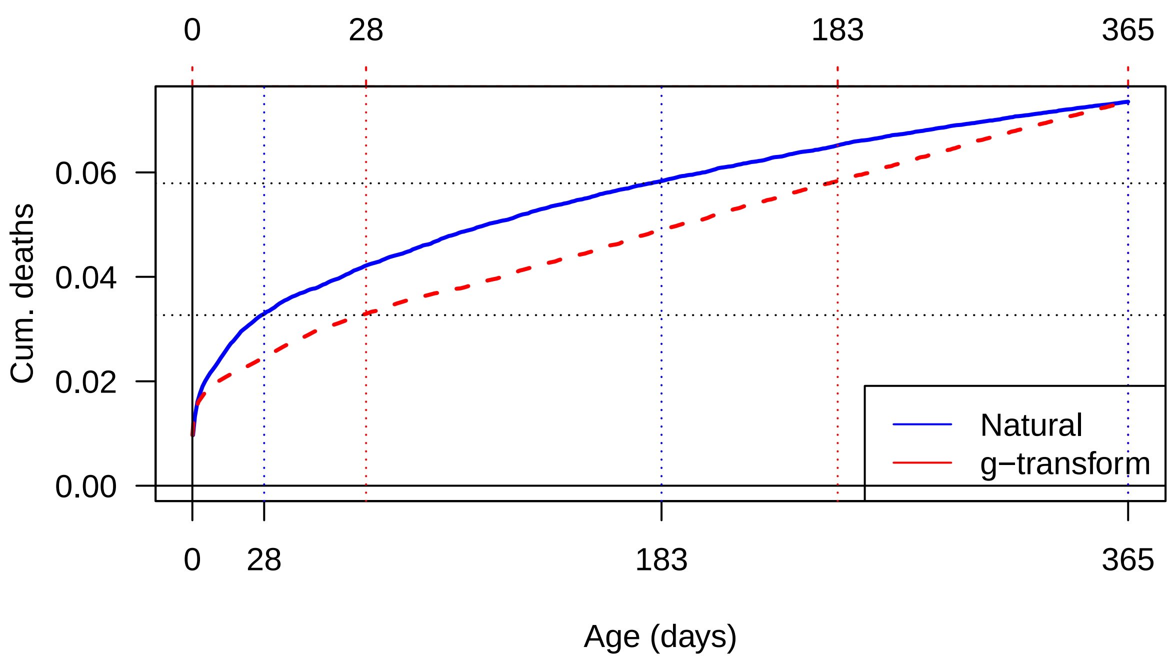


Figure 2: B-B plots with and without time transform, Umeå-Skellefteå 1895-1950.

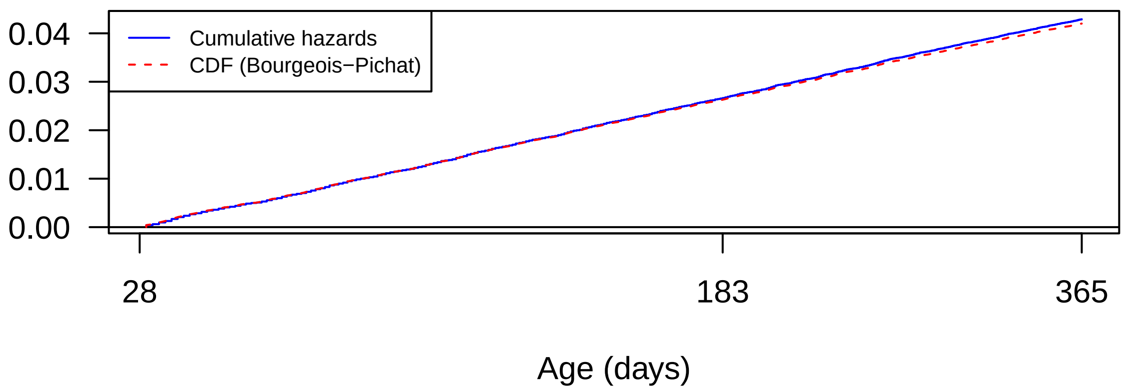


Figure 3: Relation between cumulative hazards and CDF for transformed postneonatal mortality data.

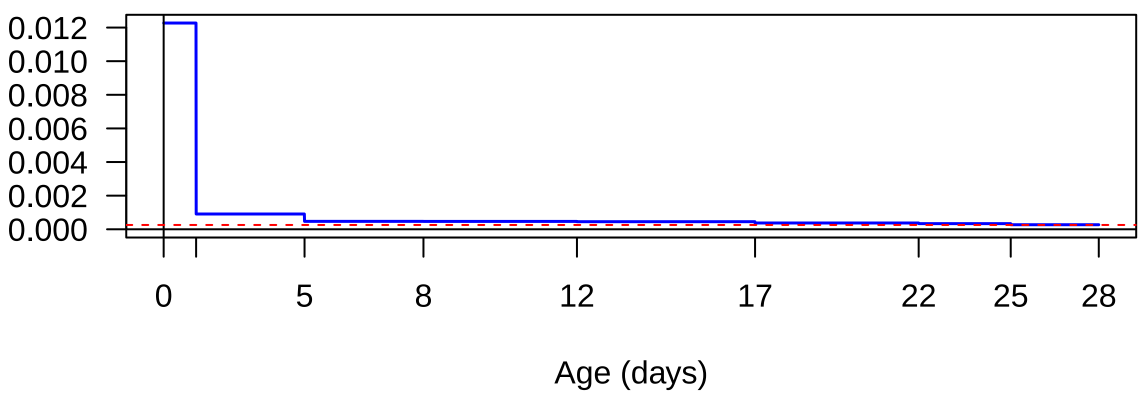


Figure 4: Neonatal hazard function with external component (dashed).

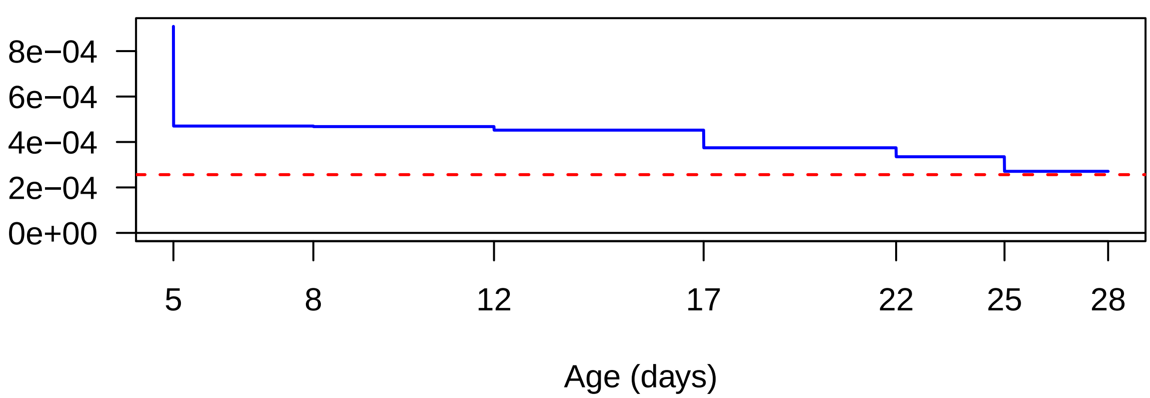


Figure 5: Neonatal hazard function with external component (dashed), right hand tail.

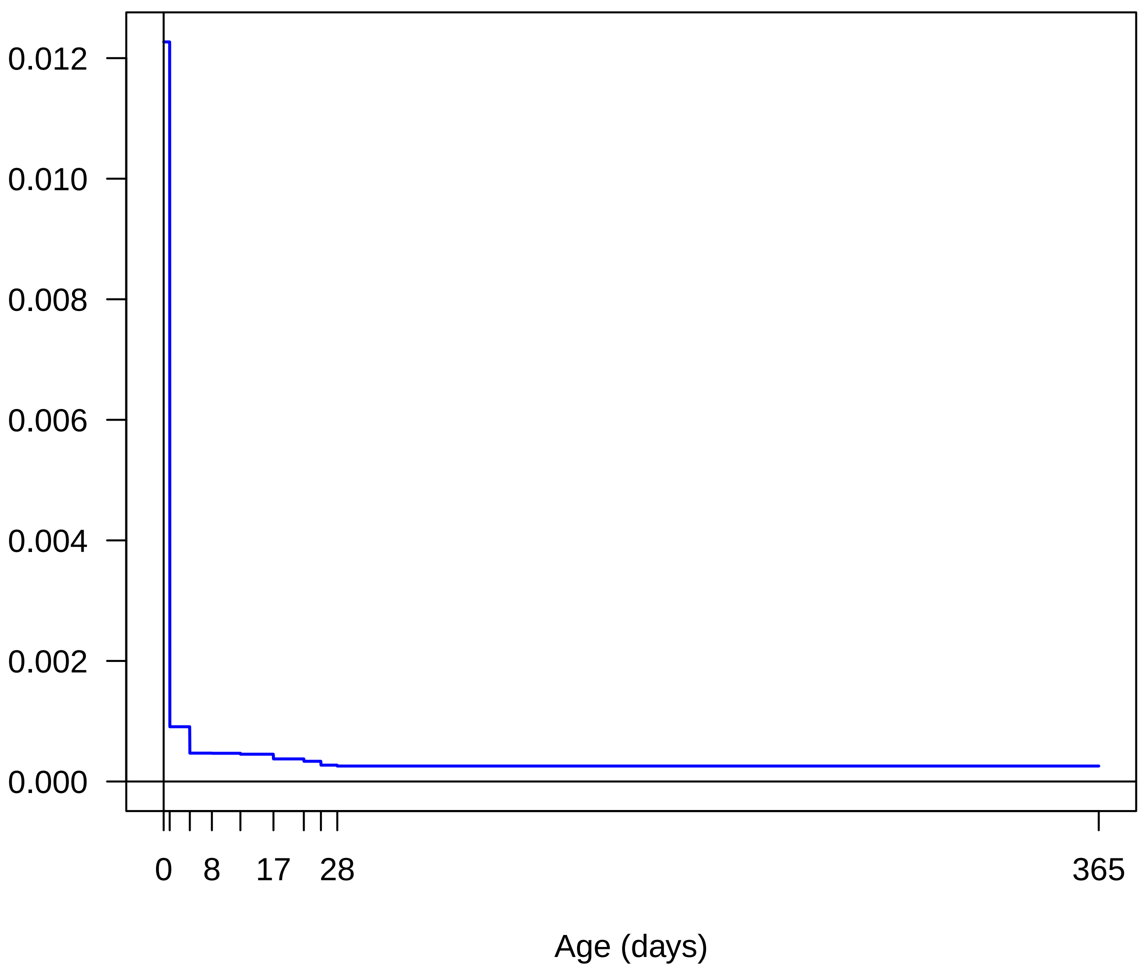


Figure 6: Hazard function, infant mortality, transformed time.

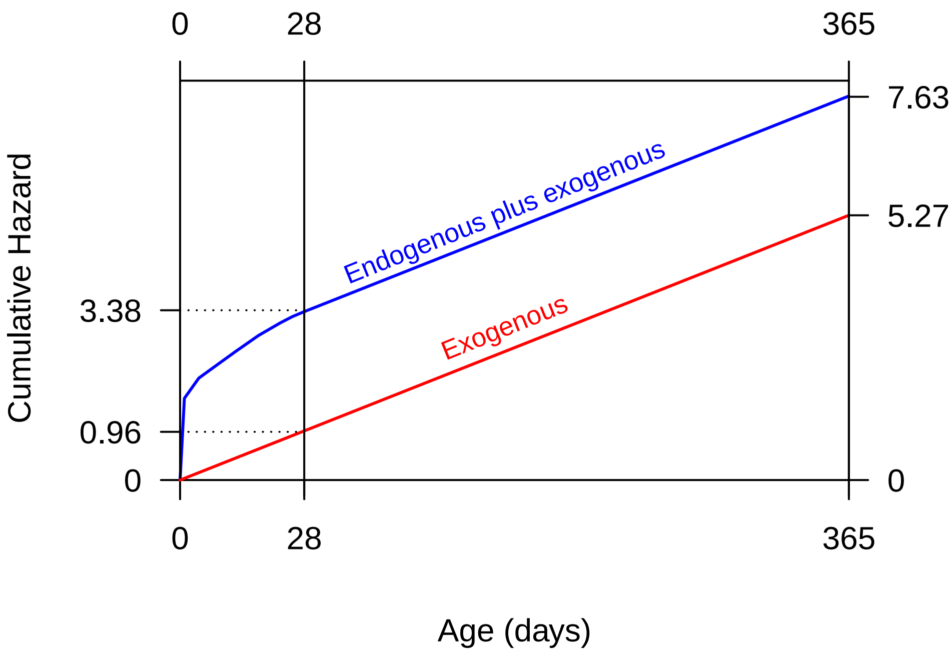


Figure 7: Separation of exogenous and endogenous infant mortality (cumulative hazards, per cent).

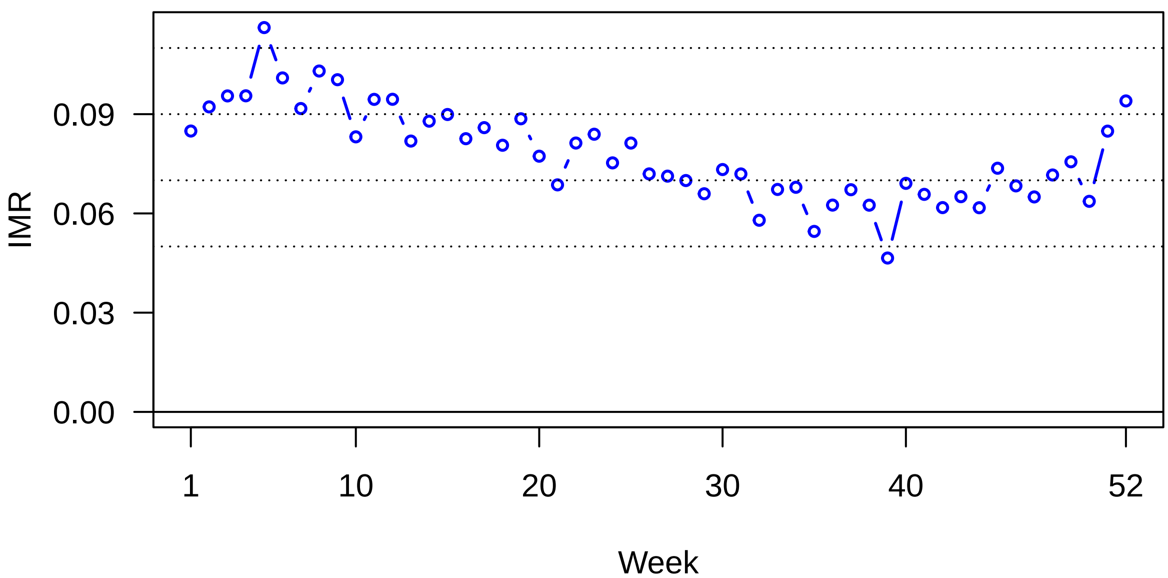


Figure 8: Crude infant mortality by week of year, Umeå/Skellefteå 1895–1950.

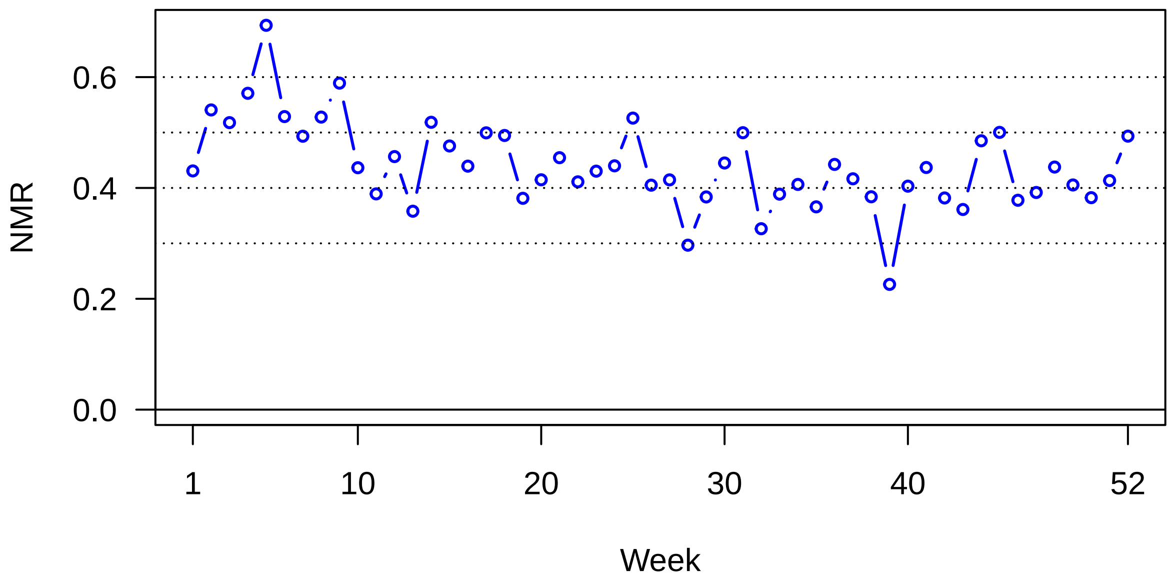


Figure 9: Crude neonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

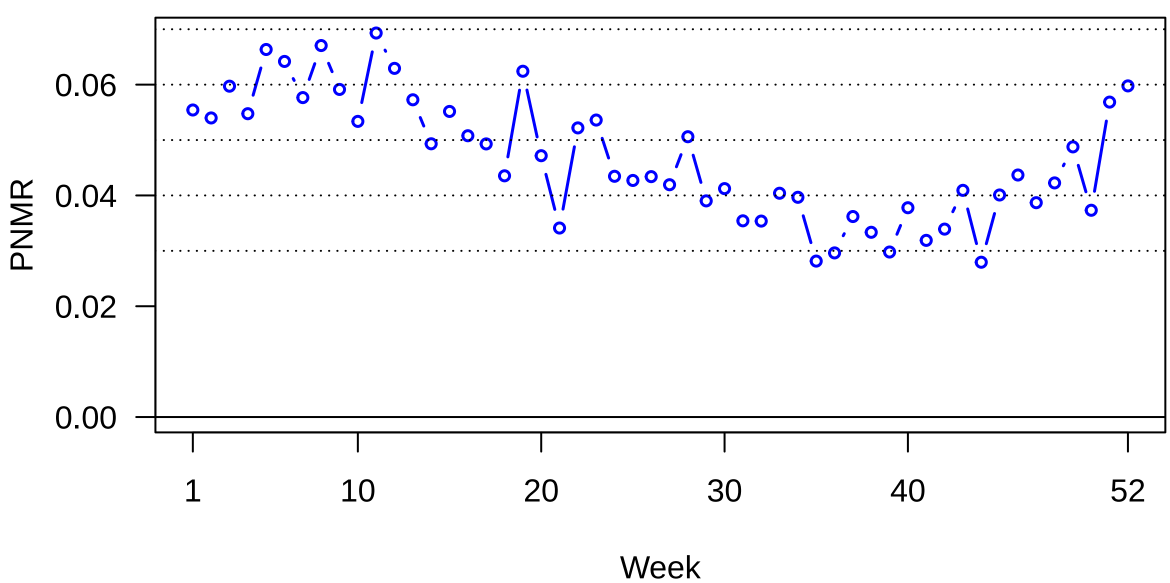


Figure 10: Crude postneonatal mortality by week of year, Umeå/Skellefteå 1895–1950.

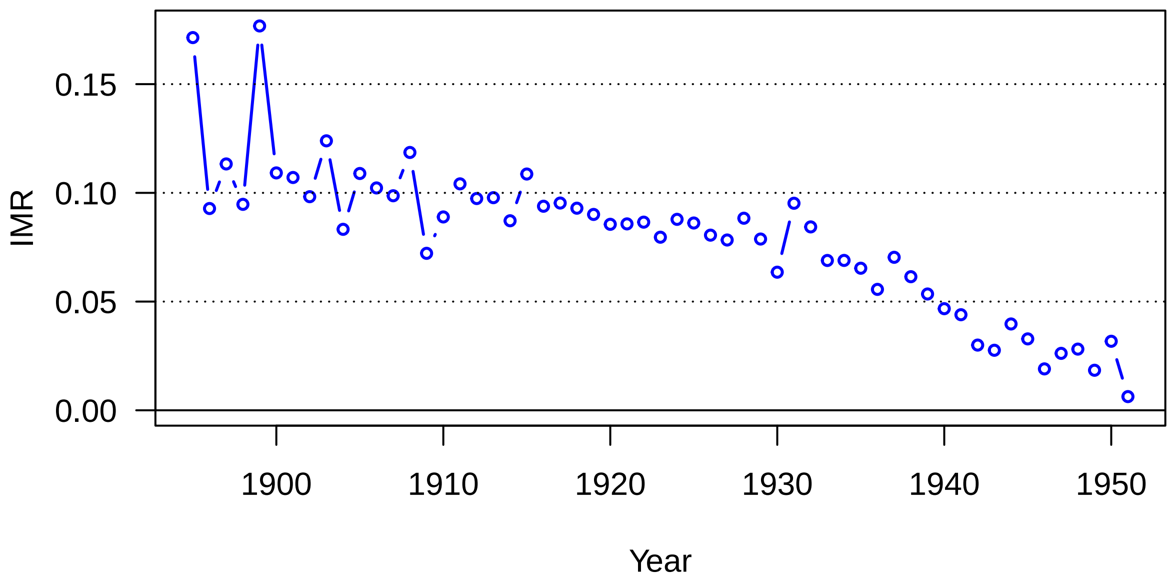


Figure 11: Crude IMR by year, Umeå-Skellefteå 1895–1950.

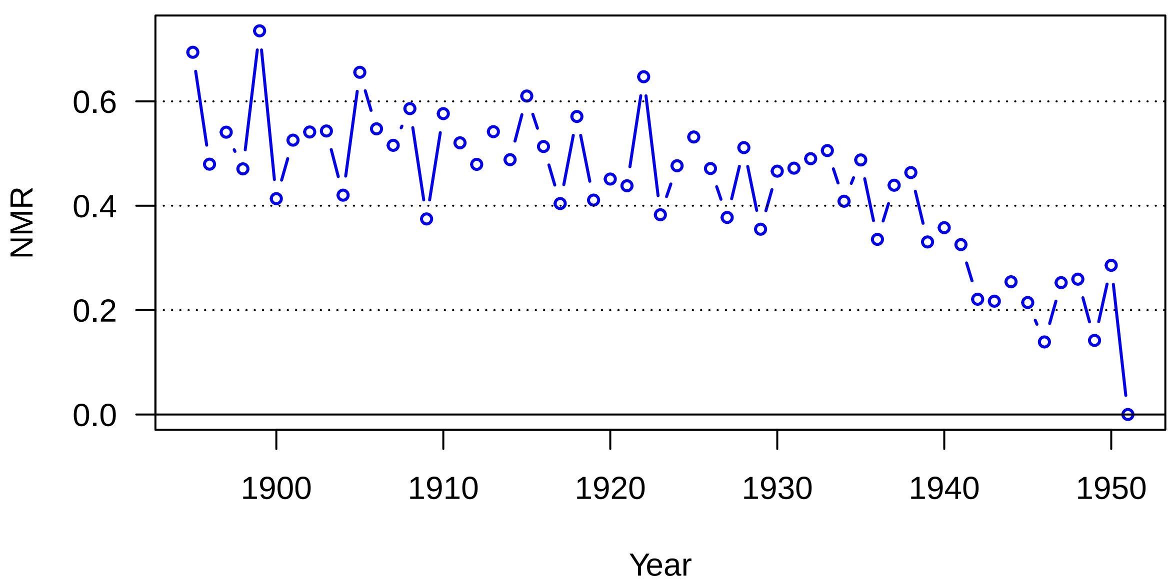


Figure 12: Crude NMR by year, Umeå-Skellefteå 1895–1950.

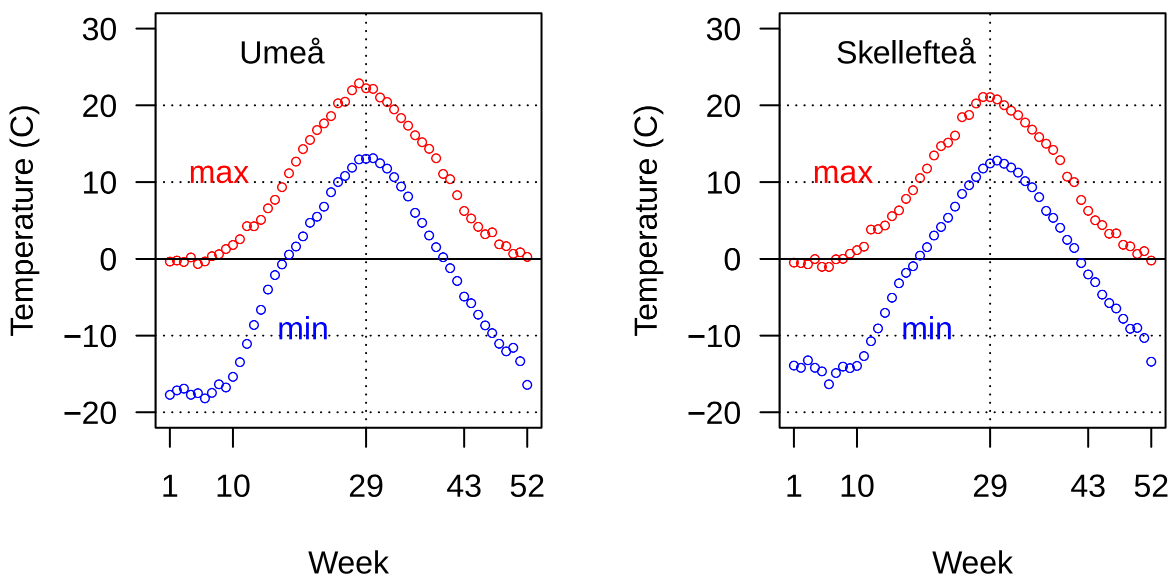


Figure 13: Weekly max and min temperature averages, 1895–1950.