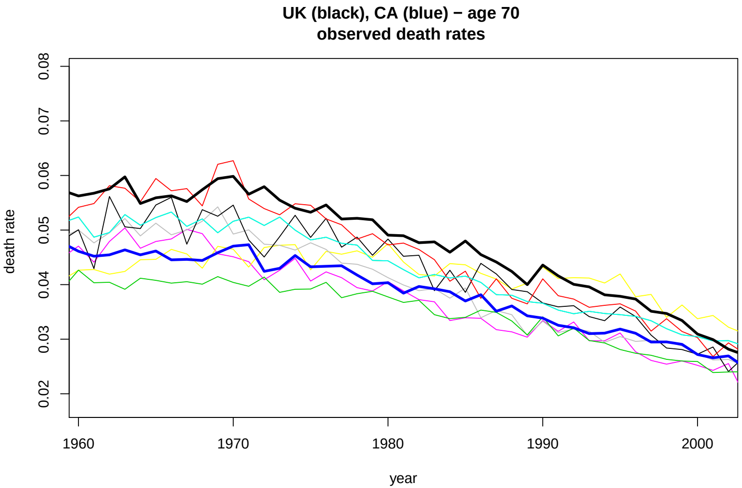
Mortality and Smoking Prevalence

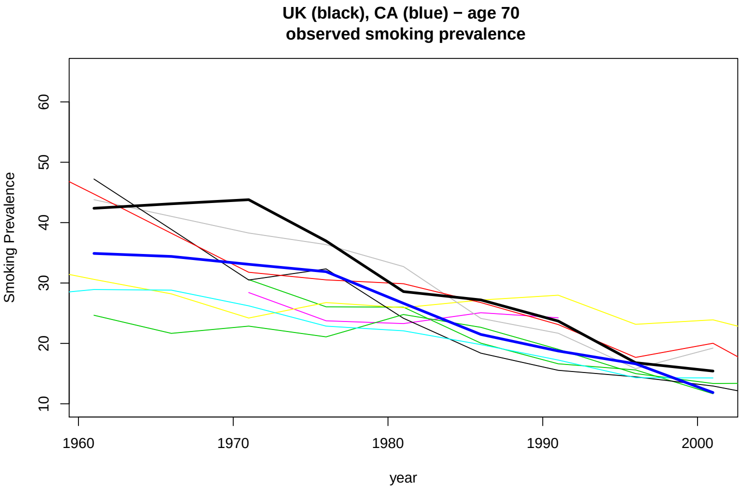
Torsten Kleinow & Andrew J.G. Cairns
Heriot-Watt University, Edinburgh

Life Conference 2011

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

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 Models for Multiple Populations

Consider k different populations (countries).

For each country i, time t (calendar year) and age x we observe

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

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

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

Mortality Models for Multiple Populations

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

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

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

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

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

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

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

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

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 - British Doctors

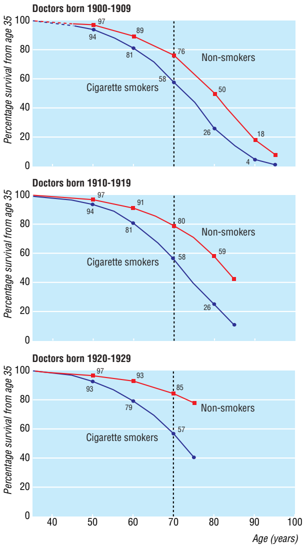
R. Doll, R. Peto, J. Boreham and I. Sutherland: “Mortality in relation to smoking: 50 years’ observations on male British Doctors”

- 34,439 British doctors,
- data about smoking habits was first obtained in 1951 and then periodically thereafter
- mortality was monitored for 50 years

Main results:

- substantial decrease in the mortality rates of non-smokers
- survival rates from age 35 for smokers are the same for cohorts born between 1900 to 1930, for non-smokers these survival rates have increased substantially

Smoking and Mortality - British Doctors



Source: R. Doll, R. Peto, J. Boreham and I. Sutherland: "Mortality in relation to smoking: 50 years' observations on male British Doctors"

Smoking and Mortality

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

$D_i(t,x)$: Number of deaths,
$D_i^N(t,x)$, $D_i^S(t,x)$ for non-smokers, smokers (not observed)
$D_i(t,x) = D_i^N(t,x) + D_i^S(t,x)$

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

$m_i(t,x)$: $= D_i(t,x)/E_i(t,x)$,
$m_i^N(t,x)$, $m_i^S(t,x)$, death rates

$s_i(t,x)$: Smoking prevalence, in $[0,1]$,
the number of smokers is $s_i(t,x)E_i(t,x)$

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

Smoking and Mortality

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

where

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

We obtain

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

Smoking and Mortality

Modelling assumptions:

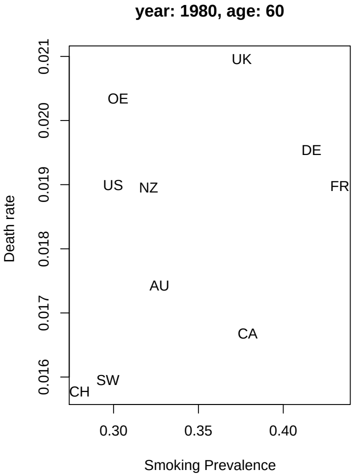
- Smoking prevalence has the same effect on mortality rates in all observed countries.
- Non-smokers' mortality in country i is the sum of general non-smokers' mortality and a "country effect"

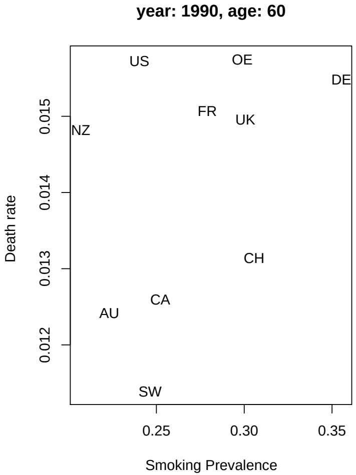
$$m_i(t,x) \quad = \quad \mathbf{m^N(t,x)} + \left[\mathbf{m^S(t,x)} - \mathbf{m^N(t,x)}\right] s_i(t,x) + C_i(t,x)$$

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

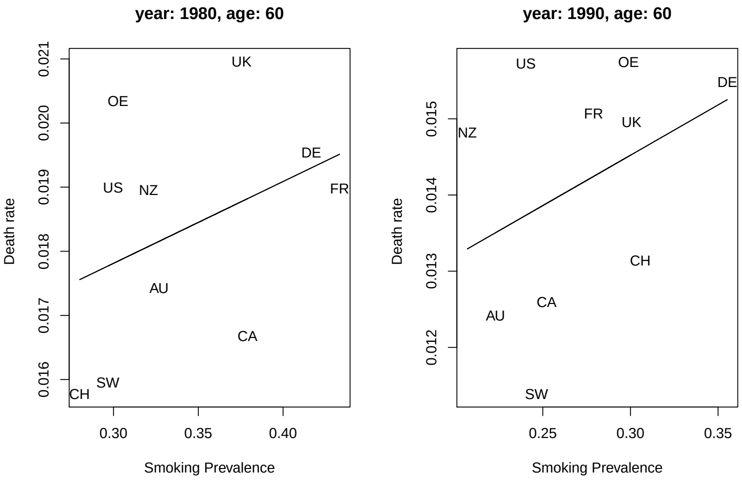
First aim: Estimate $\mathbf{m^N(t,x)}$ and $\mathbf{m^S(t,x)}$.

Smoking and Mortality





Smoking and Mortality



Simplifying Assumptions

$$m_i(t,x) \quad = \quad m^N(t,x) + \left[\mathbf{m^S(t,x)} - m^N(t,x)\right] s_i(t,x) + C_i(t,x)$$

Motivated by the findings for British doctors, we assume that there is:

no improvement in smokers' mortality rates

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

$$m_i(t,x) \quad = \quad m^N(t,x) + \left[\mathbf{m^S(x)} - m^N(t,x)\right] s_i(t,x) + C_i(t,x)$$

Constant smoker's mortality over time

Least-Square Estimation for a fixed age x:

$$\begin{aligned} \mathrm{MSE}_x(m^S, m^N) &= \sum_t \sum_i \left(C_i(t,x)\right)^2 \\ &= \sum_t \sum_i \Big(m_i(t,x) - m^N(t,x) - \big[m^S(x) - m^N(t,x)\big] s_i(t,x)\Big)^2 \end{aligned}$$

Note: $m^N = (m^N(1,x), \ldots, m^N(T,x))$
Choose m^S, m^N such that

$$\mathrm{MSE}_x(m^S, m^N) \longrightarrow min$$

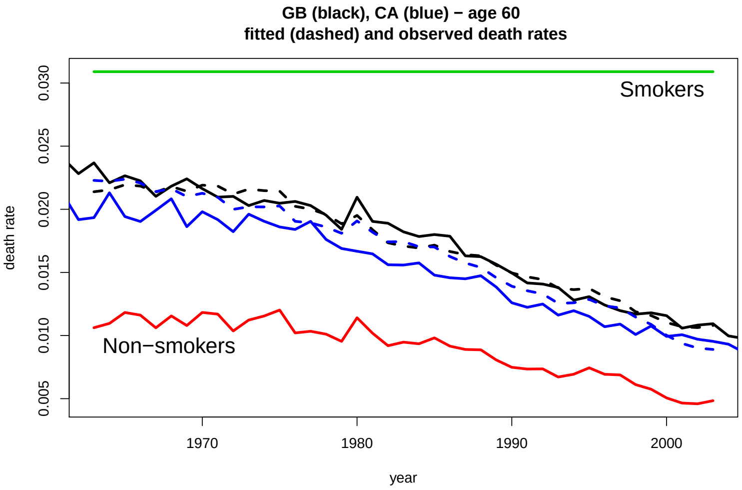
Constant smokers' mortality over time

Explicit solution for fixed age x is the solution of the following linear system of equations:

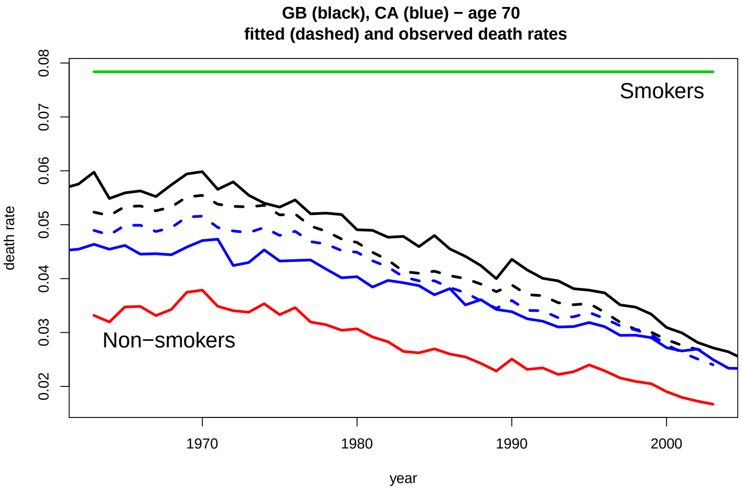
$$m^S = \frac{1}{\sum_t \sum_i s_i^2(t)} \sum_t \sum_i s_i(t) \underbrace{\left[m_i(t) - m^N(t)(1 - s_i(t))\right]}_{m^S s_i(t) + C_i(t)}$$

$$m^N(t) = \frac{\sum_i (1 - s_i(t)) m_i(t)}{\sum_i (1 - s_i(t))^2} - m^S \frac{\sum_i (1 - s_i(t)) s_i(t)}{\sum_i (1 - s_i(t))^2}$$

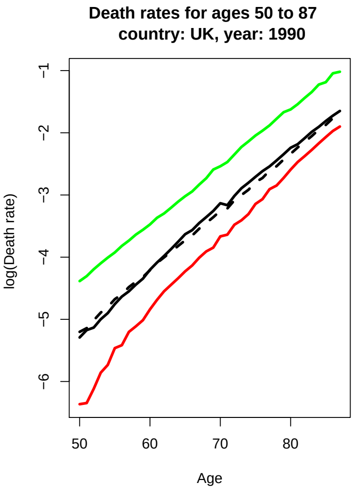
Constant smokers' mortality over time

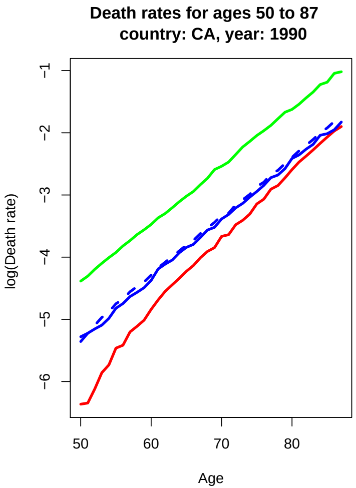


Constant smokers' mortality over time

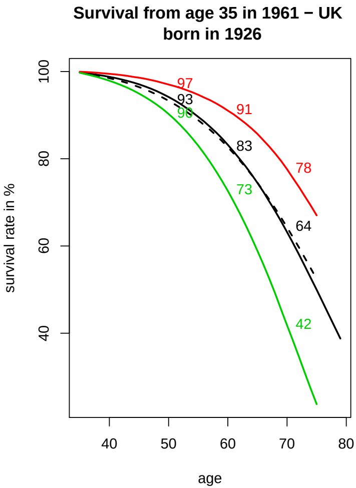


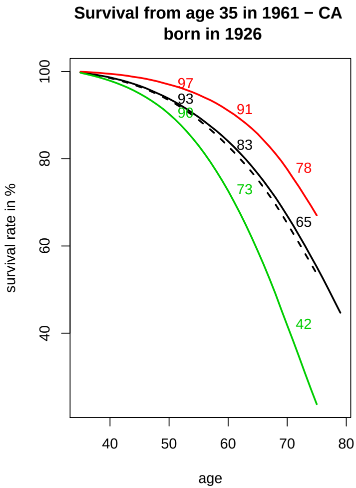
Constant smokers' mortality over time





Comparison with British Doctors





Comparison with British Doctors

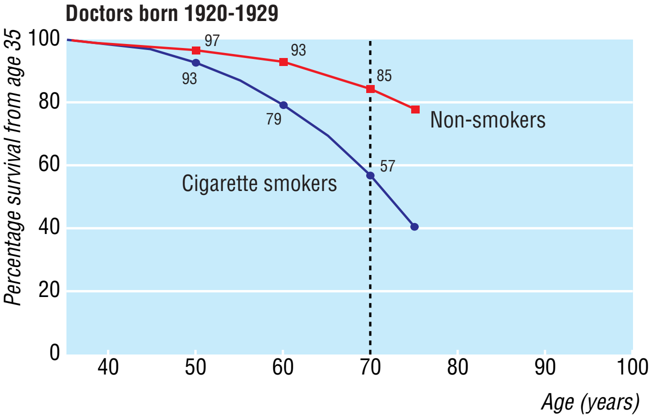


Figure: Source: R. Doll, R. Peto, J. Boreham and I. Sutherland: "Mortality in relation to smoking: 50 years' observations on male British Doctors"

Modelling the Country effect

Model for $m_i(t,x)$:

$$m_i(t,x) \quad = \quad m^N(t,x) + \left[m^S(x) - m^N(t,x)\right] s_i(t,x) + C_i(t,x)$$

"Core Hypothesis"

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

Since $m^S(x)$ is constant over time, the core hypothesis can only be fulfilled if $m^N(t,x) \to K(x) > 0$.

Modelling the Country effect

Since we consider a covariate (smoking) we change the core hypothesis to:
For any $i \neq j$ and any fixed age x holds:

$$s_i(t,x) = s_j(t,x) \, \forall \, t \qquad \Rightarrow \qquad m_i(t,x)/m_j(t,x) \not\to \infty$$

for $t \to \infty$

If the smoking prevalence is the same in any two countries in all future years, then the mortality rates should not diverge.

Scenarios

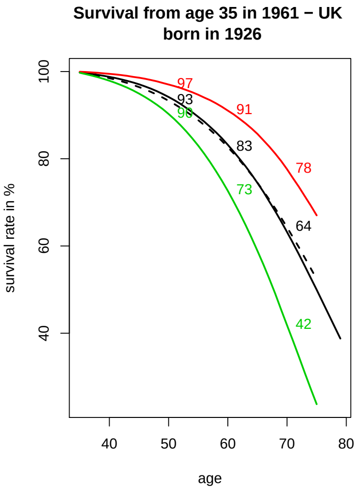
We can now investigate the effect of smoking on survival rates. With the estimates obtained earlier we consider

$$m_i(t,x) \;\; = \;\; m^N(t,x) + \mathbf{0.75}\left[m^S(x) - m^N(t,x)\right] s_i(t,x)$$

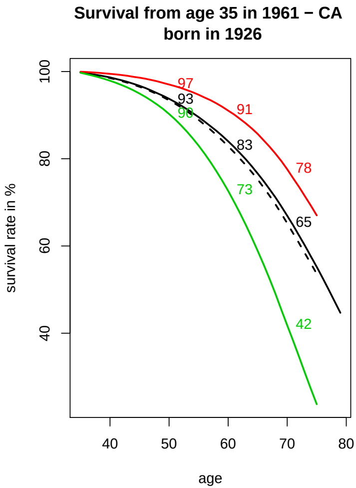
for the cohort aged 35 in 1961.
Rate of survival to age $x > 35$ for the cohort aged 35 in 1961:

$$\mathcal{S}(x, 1961, 35) = \prod_{j=1}^{x-35} \Big(1 - m_i(1961+j, 35+j)\Big)$$

Scenarios

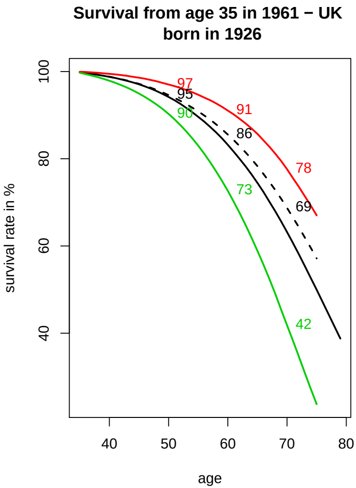


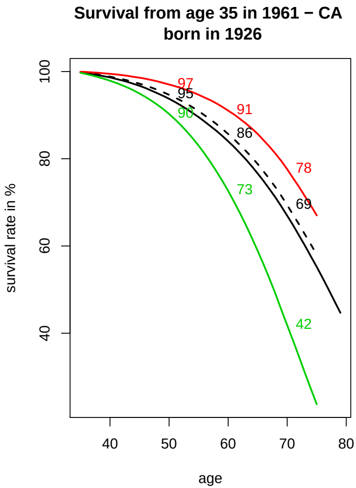
Survival from age 35 in 1961 − UK
born in 1926
survival rate in %
100
80
60
40
40
50
60
70
80
age
97
93
90
91
83
78
73
64
42



Survival from age 35 in 1961 − CA
born in 1926
survival rate in %
100
80
60
40
40
50
60
70
80
age
97
93
90
91
83
78
73
65
42

Smoking Prevalence reduced by 25%





Smoking and Mortality - Discussion

- there is empirical evidence that smoking prevalence can be used to model death rates for entire countries and explain differences in country-specific mortality rates
- there are also other country-specific factors that have an impact on mortality
- there is only one "trend" component (non-smokers' mortality) in our model
- we require an assumption about the relationship between mortality rates of smokers and non-smokers when no cessation data are available
- the assumption of constant smokers' mortality rates is very strong, and other assumptions should be investigated

What to model?

Basic model:

$$m_i(t,x) \quad = \quad m^N(t,x) + \left[m^S(x) - m^N(t,x)\right] s_i(t,x) + C_i(t,x)$$

To generate future mortality scenarios we need to model:

- $m^N(t,x)$ - any mortality model can be used
- $C_i(t,x)$? - Core hypothesis
- $s_i(t,x)$

Incorporating Smoking into a Mortality Model

Predictors of mortality:

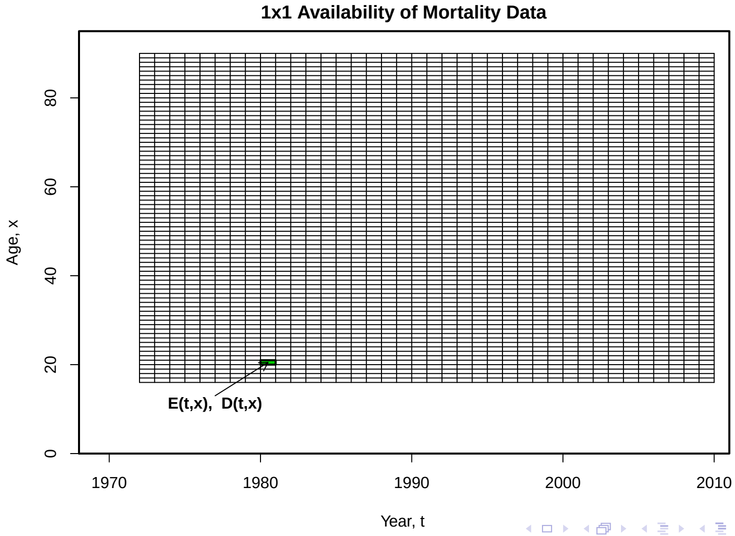
- Age
- Calendar year
- Year of birth (cohort effect)

Incorporating Smoking into a Mortality Model

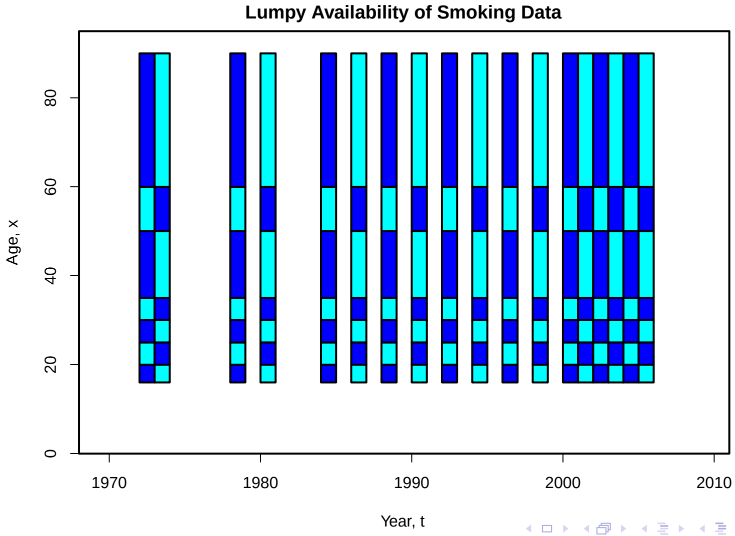
Predictors of mortality:

- Age
- Calendar year
- Year of birth (cohort effect)
- Smoking prevalence (possible cohort effect)

Available Data



Available Data



Smoking Prevalence Model

- $s(t, x) =$ smoking prevalence in year t age x
- $P(t, x) = \text{logit } s(t, x) = \log \left[s(t, x) / \big(1 - s(t, x)\big) \right]$
- $P(t, x) \in (-\infty, \infty)$
- $P(t, x)$ not directly observeable
  $\Rightarrow$ must be inferred from limited data

Smoking Prevalence Model

$$P(t, x) = \quad P(t-1, x-1) \qquad \text{status quo}$$

Smoking Prevalence Model

$$
\begin{aligned}
P(t,x) = \quad & P(t-1, x-1) && \text{status quo} \\
& +\delta\ C(t-1, x-1) && \text{diffusion}
\end{aligned}
$$

Smoking Prevalence Model

$$
\begin{aligned}
P(t,x) = \quad & P(t-1, x-1) && \text{status quo} \\
& +\delta\ C(t-1, x-1) && \text{diffusion} \\
& +\mu_x && \text{drift}
\end{aligned}
$$

Smoking Prevalence Model

$$\begin{aligned} P(t,x) = \quad & P(t-1, x-1) && \text{status quo} \\ & +\delta\ C(t-1, x-1) && \text{diffusion} \\ & +\mu_x && \text{drift} \\ & +Z_x(t) && \text{randomness} \end{aligned}$$

where $C(t-1, x-1) =$
$\left(\frac{1}{2}P(t-1, x-2) - P(t-1, x-1) + \frac{1}{2}P(t-1, x)\right)$ =convexity

Smoking Prevalence Model

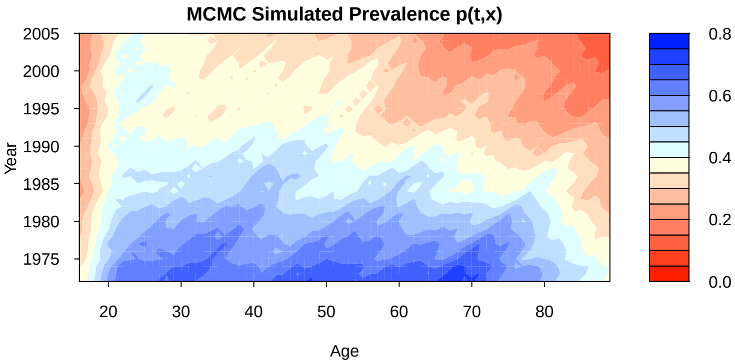
$$\begin{aligned} P(t,x) = \quad & P(t-1, x-1) && \text{status quo} \\ & +\delta\ C(t-1,x-1) && \text{diffusion} \\ & +\mu_x && \text{drift} \\ & +Z_x(t) && \text{randomness} \end{aligned}$$

where $C(t-1,x-1) =$
$\left(\frac{1}{2}P(t-1,x-2) - P(t-1,x-1) + \frac{1}{2}P(t-1,x)\right)$ =convexity

Estimation of the $P(t,x)$

- Uses Bayesian statistics
- Markov chain Monte Carlo (MCMC)

Smoking Prevalence Model



- Declining smoking by calendar year
- Age profile
- Cohort effect (by design)

Mortality Model with Smoking Prevalence

Cairns et al. (2009) (M7):

$$\operatorname{logit} q(t,x) = \kappa^{(1)}(t)+\kappa^{(2)}(t)(x-\bar{x})+\kappa^{(2)}(t)\left((x-\bar{x})^2-\sigma_X^2\right)+\gamma_4(t-x)$$

Mortality Model with Smoking Prevalence

Cairns et al. (2009) (M7):

$$\text{logit } q(t,x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t)\left((x - \bar{x})^2 - \sigma_X^2\right) + \gamma_4(t-x)$$

CBD-Plat

$$\text{logit } q(t,x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t)\left((x - \bar{x})^2 - \sigma_X^2\right) + \gamma_4(t-x)$$

$$+\beta^{(0)}(x)$$

Mortality Model with Smoking Prevalence

Cairns et al. (2009) (M7):

$$\text{logit } q(t,x) = \kappa^{(1)}(t)+\kappa^{(2)}(t)(x-\bar{x})+\kappa^{(2)}(t)\left((x-\bar{x})^2-\sigma_X^2\right)+\gamma_4(t-x)$$

CBD-Plat

$$\text{logit } q(t,x) = \kappa^{(1)}(t)+\kappa^{(2)}(t)(x-\bar{x})+\kappa^{(2)}(t)\left((x-\bar{x})^2-\sigma_X^2\right)+\gamma_4(t-x)$$

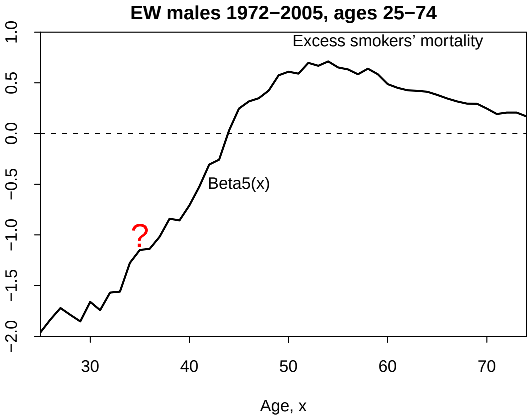
$$+\beta^{(0)}(x)$$

CBD-P-Smoking

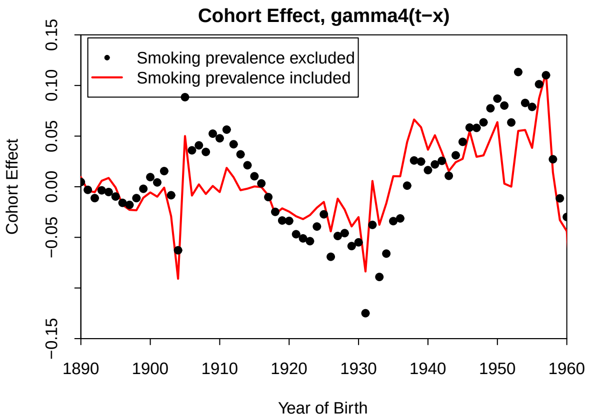
$$\log m(t,x) = \kappa^{(1)}(t)+\kappa^{(2)}(t)(x-\bar{x})+\kappa^{(2)}(t)\left((x-\bar{x})^2-\sigma_X^2\right)+\gamma_4(t-x)$$

$$+\beta^{(0)}(x) + \beta^{(5)}(x)s(t,x)$$

"Excess" Smoker Mortality



Fitted Cohort Effect



Smoking Prevalence Model: Discussion

- Smoking prevalence as a covariate
  $\beta_x^{(5)}$ $<$ excess mortality due to regular smoking
  $\Rightarrow$ less impact than we might expect
  - Smoking prevalence $>$ Prevalence of regular smokers
  - Non-smokers include recent *quitters*
- Younger adults $\Rightarrow$ smoking is "beneficial"
  BUT linked with lifestyle:
  - Smoking $\Rightarrow$ (??) less likely to engage in hazardous activities