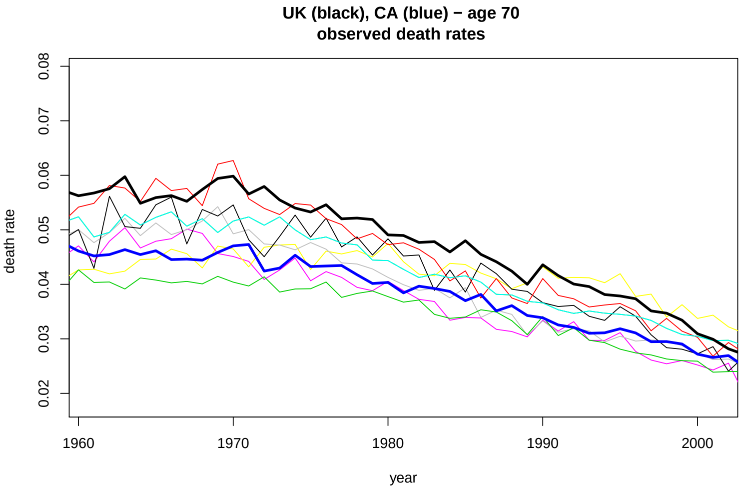
Mortality and Smoking Prevalence

Torsten Kleinow
joint work with Andrew J.G. Cairns
Heriot-Watt University, Edinburgh

AFIR - Mexico - 2012

Mortality



UK (black), CA (blue) − age 70
observed death rates
death rate
0.02
0.03
0.04
0.05
0.06
0.07
0.08
1960
1970
1980
1990
2000
year

Mortality Models for Multiple Populations

Consider k different populations (countries).
For each country i, time t (calendar year) and age x we observe

$D_i(x,t)$: Number of deaths,

$E_i(x,t)$: Exposure-to-risk

$m_i(x,t)$: $= D_i(x,t)/E_i(x,t)$, death rate

Core Hypothesis, Li & Lee (2005), Cairns et al. (2011): For all ages x and all i and j:

$$m_i(x,t)/m_j(x,t) \not\to \infty \text{ for } t \to \infty$$

Covariates

Covariates influencing individual life expectancy and disability-free life expectancy:

- life style (obesity, smoking, alcohol consumption, physical exercise, ...)
- socio-economic variables (income, wealth, Housing tenure, education, ...)
- genetic factors?

Covariates

Covariates influencing individual life expectancy and disability-free life expectancy:

- life style (obesity, smoking, alcohol consumption, physical exercise, ...)
- socio-economic variables (income, wealth, Housing tenure, education, ...)
- genetic factors?

Can these covariates be used to model country specific mortality rates?

Covariates

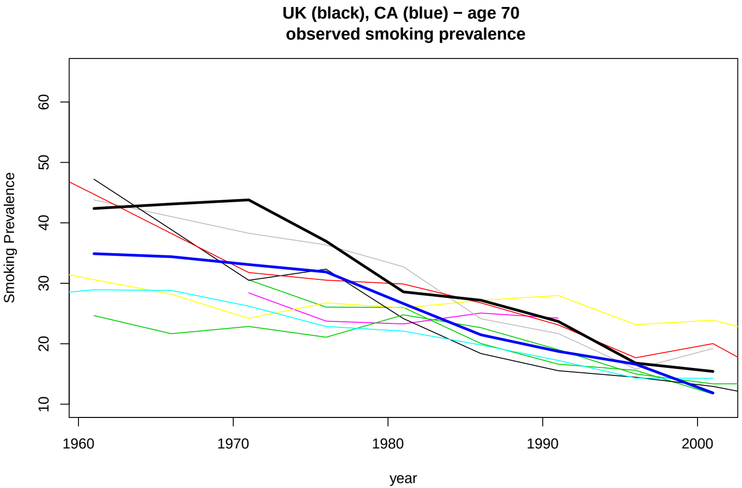
Covariates influencing individual life expectancy and disability-free life expectancy:

- life style (obesity, smoking, alcohol consumption, physical exercise, ...)
- socio-economic variables (income, wealth, Housing tenure, education, ...)
- genetic factors?

Can these covariates be used to model country specific mortality rates?

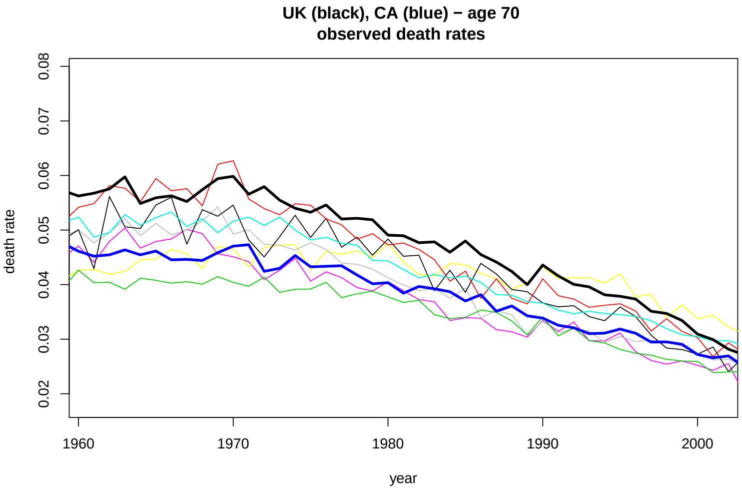
Example: Smoking prevalence

Smoking



UK (black), CA (blue) − age 70
observed smoking prevalence
Smoking Prevalence
10
20
30
40
50
60
1960
1970
1980
1990
2000
year

Mortality


UK (black), CA (blue) − age 70
observed death rates
death rate
0.02
0.03
0.04
0.05
0.06
0.07
0.08
1960
1970
1980
1990
2000
year

Available data

What we observe:

- death rates (www.mortality.org), "1$\times$1-table"
- smoking prevalence (International Smoking Statistics, P N Lee Statistics and Computing Ltd)
  there are different definitions (total cigarettes, manufactured cigarettes, any tobacco products), and different frequencies (age groups and year groups)
  based on surveys, different organisations focus on different age groups

Available data

What we observe:

- death rates (www.mortality.org), "1×1-table"
- smoking prevalence (International Smoking Statistics, P N Lee Statistics and Computing Ltd)
  there are different definitions (total cigarettes, manufactured cigarettes, any tobacco products), and different frequencies (age groups and year groups)
  based on surveys, different organisations focus on different age groups

What we do not observe:

- death rates for smokers and non-smokers, separately
- Cessation data
- "1×1-table", in general, prevalence data are only available for age groups

Smoking and Mortality

For each country i, time t and age x we define

- $D_i(x,t)$: Number of deaths,
  $D_i^N(x,t)$, $D_i^S(x,t)$ for non-smokers, smokers (not observed)
  $D_i(x,t) = D_i^N(x,t) + D_i^S(x,t)$
- $E_i(x,t)$: Exposure-to-risk
- $m_i(x,t)$: $= D_i(x,t)/E_i(x,t)$,
  $m_i^N(x,t)$, $m_i^S(x,t)$, death rates
- $s_i(x,t)$: Smoking prevalence, in $[0,1]$,
  the number of smokers is $s_i(x,t)E_i(x,t)$

We do not distinguish between life-long non-smokers and non-smokers who used to smoke.

Smoking and Mortality

$$
\begin{aligned}
D_i(x,t) &= D_i^N(x,t) + D_i^S(x,t) \\
&= m_i^N(x,t)\left[1 - s_i(x,t)\right] E_i(x,t) + m_i^S(x,t) s_i(x,t) E_i(x,t)
\end{aligned}
$$

where

$$
\begin{aligned}
m_i^N(x,t) &= \frac{D_i^N(x,t)}{\left[1 - s_i(x,t)\right] E_i(x,t)} \\
m_i^S(x,t) &= \frac{D_i^S(x,t)}{s_i(x,t) E_i(x,t)}
\end{aligned}
$$

We obtain

$$
m_i(x,t) = \frac{D_i(x,t)}{E_i(x,t)} = m_i^N(x,t) + \left[m_i^S(x,t) - m_i^N(x,t)\right] s_i(x,t)
$$

Smoking and Mortality

Modelling assumptions:

- Smoking prevalence has the same effect on mortality rates in all observed countries.
- Total mortality in country i is the weighted average of non-smokers' and smokers' mortality adjusted by a "country effect"

$$m_i(x,t) \;=\; \Big[\mathbf{m}^{\mathbf{N}}(\mathbf{x},\mathbf{t}) + \big[\mathbf{m}^{\mathbf{S}}(\mathbf{x},\mathbf{t}) - \mathbf{m}^{\mathbf{N}}(\mathbf{x},\mathbf{t})\big]\, s_i(x,t)\Big] \exp(\Gamma_i(x,t))$$

where $\Gamma_i(x,t)$ is a country specific effect.

First aim: Estimate $\mathbf{m}^{\mathbf{N}}(\mathbf{x},\mathbf{t})$ and $\mathbf{m}^{\mathbf{S}}(\mathbf{x},\mathbf{t})$.

Maximum Likelihood

We assume for all x and t:

$$\Gamma_i(x,t) \sim N\left(0, \sigma_\Gamma^2(x,t)\right)$$

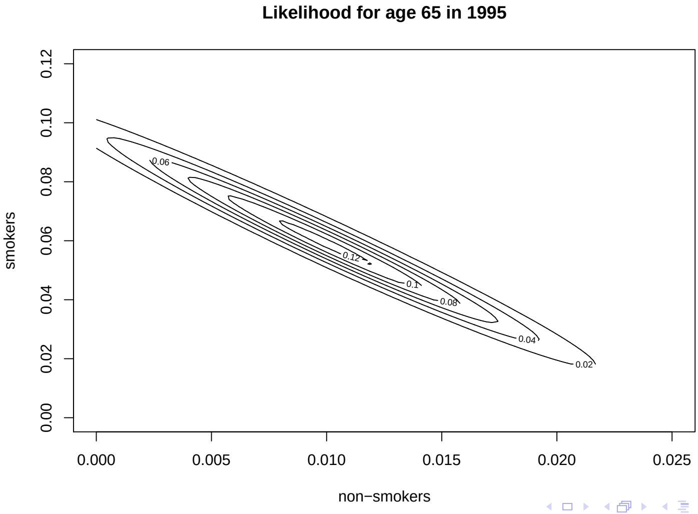
Model for log mortality: $\mu_i(x,t) = \log(m_i(x,t))$ is normal with expectation

$$\log\left\{m^N(x,t) + \left[m^S(x,t) - m^N(x,t)\right] s_i(x,t)\right\}$$

and variance $\sigma_\Gamma^2(x,t)$
likelihood function

$$L_0(m^N(x,t), m^S(x,t)) = \prod_i f\left(\mu_i(x,t) \middle| m^N(x,t), m^S(x,t)\right)$$

Maximum Likelihood



Adjusted Likelihood

For young ages we can argue that $\mu^N \approx \mu^S \approx \bar{\mu}(x,t)$.
Therefore, we estimate $\mu^N(x_0,t)$ and $\mu^S(x_0,t)$ by maximising the modified likelihood function

$$L_0(m^N(x,t), m^S(x,t)) = \prod_i f\left(\mu_i(x,t)\middle| m^N(x,t), m^S(x,t)\right)\psi(\mu^N(x,t))$$

where f is the density of the normal distribution with expectation

$$\log\left\{m^N(x,t) + \left[m^S(x,t) - m^N(x,t)\right]s_i(x,t)\right\}$$

and variance σ_Γ^2, and ψ is a normal density with expectation $\bar{\mu}(x,t)$ and variance σ_Γ^2.

Adjusted Likelihood

The variance can be estimated from the data by the sample variance:

$$\hat{\sigma}_{\Gamma}^2(x,t) = \frac{1}{|\mathcal{C}(t)| - 1} \sum_{i \in \mathcal{C}(t)} (\mu_i(x,t) - \bar{\mu}(x,t))^2$$

where $\bar{\mu}(x,t)$ denotes the average log mortality rate for age x in year t, and $\mathcal{C}(t)$ is the set of countries included in the study in year t.

Note that estimating the variance in that way is justified here by the assumption that $m^S(x,t) - m^N(x,t) \approx 0$ for young ages.

Adjusted Likelihood

These estimates are the basis for obtaining estimates at older ages. We assume that log mortality rates μ^N and μ^S for a given cohort will approximately follow a two-dimensional random walk with normally distributed increments
$\Delta\mu^N(x,t) = \mu^N(x,t) - \mu^N(x-1,t-1)$ and $\Delta\mu^S(x,t)$.

Adjusted Likelihood

The modified likelihood function for older ages we obtain in this way is given by

$$L\left(m^N, m^S\right) = \prod_i f\left(\mu_i(x,t) | m^N(x,t), m^S(x,t)\right) p\left(\left(\begin{array}{c} \Delta\mu^N(x,t) \\ \Delta\mu^S(x,t) \end{array}\right)\right)$$

where f is defined as above, and p is the density of a two-dimensional normal distribution with expectation $(M^N(x,t), M^S(x,t))$ and covariance matrix Σ_M. We estimate these parameters from the increments of the average mortality rates at age x in year t, that is,

$$\hat{M}^N(x,t) = \hat{M}^S(x,t) = \frac{1}{|\mathcal{C}(t)|} \sum_{i \in \mathcal{C}(t)} \Delta\mu_i(x,t)$$

Adjusted Likelihood

The variance matrix Σ_M needs to be specified:
we assume:

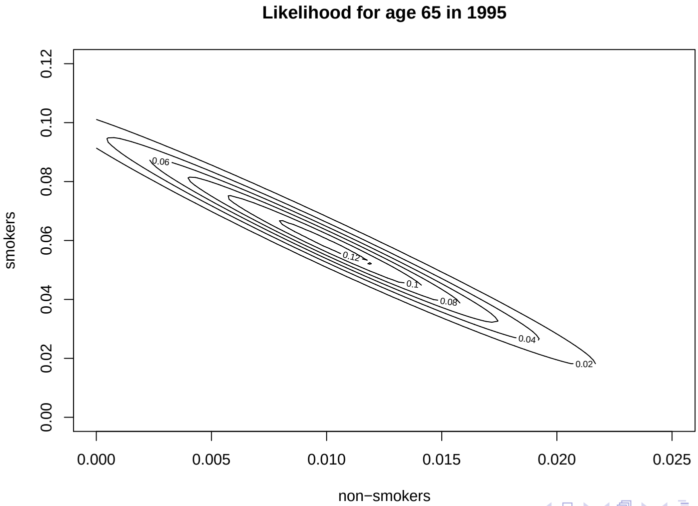
$$\mathrm{Var}(\Delta\mu^N(x,t)) = \mathrm{Var}(\Delta\mu^S(x,t)) = K\mathrm{Var}(\Delta\mu_i(x,t))$$

and the

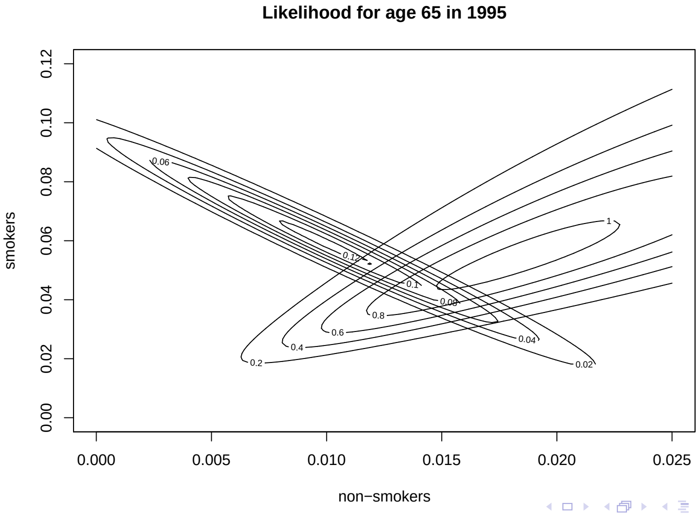
$$\mathrm{Corr}(\Delta\mu^N(x,t), \Delta\mu^S(x,t)) = \rho > 0$$

In the following we choose $K = 100$.

Adjusted Likelihood

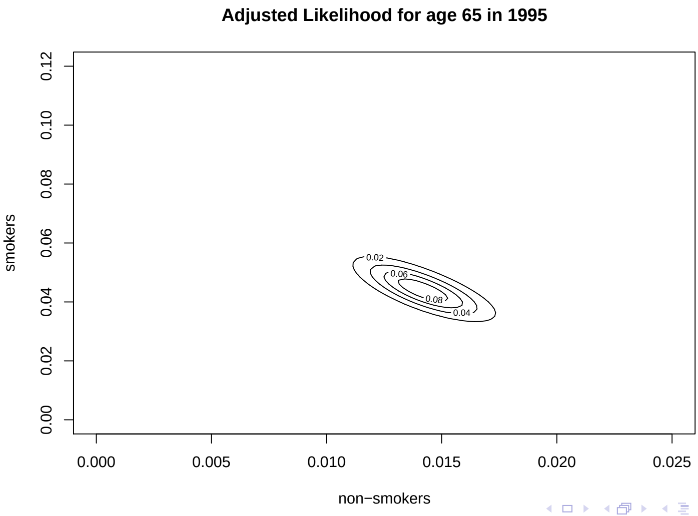


Adjusted Likelihood

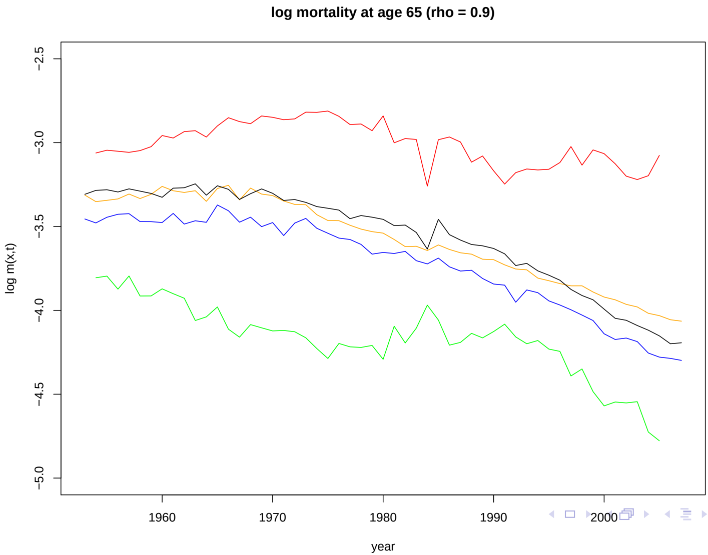


$\rho = 0.9$

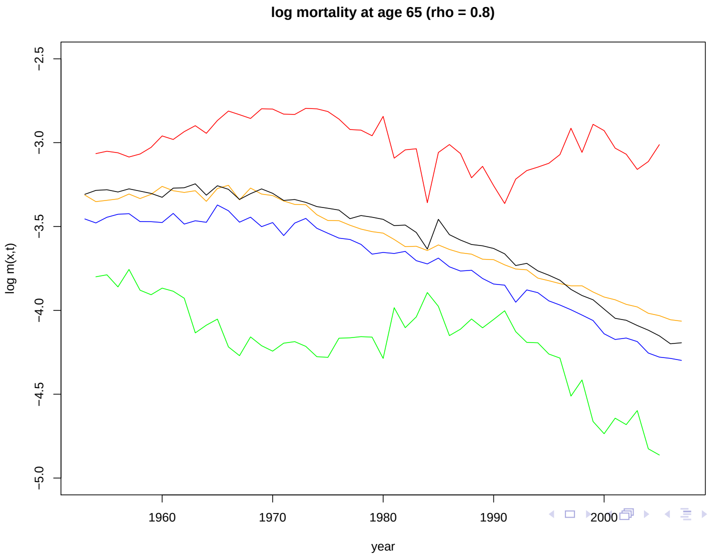
Adjusted Likelihood



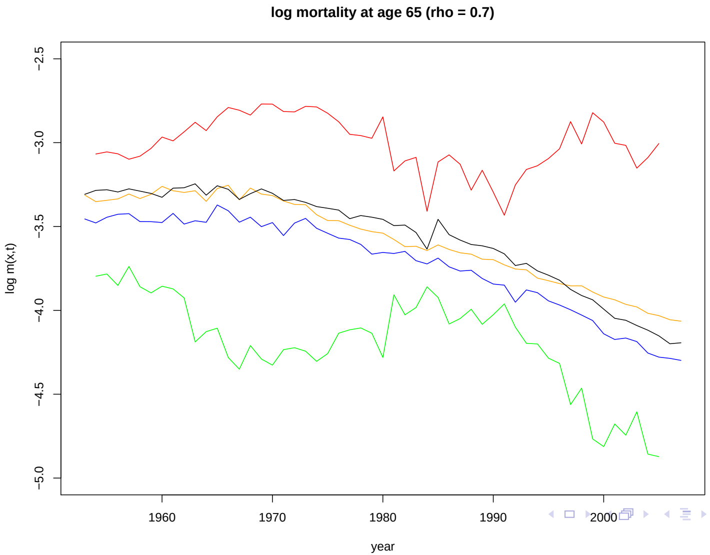
Estimated Log Mortality Rates



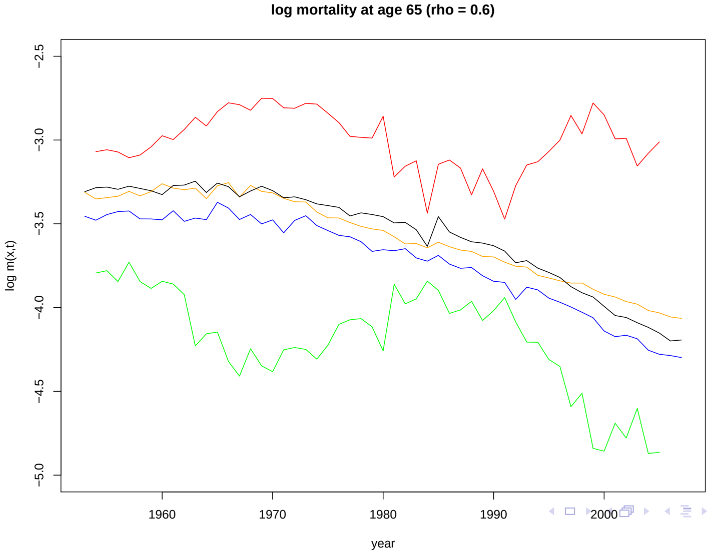
Estimated Log Mortality Rates



Estimated Log Mortality Rates



Estimated Log Mortality Rates



Simplifying Assumptions

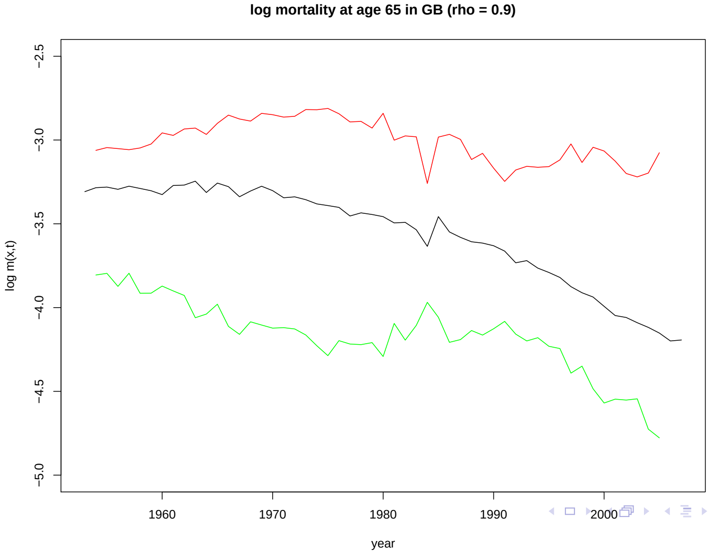
no improvement in smokers' mortality rates

$$m^S(x,t) = m^S(x)$$

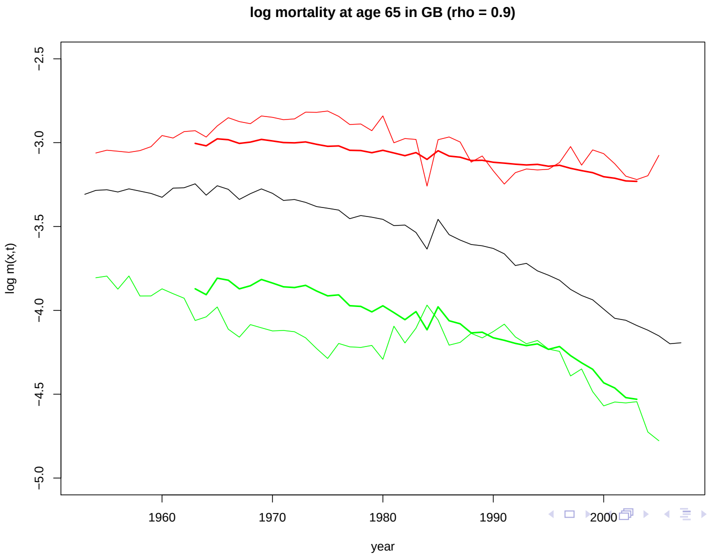
constant difference between smokers' and non-smokers' mortality rates

$$m^S(x,t) - m^N(x,t) = m^S(x) - m^N(x)$$

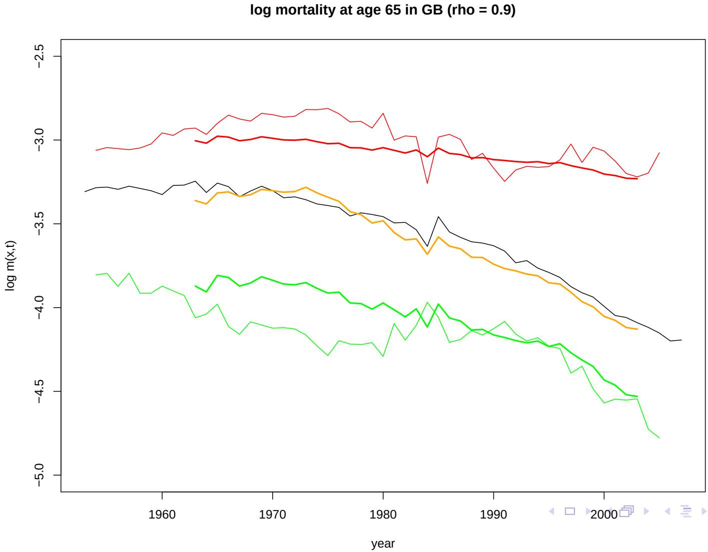
Constant Difference



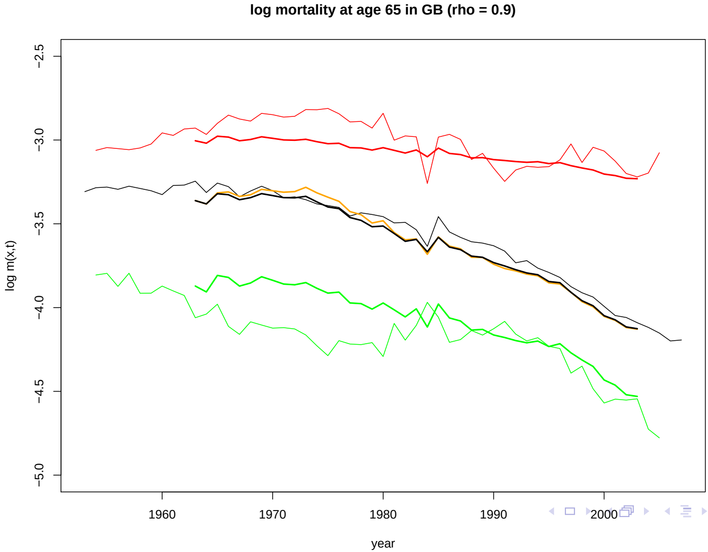
Constant Difference



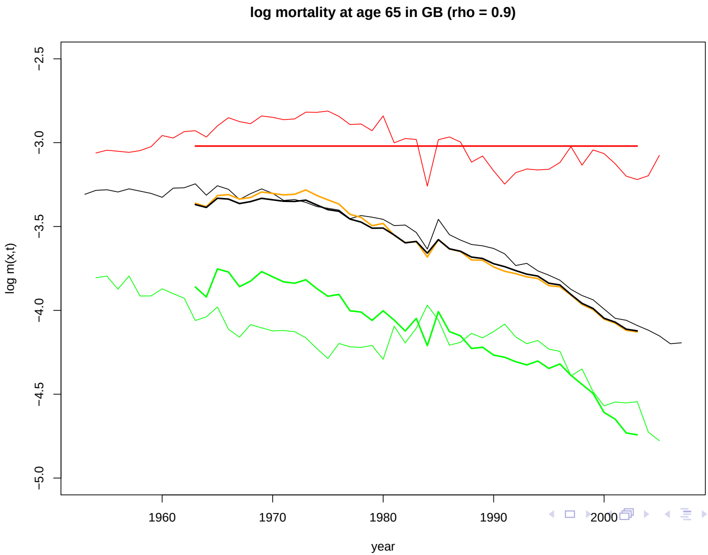
Constant Difference



Constant Difference



Constant Smokesrs' Mortality



Concluding remarks

- there is empirical evidence that smoking prevalence explains differences in country-specific mortality rates
- estimating this effect requires additional assumptions about the behaviour of non-smokers' and smokers' mortality rates (correlation)
- there is no clear downward trend in smokers' mortality rates, this corresponds to findings for British doctors by R. Doll, R. Peto, J. Boreham and I. Sutherland: "Mortality in relation to smoking: 50 years' observations on male British Doctors"
- mortality improvements are coming from reduced smoking prevalence and decreasing non-smokers' mortality
- the same approach can be used for other covariates that split the population into subgroups (marital status, ...)