

Fitting a TOPALS mortality model with age-grouped data, by Penalized Iteratively Weighted Least Squares (PIRLS)

Carl Schmertmann

11 Aug 2019

Main Idea

The objective is to fit a TOPALS model mortality schedule for A single years of age $x = 0, 1, \dots, (A - 1)$ from closed age-group data on exposure $N_1 \dots N_G$ and deaths $D_1 \dots D_G$, where the $G \leq A$ and the age groups are defined by G upper and lower bounds $[L_g, U_g)$ for $g = 1 \dots G$.

Sample Data

Observed data consists of deaths D_g and exposure N_g for closed, non-overlapping age groups $g = 1 \dots G$ delimited by lower and upper bounds $[L_g, U_g)$. Denote X_g as the set of integer ages that belong to group g , and n_g as the number of ages in X_g .

A typical age grouping is $L = (0, 1, 5, 10, \dots, 80)$ and $U = (1, 5, 10, 15, \dots, 85)$. But if single-year death and exposure data is available then the “groups” could be simply integer ages, in which case $L = (0, 1, 2, \dots, (A - 1))$ and $U = (1, 2, 3, \dots, A)$.

Poisson Log Likelihood for Grouped Data

Regardless of the age grouping in the data, we assume that there is a latent schedule for single-year ages. Deaths at each single-year age have a Poisson distribution with expected value equal to the (possibly unobserved) exposure N_x times the mortality rate:

$$D_x \sim \text{Poisson}(N_x \mu_x) \quad x = 0, 1, 2, \dots, (A - 1)$$

Under the standard assumption that age-specific deaths are statistically independent, this implies that deaths are also Poisson distributed within each age group:

$$D_g \sim \text{Poisson} \left(\sum_{x \in X_g} N_x \mu_x \right) \quad g = 1, 2, \dots, G$$

Expressing this in terms of the total *observed* exposure in each age group,

$$D_g \sim \text{Poisson} \left(N_g \sum_{x \in X_g} \left[\frac{N_x}{N_g} \right] \mu_x \right)$$

or

$$D_g \sim \text{Poisson} (N_g M_g)$$

where M_g represents the exposure-weighted average mortality rate across ages in group g .

In the absence of single-year exposure data, we approximate with $M_g = [\frac{1}{n_g}] \sum_{x \in X_g} \mu_x$.¹

The entire vector of G averaged mortality rates by age group is therefore

$$\mathbf{M} = \begin{pmatrix} n_1^{-1} \dots n_1^{-1} & 0 \dots 0 & \dots & 0 \dots 0 \\ 0 \dots 0 & n_2^{-1} \dots n_2^{-1} & \dots & 0 \dots 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 \dots 0 & 0 \dots 0 & \dots & n_G^{-1} \dots n_G^{-1} \end{pmatrix} \begin{pmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_{A-1} \end{pmatrix} = \mathbf{W} \boldsymbol{\mu}$$

where \mathbf{W} is a $G \times A$ matrix of weights with row sums all equal to one.²

The sample log likelihood for a single-year mortality schedule $\boldsymbol{\mu}$ is thus

$$\begin{aligned} \ln L(\boldsymbol{\mu}) &= c + \sum_{g=1}^G (D_g \ln M_g - N_g M_g) \\ &= c + \sum_{g=1}^G (D_g \ln M_g - \hat{D}_g) \end{aligned}$$

where $\hat{D}_g = N_g M_g$ represents the expected number of deaths in group g .

¹An alternative approach might interpolate single-year exposure (N_0, N_1, \dots, N_{A-1}) from the available N_g to construct more nuanced weights for intra-group averaging of mortality rates. In most populations this would have only minor effects on estimates.

²When "groups" correspond to single years, $\mathbf{W} = \mathbf{I}_A$ and $\mathbf{M} = \boldsymbol{\mu}$.

TOPALS Model Mortality Schedule

The TOPALS model for single-year mortality rates is

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}^* + \mathbf{B}\boldsymbol{\alpha}$$

where $\boldsymbol{\lambda}$ is an $A \times 1$ vector of age-specific log mortality rates for ages $0 \dots (A-1)$, $\boldsymbol{\lambda}^*$ is a vector with fixed constants representing a standard schedule for those ages, \mathbf{B} is a $A \times K$ matrix of linear B-spline constants³, and $\boldsymbol{\alpha}$ is a K -dimensional vector representing deviations from the standard log mortality schedule at specified ages.

In this model the log mortality rate at age $x \in \{0 \dots (A-1)\}$ is

$$\ln \mu_x = \lambda_x = \lambda_x^* + \mathbf{b}_x' \boldsymbol{\alpha}$$

and the mortality rate is

$$\mu_x = \exp(\lambda_x^* + \mathbf{b}_x' \boldsymbol{\alpha})$$

where \mathbf{b}_x' is the $1 \times K$ row of \mathbf{B} that corresponds to age x .

TOPALS derivatives

Derivatives of age-specific rates with respect to TOPALS parameters $\boldsymbol{\alpha}$ are $K \times 1$ vectors:

$$\frac{\partial \ln \mu_x}{\partial \boldsymbol{\alpha}} = \mathbf{b}_x \quad , \quad \frac{\partial \mu_x}{\partial \boldsymbol{\alpha}} = \mu_x \mathbf{b}_x$$

For the entire $G \times 1$ vector of group mortality rates, this implies

$$\begin{aligned} \frac{\partial \mathbf{M}'}{\partial \boldsymbol{\alpha}} &= \frac{\partial}{\partial \boldsymbol{\alpha}} (\boldsymbol{\mu}' \mathbf{W}') \\ &= \left(\frac{\partial \ln \mu_0}{\partial \boldsymbol{\alpha}} \quad \dots \quad \frac{\partial \ln \mu_{A-1}}{\partial \boldsymbol{\alpha}} \right) \mathbf{W}' \\ &= \left(\mathbf{b}_0 \mu_0 \quad \dots \quad \mathbf{b}_{A-1} \mu_{A-1} \right) \mathbf{W}' \\ &= \mathbf{B}' \text{diag}(\boldsymbol{\mu}) \mathbf{W}' \end{aligned}$$

³Typically $K = 7$, with spline knots fixed at ages $t = 0, 1, 10, 20, 40, 70, 99$. Values in the k th column of \mathbf{B} are

$$B_{xk} = \begin{cases} \frac{x-t_{k-1}}{t_k-t_{k-1}} & \text{for } x \in [t_{k-1}, t_k] \\ \frac{t_{k+1}-x}{t_{k+1}-t_k} & \text{for } x \in [t_k, t_{k+1}] \\ 0 & \text{otherwise} \end{cases}$$

Abbreviate this as

$$\frac{\partial \mathbf{M}'}{\partial \boldsymbol{\alpha}} = \mathbf{X}'$$

remembering that the $G \times K$ matrix $\mathbf{X} = \mathbf{W} \text{diag}(\boldsymbol{\mu}) \mathbf{B}$ varies with parameters $\boldsymbol{\alpha}$ via the μ terms in the central diagonal matrix.

The $G \times 1$ vector of expected deaths in groups $1 \dots G$ is

$$\hat{\mathbf{D}} = \begin{pmatrix} N_1 M_1 \\ \vdots \\ N_G M_G \end{pmatrix} = \text{diag}(\mathbf{N}) \mathbf{M}$$

and the derivative of its transpose with respect to TOPALS parameters $\boldsymbol{\alpha}$ is a $G \times K$ matrix:

$$\frac{\partial \hat{\mathbf{D}}'}{\partial \boldsymbol{\alpha}} = \frac{\partial \mathbf{M}'}{\partial \boldsymbol{\alpha}} \text{diag}(\mathbf{N}) = \mathbf{X}' \text{diag}(\mathbf{N}) \quad (1)$$

Penalized Log Likelihood for TOPALS parameters

With the TOPALS parameterization the log likelihood for a sample $\{D_g, N_g\}$ is

$$\begin{aligned} \ln L(\boldsymbol{\alpha}) &= c + \sum_{g=1}^G \left(D_g \ln M_g(\boldsymbol{\alpha}) - \hat{D}_g(\boldsymbol{\alpha}) \right) \\ &= c + [\ln M_1 \cdots \ln M_G] \mathbf{D}_{obs} - \hat{\mathbf{D}}' \mathbf{1} \end{aligned}$$

To stabilize estimates in small populations with few deaths, we add a small penalty term to the log likelihood:

$$\begin{aligned} Q(\boldsymbol{\alpha}) &= \ln L(\boldsymbol{\alpha}) - \text{penalty}(\boldsymbol{\alpha}) \\ &= c + [\ln M_1 \cdots \ln M_G] \mathbf{D} - \hat{\mathbf{D}}' \mathbf{1} - \frac{1}{2} \boldsymbol{\alpha}' \mathbf{P} \boldsymbol{\alpha} \end{aligned} \quad (2)$$

where \mathbf{P} is a $K \times K$ matrix of constants selected so that the penalty term equals the sum of squared differences between consecutive α parameters – i.e., $\frac{1}{2} \boldsymbol{\alpha}' \mathbf{P} \boldsymbol{\alpha} = \sum_{k=2}^K (\alpha_k - \alpha_{k-1})^2$. Adding the penalty gives priority to sets of TOPALS parameters $(\alpha_1 \dots \alpha_K)$ that are similar to one another, and thus to log mortality schedules that look more like up-and-down vertical shifts of the standard schedule. For all but the smallest populations the penalty term has virtually no effect on parameter estimates. For very small populations with zero deaths in some age groups, the addition of the penalty stabilizes estimated schedules by borrowing strength across groups.

Maximizing via Iteratively Reweighted Regression

We want to select α to maximize Q . This requires setting a vector of derivatives equal to zero: $\frac{\partial Q}{\partial \alpha} = 0 \in \mathbb{R}^K$.

Differentiating Eq. (2) with respect to the TOPALS parameters and substituting some of the results above produces

$$\begin{aligned}
\frac{\partial Q}{\partial \alpha} &= \left[\frac{1}{M_1} \frac{\partial M_1}{\partial \alpha} \dots \frac{1}{M_G} \frac{\partial M_G}{\partial \alpha} \right] D_{obs} - X' \text{diag}(N) \mathbf{1} - P\alpha \\
&= \frac{\partial M'}{\partial \alpha} \text{diag}\left(\frac{1}{M}\right) D_{obs} - X' N - P\alpha \\
&= X' \text{diag}\left(\frac{1}{M}\right) D_{obs} - X' N - P\alpha \\
&= X' \text{diag}\left(\frac{1}{M}\right) D_{obs} - X' \text{diag}\left(\frac{1}{M}\right) \hat{D} - P\alpha \\
&= X' \text{diag}\left(\frac{D_{obs} - \hat{D}}{M}\right) - P\alpha
\end{aligned}$$

For reasons that will become clear in a moment, it is convenient to define the abbreviation $A = \text{diag}\left(\frac{N_1}{M_1} \dots \frac{N_G}{M_G}\right)$. Using this new abbreviation,

$$\frac{\partial Q}{\partial \alpha} = X' A \text{diag}\left(\frac{D_{obs} - \hat{D}}{N}\right) - P\alpha = \mathbf{0} \in \mathbb{R}^K \quad (3)$$

Rearranging Eq. (3) and placing hats over the α -dependent terms yields a system of K equations in K variables α

$$\hat{X}' \hat{A} \text{diag}\left(\frac{\hat{D}}{N}\right) + P\hat{\alpha} = \hat{X}' \hat{A} \text{diag}\left(\frac{D_{obs}}{N}\right) \quad (4)$$

The iteratively reweighted least squares approach to solving the system of equations in (4) begins by approximating how predicted values \hat{D} would change if the vector of TOPALS parameters changed from $\hat{\alpha}$ to some other value α^* , while (temporarily) keeping matrices \hat{X} and \hat{A} constant:

$$\hat{X}' \hat{A} \text{diag}\left(\frac{1}{N}\right) \left[\hat{D} + \frac{\partial \hat{D}}{\partial \alpha'} (\alpha^* - \hat{\alpha}) \right] + P\alpha^* = \hat{X}' \hat{A} \text{diag}\left(\frac{D_{obs}}{N}\right) \quad (5)$$

which from Eq. (1) expands to

$$\hat{X}' \hat{A} \text{diag}\left(\frac{1}{N}\right) \left[\hat{D} + \text{diag}(N) \hat{X} (\alpha^* - \hat{\alpha}) \right] + P\hat{\alpha} = \hat{X}' \hat{A} \text{diag}\left(\frac{1}{N}\right) D_{obs} \quad (6)$$

which can be rearranged as

$$\hat{\mathbf{X}}' \hat{\mathbf{A}} \hat{\mathbf{X}} (\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}) + \mathbf{P} \boldsymbol{\alpha}^* = \hat{\mathbf{X}}' \hat{\mathbf{A}} \text{diag}\left(\frac{\mathbf{D}_{obs} - \hat{\mathbf{D}}}{\mathbf{N}}\right)$$

and further simplified as

$$\left[\hat{\mathbf{X}}' \hat{\mathbf{A}} \hat{\mathbf{X}} + \mathbf{P} \right] \boldsymbol{\alpha}^* = \hat{\mathbf{X}}' \hat{\mathbf{A}} \left[\text{diag}\left(\frac{\mathbf{D}_{obs} - \hat{\mathbf{D}}}{\mathbf{N}}\right) + \hat{\mathbf{X}} \hat{\boldsymbol{\alpha}} \right]$$

and finally simplified to

$$\left[\hat{\mathbf{X}}' \hat{\mathbf{A}} \hat{\mathbf{X}} + \mathbf{P} \right] \boldsymbol{\alpha}^* = \hat{\mathbf{X}}' \hat{\mathbf{A}} \hat{\mathbf{y}} \quad (7)$$

where $\hat{\mathbf{y}} = \text{diag}\left(\frac{\mathbf{D}_{obs} - \hat{\mathbf{D}}}{\mathbf{N}}\right) + \hat{\mathbf{X}} \hat{\boldsymbol{\alpha}}$.

Eq. (7) is analogous to a (penalized) weighted least squares regression of $\hat{\mathbf{X}}$ on $\hat{\mathbf{y}}$ with weights $\hat{\mathbf{A}}^{-1}$. Starting from an arbitrary vector $\hat{\boldsymbol{\alpha}}$, one can solve for the vector that maximizes penalized likelihood $Q(\boldsymbol{\alpha})$ by (1) calculating the values of $(\hat{\mathbf{X}}, \hat{\mathbf{A}}, \hat{\mathbf{D}})$ at the current parameter value, (2) solving eq. (7) for $\boldsymbol{\alpha}^*$, and repeating until convergence. This is the algorithm used below.

Example: A TOPALS fit from five-year age groups for Italian Females in 1980

Age groups for this exercise are 0, 1–4, 5–9, ..., 80–85.

```
library(tidyverse, quietly = TRUE)
library(splines, quietly = TRUE)
```

```
## Italian female 1980 HMD data for age groups
```

```
L = c(0,1,seq(5,80,5))
U = c(tail(L,-1),85)
```

```
N = c(312106.85, 1423566.3, 2105814.63, 2249555.41, 2230885.74, 1983157.8,
1874479.58, 1989351.99, 1772133.73, 1896866.51, 1836597.36, 1834496.64,
1811178.38, 1192763.85, 1498384.03, 1223810.9, 863725.92, 537720.77)
```

```
D = c(3889, 716, 587, 589, 791, 816, 832, 1257, 1651, 2721, 4310,
```

```

6636, 10536, 11043, 23312, 34945, 44537, 50392)

## save the real data in "big" versions -- useful for sampling later
bigD = D
bigN = N

## standard schedule is (arbitrarily) the 1959 period schedule
## for Canadian females at ages 0,1,...,99
## these are the log mortality rates
std = c(-3.8933, -5.7776, -6.8474, -7.3298, -7.4519, -7.4408, -7.4807,
-7.5845, -7.7219, -7.8628, -7.9771, -8.041, -8.0568, -8.0329,
-7.9779, -7.9005, -7.8088, -7.7101, -7.6113, -7.5195, -7.4415,
-7.3823, -7.3393, -7.308, -7.2837, -7.2619, -7.238, -7.2082,
-7.1712, -7.1264, -7.0735, -7.0118, -6.9414, -6.8648, -6.7849,
-6.7047, -6.6272, -6.5544, -6.4845, -6.4147, -6.3423, -6.2645,
-6.1791, -6.0872, -5.9904, -5.8903, -5.7887, -5.6869, -5.586,
-5.4866, -5.3895, -5.2953, -5.2049, -5.1186, -5.0347, -4.9513,
-4.8664, -4.778, -4.6847, -4.5877, -4.4887, -4.3895, -4.2918,
-4.1969, -4.1041, -4.0122, -3.9199, -3.8261, -3.7296, -3.6303,
-3.5278, -3.4221, -3.3129, -3.2004, -3.0861, -2.9716, -2.8589,
-2.7497, -2.6458, -2.5482, -2.4556, -2.3659, -2.2771, -2.187,
-2.0942, -1.9991, -1.9028, -1.8062, -1.7105, -1.6164, -1.5242,
-1.434, -1.3458, -1.2596, -1.1758, -1.0958, -1.0212, -0.9535,
-0.8944, -0.8454)

## single-year log mortality rates from HMD
## these are the targets for TOPALS estimation
ITA_HMD_logmx =
  c(-4.3852, -7.1185, -7.6009, -7.7517, -8.1117, -8.1456, -8.1456,
-8.1456, -8.294, -8.2171, -8.4684, -8.294, -8.3349, -8.1456,
-8.0789, -7.9866, -7.9866, -8.0164, -7.902, -7.824, -7.7753,
-7.7753, -7.7753, -7.7753, -7.8753, -7.7063, -7.7063, -7.6628,
-7.6417, -7.8753, -7.4876, -7.4354, -7.2644, -7.3233, -7.3385,
-7.2644, -7.0021, -6.959, -6.959, -6.7855, -6.8216, -6.5713,
-6.5225, -6.4956, -6.3539, -6.2712, -6.2196, -6.0035, -5.9835,
-5.8569, -5.7992, -5.7169, -5.6694, -5.5315, -5.433, -5.3247,

```

```

-5.2514, -5.1814, -5.0625, -4.9533, -4.8783, -4.7915, -4.6767,
-4.5923, -4.4945, -4.3836, -4.2992, -4.1825, -4.0513, -3.9409,
-3.8135, -3.6913, -3.5332, -3.4455, -3.2966, -3.2069, -3.0614,
-2.9677, -2.8466, -2.7201, -2.5974, -2.4617, -2.3462, -2.2249,
-2.1253, -1.9713, -1.8905, -1.7861, -1.6842, -1.5945, -1.4583,
-1.3792, -1.297, -1.2087, -1.1393, -1.0245, -0.9444, -0.8681,
-0.7958, -0.7276)

## single-year ages for TOPALS schedule
age = 0:99
knot_positions = c(0,1,10,20,40,70)

## B is an A x 7 matrix. Each column is a linear B-spline basis function
B = bs( age, knots=knot_positions, degree=1 )
K = ncol(B)

## penalty
D1 = diff( diag(K), diff=1)
P = 2 * crossprod(D1)

## weighting matrix W
A = length(age) # number of single-year ages
G = length(L)   # number of age groups

W = matrix(0, nrow=G, ncol=A, dimnames=list(L,age))

W['0','0'] = 1
W['1',2:5] = 0.25
for (i in 3:G) W[i, 5*(i-2)+1:5] = 0.2

```

Define a function that takes a TOPALS $\hat{\alpha}$ vector as input and solves (7) to produce an updated value α^* .

```

next_alpha = function(alpha, D, N) {
  mu = as.vector( exp( std + B %*% alpha))

```



```

M = as.vector( W %*% mu)

Dhat = N * M

X = W %*% diag(mu) %*% B
A = diag(N/M)

y = (D-Dhat)/N + X %*% alpha

updated_alpha = solve( t(X) %*% A %*% X + P, t(X) %*% A %*% y)
return(updated_alpha)
}

## (preliminary & rudimentary) function to fit and plot a TOPALS schedule

fit_TOPALS = function(D,N) {

  alpha = matrix(NA,K,6, dimnames=list(c(knot_positions,99), paste0('iter',0:5)))

  alpha[,1] = rep(0,K)

  for (i in 2:ncol(alpha)) {
    alpha[,i] = next_alpha( alpha[,i-1], D, N)
  }

  round( alpha,4)

  ## plot the data, true rates, and PIRLS fit

  df_grouped = data.frame(
    L = L,
    U = U,
    N = N,
    D = D

```

```

    ) %>%
    mutate(logmx_obs = log(D/N))

df_single = data.frame(
  age=age+.50,
  std = std,
  logmx_true = ITA_HMD_logmx,
  logmx_fit = std + B %*% alpha[,6]
)

this_plot =
  ggplot(data = df_single, aes(x=age,y=logmx_true)) +
  geom_line(aes(x=age,y=std), color='black', lwd=0.5) +
  geom_line(aes(x=age,y=logmx_fit), color='red', lwd=3, alpha=.40) +
  geom_segment(data=df_grouped,aes(x=L,xend=U,
                                   y=logmx_obs,
                                   yend=logmx_obs),
              color='red',lwd=1, alpha=.90) +
  geom_point(size=0.60) +
  labs(x='Age',y='Log Mortality Rate',
       title='Italy Females 1980',
       subtitle = paste(sum(D),'deaths to',round(sum(N)),'women')) +
  scale_x_continuous(breaks=c(0,1,seq(5,100,5)),minor_breaks = NULL) +
  theme_bw()

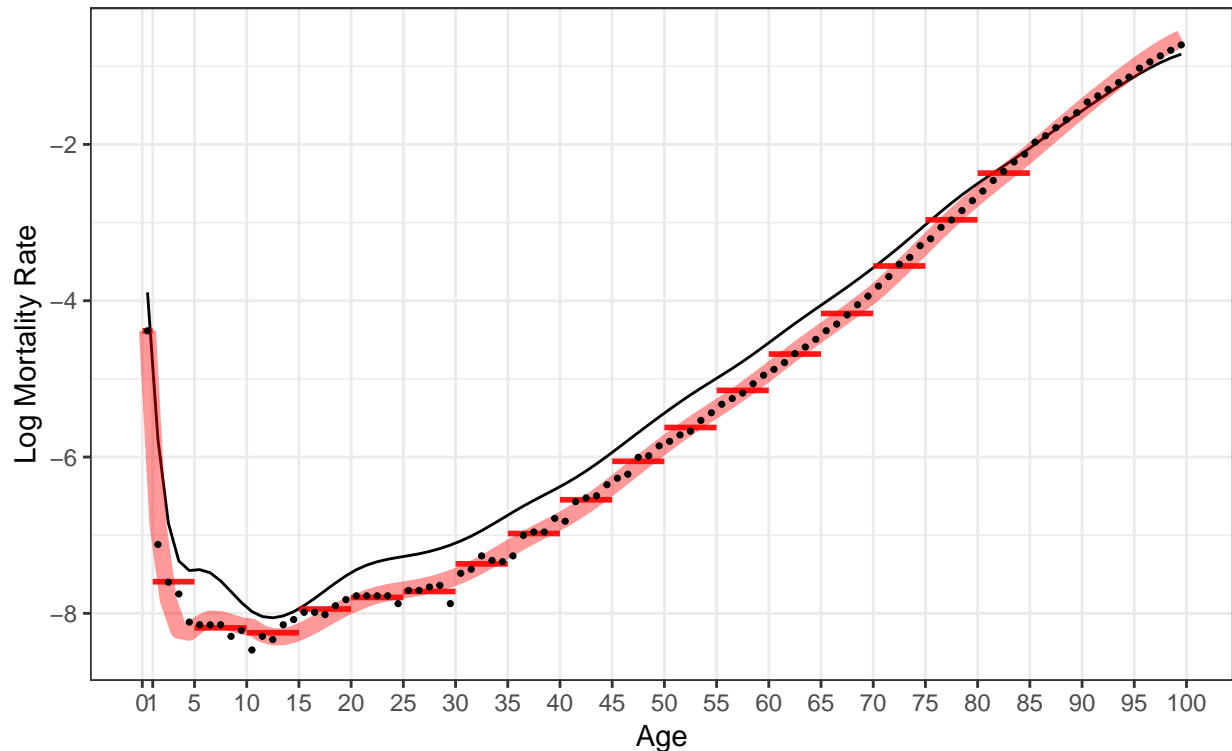
print(this_plot)
return(NULL) # temporary -- will want detailed output
}

fit_TOPALS(bigD,bigN)

```

Italy Females 1980

199560 deaths to 28636596 women



NULL

In this figure the dots are the true rates from the HMD, the horizontal bars are the $\log(N_g/D_g)$ values from the grouped data, the black curve is the standard, and the red curve is the TOPALS single-year fit from PIRLS.

Example: Very small populations with Italy 1980 rates

Repeat the exercise in the last section but with simulated deaths for a much smaller population subject to the same rates. Suppose there were only 10000 Italian women and simulate deaths:

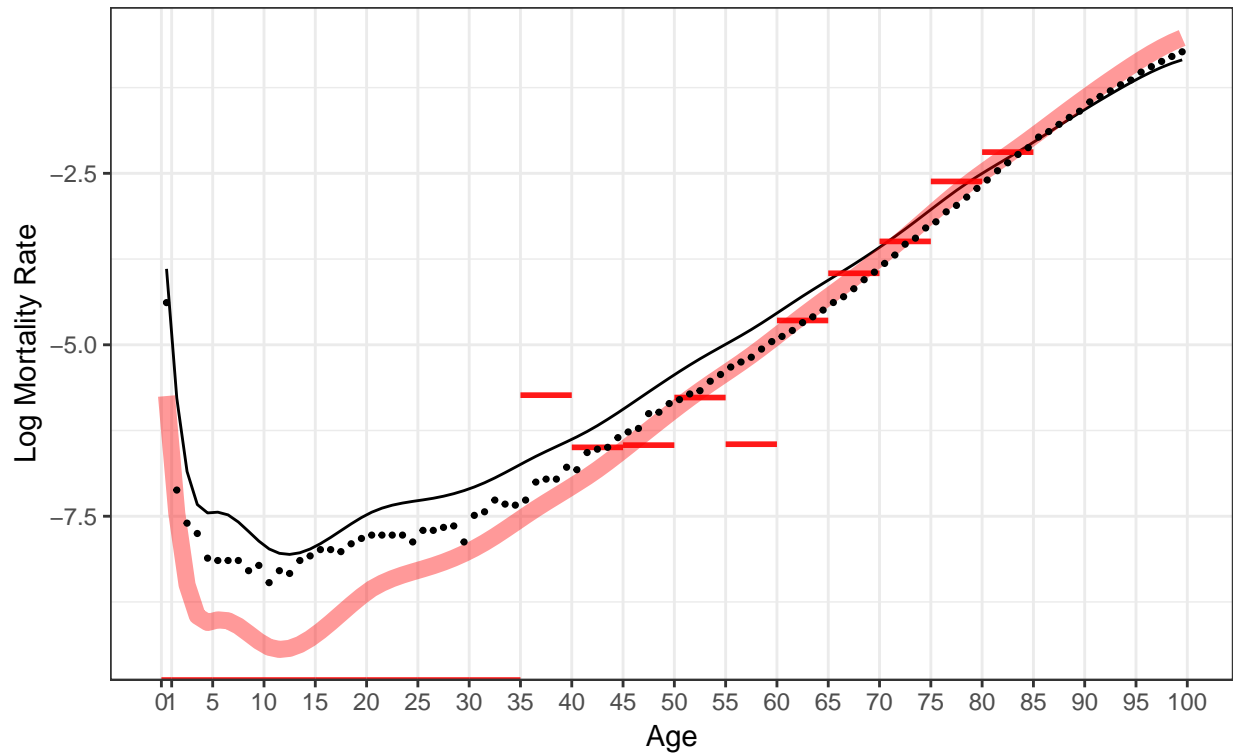
```
for (i in 1:5) {  
  ## simulate a small population  
  target_pop = 10000  
  N = bigN * target_pop/sum(bigN)  
  D = rpois(length(N), N* bigD/bigN)
```

```
fit_TOPALS(D,N)
```

```
}
```

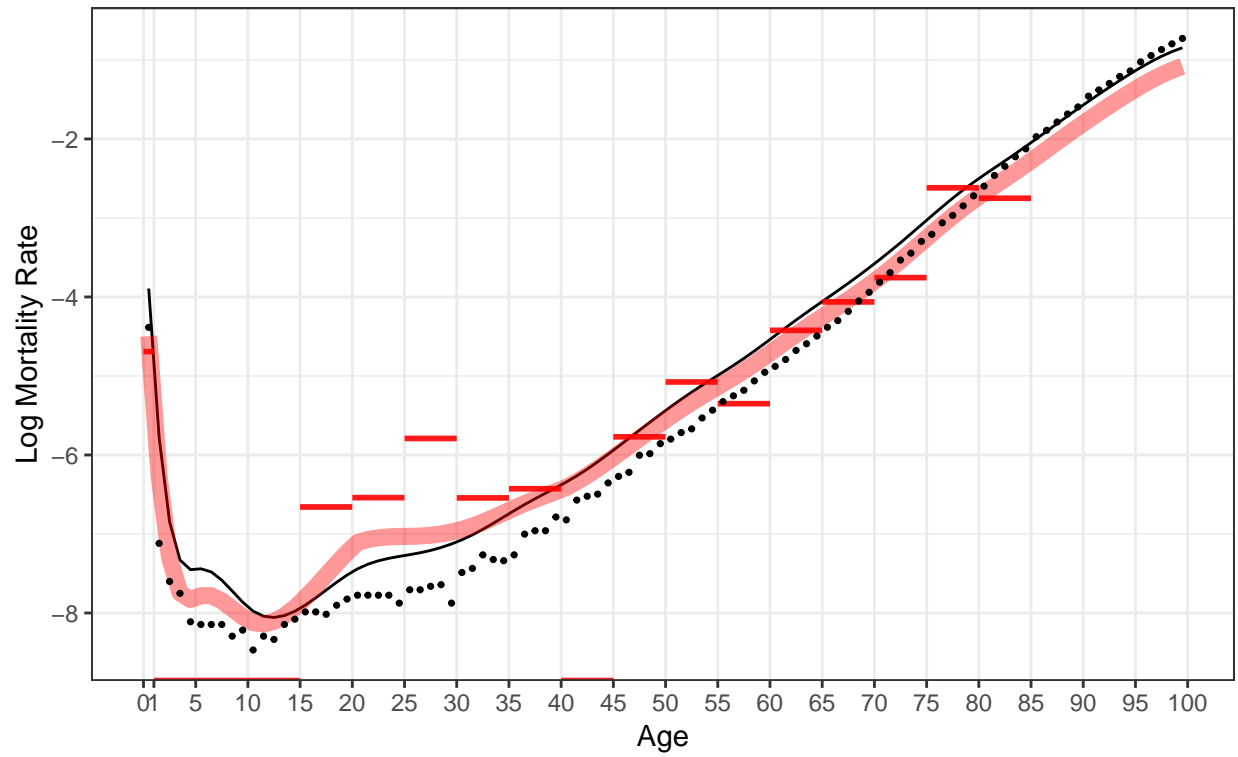
Italy Females 1980

77 deaths to 10000 women

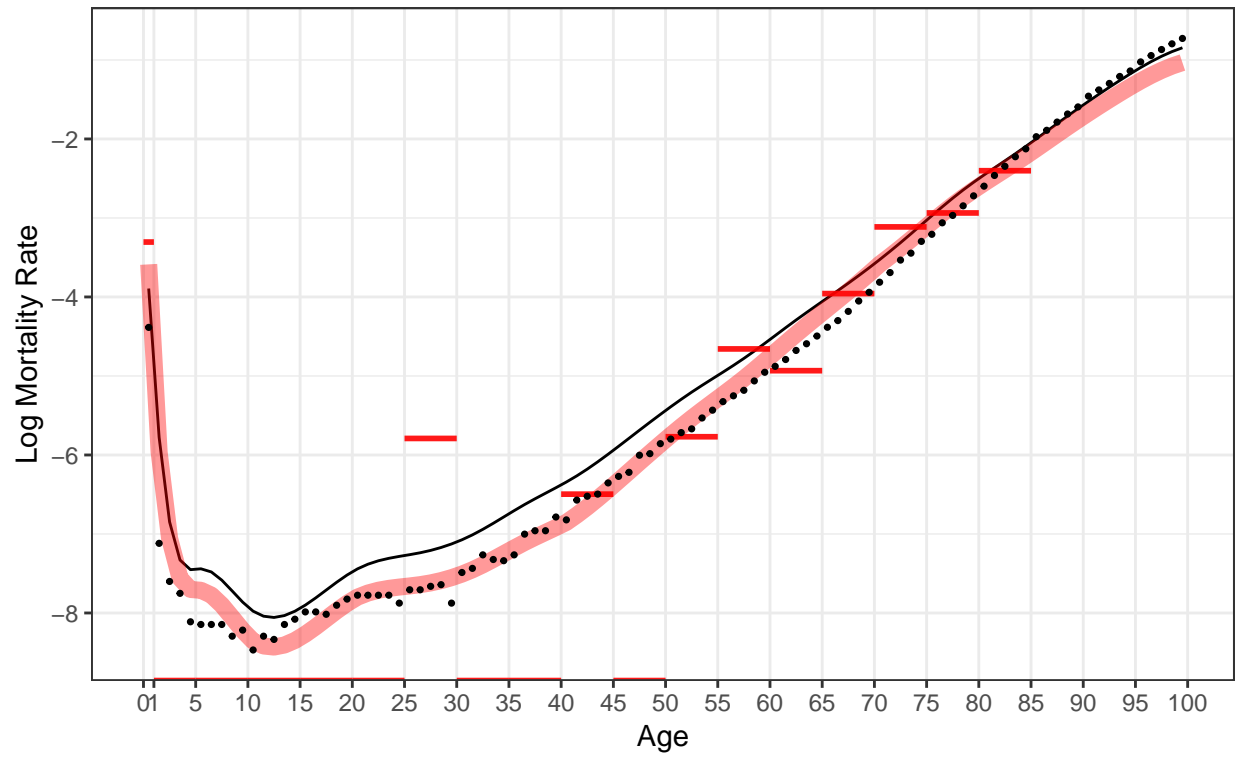


Italy Females 1980

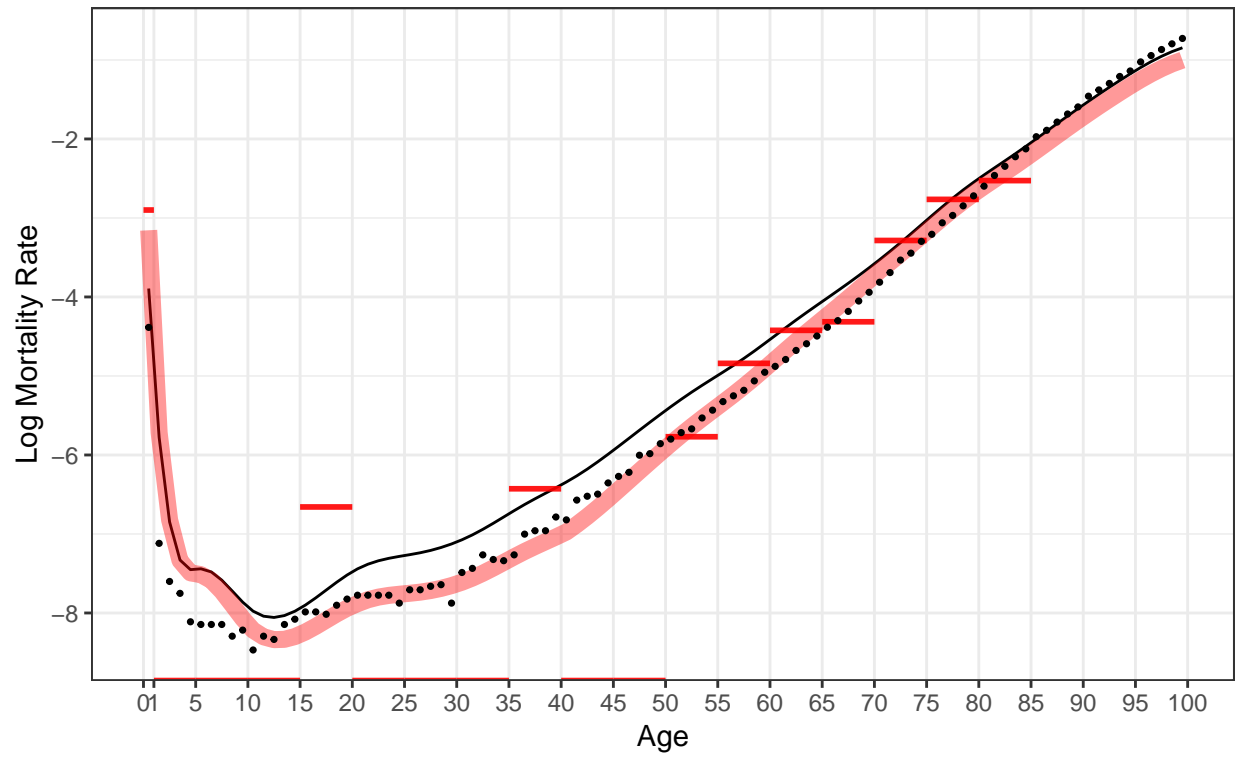
74 deaths to 10000 women



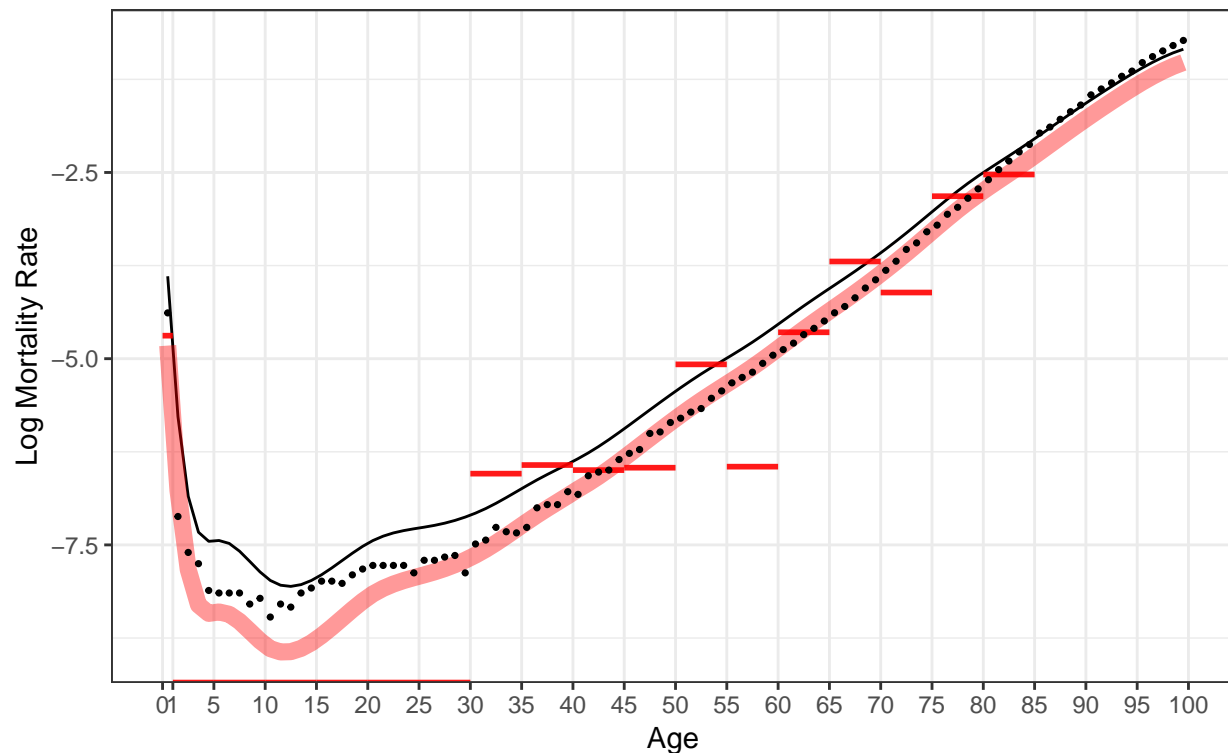
Italy Females 1980
80 deaths to 10000 women



Italy Females 1980
77 deaths to 10000 women



Italy Females 1980
67 deaths to 10000 women



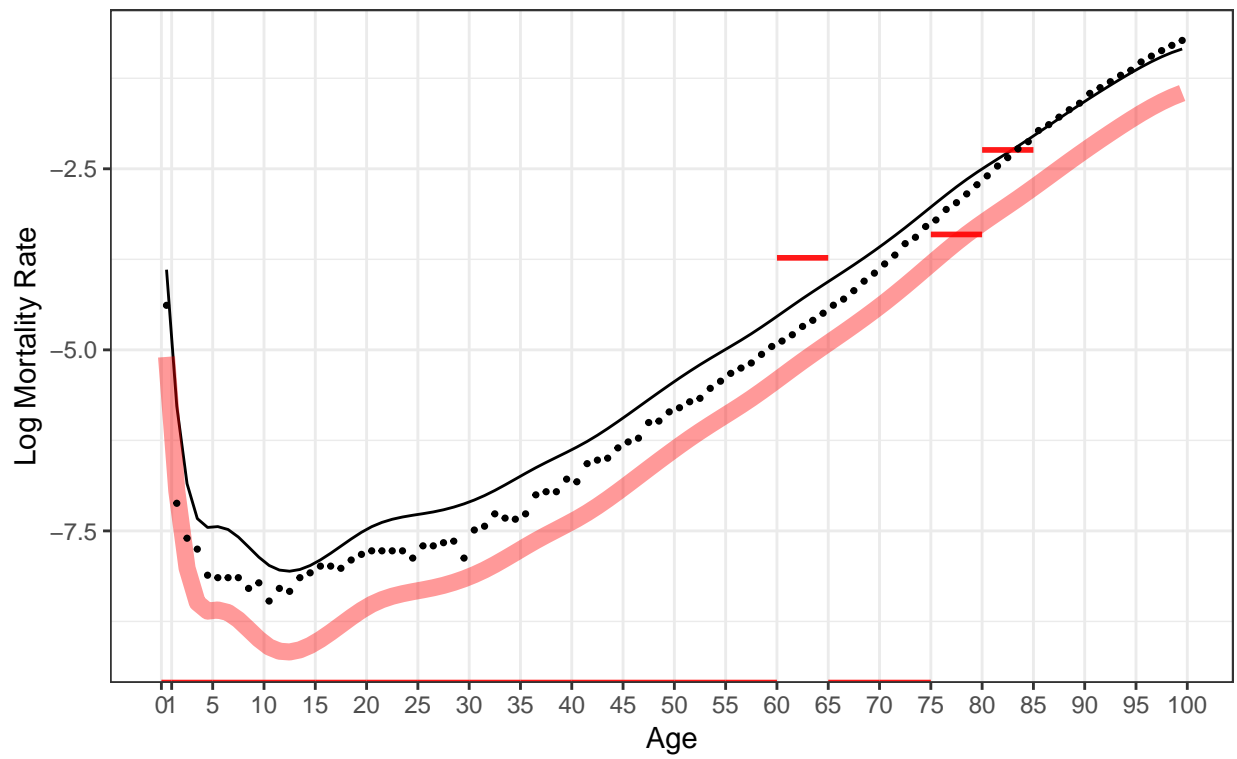
Example: Super-small populations with Italy 1980 rates

Repeat the exercise in the last section but with simulated deaths for a much smaller population subject to the same rates. Suppose there were only 1000 Italian women and simulate deaths:

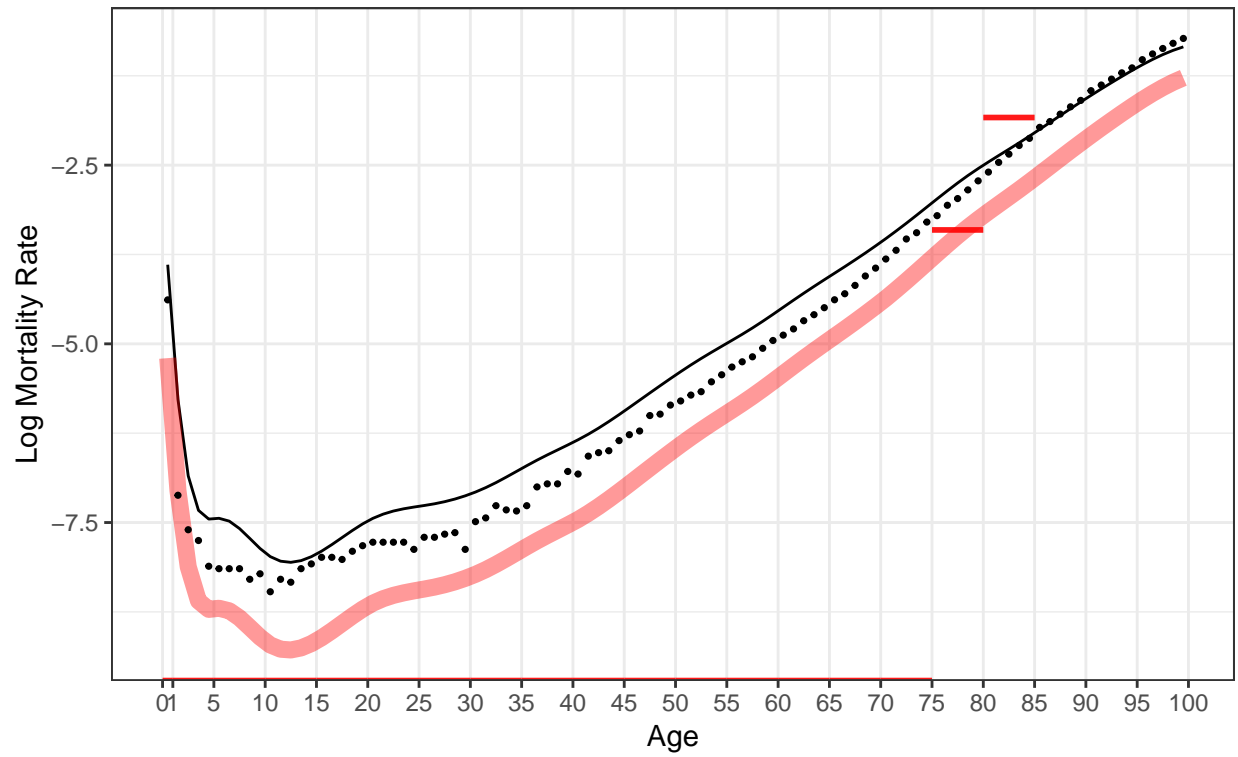
```
for (i in 1:5) {
  ## simulate a small population
  target_pop = 1000
  N = bigN * target_pop/sum(bigN)
  D = rpois(length(N), N* bigD/bigN)

  fit_TOPALS(D,N)
}
```

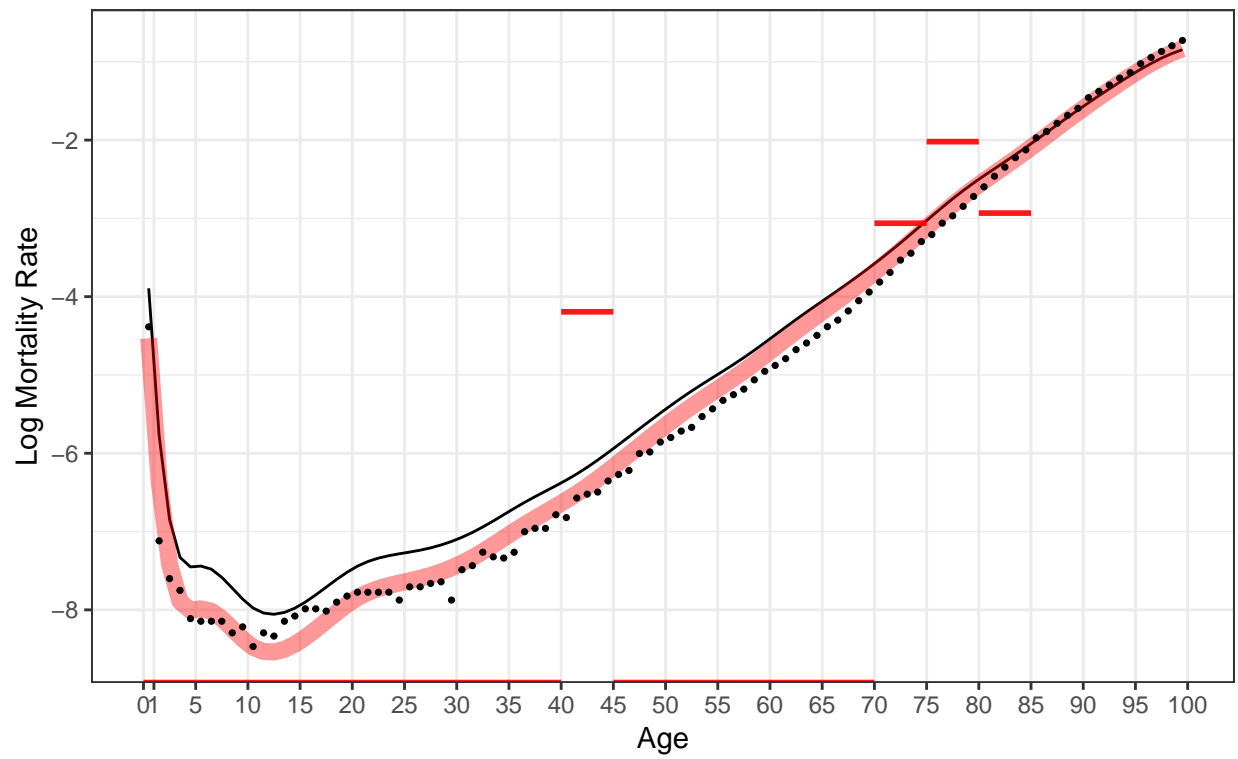

Italy Females 1980
4 deaths to 1000 women



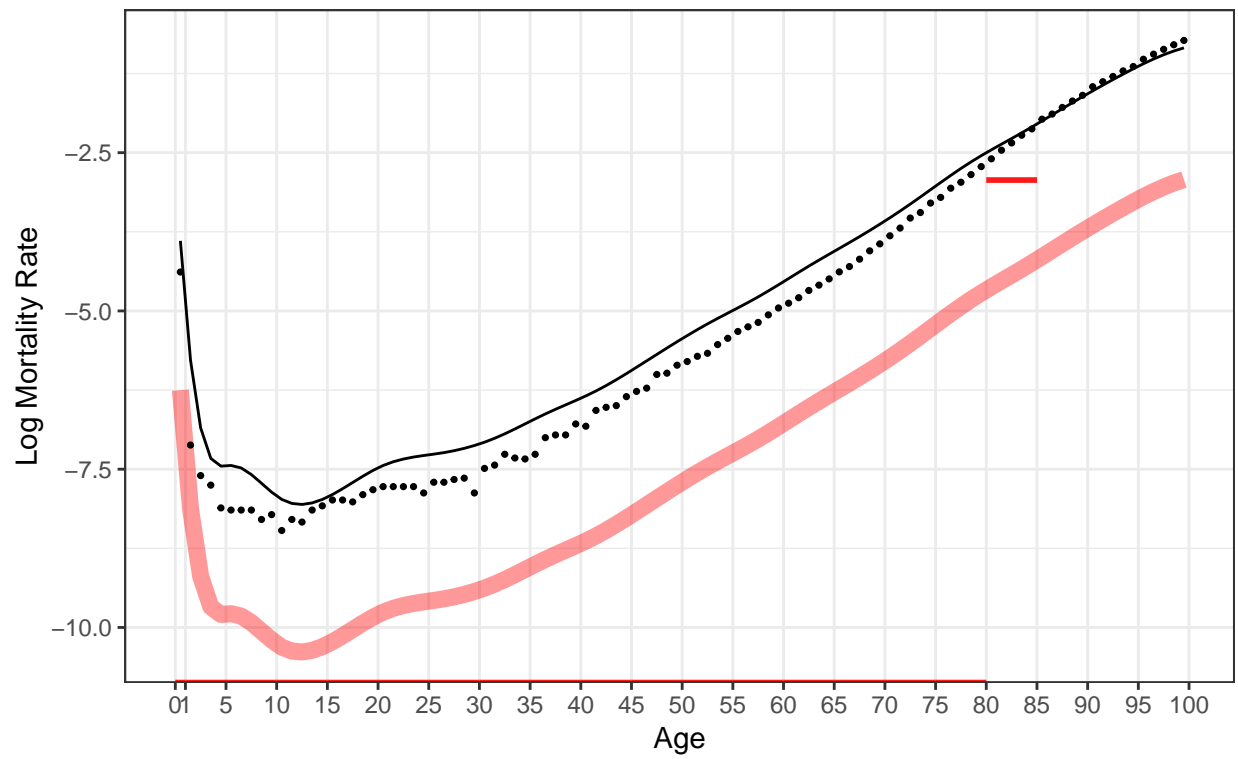
Italy Females 1980
4 deaths to 1000 women



Italy Females 1980
8 deaths to 1000 women



Italy Females 1980
1 deaths to 1000 women



Italy Females 1980
9 deaths to 1000 women

