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
Smoking

UK (black), CA (blue) – age 70 observed smoking prevalence

year

Smoking Prevalence

10 20 30 40 50 60
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

\[ D_i(t, x) : \text{Number of deaths}, \]

\[ E_i(t, x) : \text{Exposure-to-risk} \]

\[ m_i(t, x) : = \frac{D_i(t, x)}{E_i(t, x)}, \text{death rate} \]

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 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

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
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

\[
D_i(t, x): \text{Number of deaths,} \\
D_i^N(t, x), D_i^S(t, x) \text{ for non-smokers, smokers (not observed)} \\
D_i(t, x) = D_i^N(t, x) + D_i^S(t, x)
\]

\[
E_i(t, x): \text{Exposure-to-risk} \\
m_i(t, x) = D_i(t, x)/E_i(t, x), \\
m_i^N(t, x), m_i^S(t, x), \text{death rates}
\]

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

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

\[ D_i(t, x) = D_i^N(t, x) + D_i^S(t, x) \]
\[ = m_i^N(t, x)[1 - s_i(t, x)] E_i(t, x) + m_i^S(t, x)s_i(t, x)E_i(t, x) \]

where

\[ m_i^N(t, x) = \frac{D_i^N(t, x)}{[1 - s_i(t, x)] E_i(t, x)} \]
\[ m_i^S(t, x) = \frac{D_i^S(t, x)}{s_i(t, x)E_i(t, x)} \]

We obtain

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

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

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

First aim: Estimate $m^N(t, x)$ and $m^S(t, x)$. 
Smoking and Mortality

year: 1980, age: 60

year: 1990, age: 60
Smoking and Mortality

year: 1980, age: 60

year: 1990, age: 60
Simplifying Assumptions

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

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

no improvement in smokers’ mortality rates
\[ m^S(t, x) = m^S(x) \]

\[ m_i(t, x) = m^N(t, x) + [m^S(x) - m^N(t, x)] s_i(t, x) + C_i(t, x) \]
Constant smoker’s mortality over time

Least-Square Estimation for a fixed age $x$:

$\text{MSE}_x(m^S, m^N) = \sum_t \sum_i (C_i(t, x))^2$

$= \sum_t \sum_i \left( m_i(t, x) - m^N(t, x) - [m^S(x) - m^N(t, x)] s_i(t, x) \right)^2$

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

$\text{MSE}_x(m^S, m^N) \rightarrow min$
Constant smokers’ mortality over time

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

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

$$m^N(t) = \frac{\sum_i (1 - s_i(t)) m_i(t)}{\sum_i (1 - s_i(t))^2} - m^S \frac{\sum_i (1 - s_i(t)) s_i(t)}{\sum_i (1 - s_i(t))^2}$$
Constant smokers’ mortality over time

GB (black), CA (blue) – age 60
fitted (dashed) and observed death rates

Smokers
Non-smokers

year
death rate

Constant smokers’ mortality over time

GB (black), CA (blue) – age 70
fitted (dashed) and observed death rates

Smokers
Non–smokers
Constant smokers’ mortality over time

Death rates for ages 50 to 87
country: UK, year: 1990

Death rates for ages 50 to 87
country: CA, year: 1990
Comparison with British Doctors

Survival from age 35 in 1961 – UK
born in 1926

Survival from age 35 in 1961 – CA
born in 1926

Survival rate in %

age

93
97
64
83
78
91
73
90
42

Survival rate in %

age

93
97
65
83
78
91
73
65
42
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 $m_i(t, x)$:

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

“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:

\[
s_i(t, x) = s_j(t, x) \forall t \quad \Rightarrow \quad m_i(t, x)/m_j(t, x) \not\to \infty
\]

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

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

for the cohort aged 35 in 1961.

Rate of survival to age \( x > 35 \) for the cohort aged 35 in 1961:

\[ S(x, 1961, 35) = \prod_{j=1}^{x-35} \left( 1 - m_i(1961 + j, 35 + j) \right) \]
Scenarios

Survival from age 35 in 1961 – UK born in 1926

Survival from age 35 in 1961 – CA born in 1926
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:

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

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

1x1 Availability of Mortality Data

E(t,x), D(t,x)

Year, t

Age, x
Available Data

Lumpy Availability of Smoking Data

Age, x

Year, t
Smoking Prevalence Model

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

\[ P(t, x) = P(t - 1, x - 1) \] status quo
Smoking Prevalence Model

\[ P(t, x) = P(t - 1, x - 1) \text{ status quo} \]
\[ + \delta C(t - 1, x - 1) \text{ diffusion} \]
Smoking Prevalence Model

\[ P(t, x) = P(t - 1, x - 1) \quad \text{status quo} \]
\[ + \delta \ C(t - 1, x - 1) \quad \text{diffusion} \]
\[ + \mu_x \quad \text{drift} \]
Smoking Prevalence Model

\[ P(t, x) = P(t - 1, x - 1) \quad \text{status quo} \]
\[ + \delta \ C(t - 1, x - 1) \quad \text{diffusion} \]
\[ + \mu_x \quad \text{drift} \]
\[ + Z_x(t) \quad \text{randomness} \]

where \( C(t - 1, x - 1) = \)
\[ (\frac{1}{2} P(t - 1, x - 2) - P(t - 1, x - 1) + \frac{1}{2} P(t - 1, x)) = \text{convexity} \]
Smoking Prevalence Model

\[ P(t, x) = P(t - 1, x - 1) \quad \text{status quo} \]
\[ + \delta \ C(t - 1, x - 1) \quad \text{diffusion} \]
\[ + \mu_x \quad \text{drift} \]
\[ + Z_x(t) \quad \text{randomness} \]

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

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):

\[
\text{logit } q(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left((x - \bar{x})^2 - \sigma_X^2\right) + \gamma_4(t - x)
\]
Mortality Model with Smoking Prevalence

Cairns et al. (2009) (M7):

\[
\logit q(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left( (x - \bar{x})^2 - \sigma_x^2 \right) + \gamma_4 (t - x)
\]

CBD-Plat

\[
\logit q(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left( (x - \bar{x})^2 - \sigma_x^2 \right) + \gamma_4 (t - x) + \beta^{(0)}(x)
\]
Mortality Model with Smoking Prevalence

Cairns et al. (2009) (M7):

\[
\text{logit } q(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left( (x - \bar{x})^2 - \sigma^2_X \right) + \gamma_4(t-x) \\
+ \beta^{(0)}(x)
\]

CBD-Plat

\[
\text{logit } q(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left( (x - \bar{x})^2 - \sigma^2_X \right) + \gamma_4(t-x) \\+ \beta^{(0)}(x)
\]

CBD-P-Smoking

\[
\text{log } m(t, x) = \kappa^{(1)}(t) + \kappa^{(2)}(t)(x - \bar{x}) + \kappa^{(2)}(t) \left( (x - \bar{x})^2 - \sigma^2_X \right) + \gamma_4(t-x) \\+ \beta^{(0)}(x) + \beta^{(5)}(x)s(t, x)
\]
“Excess” Smoker Mortality


Excess smokers’ mortality

Beta5(x)

Age, x
Fitted Cohort Effect

Cohort Effect, $\gamma_4(t-x)$

- Smoking prevalence excluded
- Smoking prevalence included
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