SUPPLEMENTARY FILE

SUPPLEMENTARY FILE 1. Detailed technical appendix

Microsimulation framework

The UK Health Forum (UKHF) simulation consists of two modules. The first module calculates the predictions of risk factor trends over time based on data from rolling cross-sectional studies. The second module performs the microsimulation of a virtual population, generated with demographic characteristics matching those of the observed data. The health trajectory of each individual from the population is simulated over time allowing them to contract, survive or die from a set of diseases or injuries related to the analysed risk factors. The detailed description of the two modules is presented below.

Module one: Predictions of smoking over time

Table 1 presents the categories used for modelling smoking over time. Other risk factors can be modelled using the UKHF model, however only smoking is presented here.

Table 1 Description of the categories used to model smoking

Risk factor (RF)	*Number of categories (N)*	*Categories*
Smoking	3	1 Never Smoker 2 Ex-smoker 3 Smoker

For each RF, let N be the number of categories for a given risk factor, e.g. $N = 3$ for smoking. Let $k = 1, 2, \ldots, N$ number these categories and $p_k(t)$ denote the prevalence of individuals with RF values that correspond to the category k at time t. We estimate $p_k(t)$ using multinomial logistic regression model with prevalence of RF category k as the outcome, and time t as a single explanatory variable. For $k < N$, we have

$$\ln\left(\frac{p_k(t)}{p_1(t)}\right) = \beta_0^k + \beta_1^k t \tag{1.1}$$

The prevalence of the first category is obtained by using the normalisation constraint $\sum_{k=1}^{N} p_k(t) = 1$. Solving equation (1.1) for $p_k(t)$, we obtain

$$p_k(t) = \frac{\exp\left(\beta_0^k + \beta_1^k t\right)}{1 + \sum_{k'=1}^{N} \exp\left(\beta_0^{k'} + \beta_1^{k'} t\right)}, \tag{1.2}$$

which respects all constraints on the prevalence values, i.e. normalisation and [0, 1] bounds.

Multinomial logistic regression

Measured data consist of sets of probabilities, with their variances, at specific time values (typically the year of the survey). For any particular time the sum of these probabilities is unity. Typically such data might be the probabilities of never smoker, ex-smoker and smoker as they are extracted from the survey data set. Each data point is treated as a normally distributed[1] random variable; together they are a set of N groups (number of years) of K probabilities $\{\{t_i, \mu_{ki}, \sigma_{ki} \,|k\in[0,K\text{-}1]\} \mid i\in[0,N\text{-}1]\}$. For each year the set of K probabilities form a distribution – their sum is equal to unity.

The regression consists of fitting a set of logistic functions $\{p_k(\mathbf{a}, \mathbf{b}, t)|k\in[0,K\text{-}1]\}$ to these data – one function for each k-value. At each time value the sum of these functions is unity. Thus, for example, when measuring smoking in the three states already mentioned, the $k = 0$ regression function represents the probability of being a never smoker over time, $k = 1$ the probability of being an ex-smoker and $k = 2$ the probability of being a smoker.

The regression equations are most easily derived from a familiar least square minimization. In the following equation set the weighted difference between the measured and predicted probabilities is written as S; the logistic regression functions $p_k(\mathbf{a},\mathbf{b};t)$ are chosen to be ratios of sums of exponentials (This is equivalent to modelling the log probability ratios, p_k/p_0, as linear functions of time.)

$$S(\mathbf{a},\mathbf{b}) = \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k(\mathbf{a},\mathbf{b};t_i)-\mu_{ki}\right)^2}{\sigma_{ki}^2} \tag{1.3}$$

$$p_k(\mathbf{a},\mathbf{b},t) \equiv \frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}} \tag{1.4}$$

$$\mathbf{a} \equiv (a_0,a_1,..,a_{K-1}), \quad \mathbf{b} \equiv (b_0,b_1,..,b_{K\text{-}1})$$
$$A_0 \equiv 0, \quad A_k \equiv a_k + b_k t$$

The parameters A_0, a_0 and b_0 are all zero and are used merely to preserve the symmetry of the expressions and their manipulation. For a K-dimensional set of probabilities there will be $2(K\text{-}1)$ regression parameters to be determined.

[1] Depending on the circumstances this assumption will be more or less accurate and more or less necessary. In general, it is both extremely useful and accurate. For simple surveys the individual Bayesian prior and posterior probabilities are Beta distributions – the likelihood being binomial. For reasonably large samples, the approximation of the beta distributions by normal distributions is both legitimate and a practical necessity. For complex, multi-PSU, stratified surveys, it is again assumed that these base probabilities are approximately normally distributed and, again, it is an assumption that makes the analysis tractable.

Depending on the nature of the raw data set it may be possible to use non-parametric statistical methods for this analysis. This is possible for the HSE and GHS data sets of this study but when this has been done the authors can report no discernible difference in the results.

For a given dimension K there are K-1 independent functions p_k – the remaining function being determined from the requirement that complete set of K form a distribution and sum to unity.

Note that the parameterization ensures that the necessary requirement that each p_k be interpretable as a probability – a real number lying between 0 and 1.
The minimum of the function S is determined from the equations

$$\frac{\partial S}{\partial a_j} = \frac{\partial S}{\partial b_j} = 0 \qquad \text{for j=1,2,....,k-1} \tag{1.5}$$

noting the relations

$$\begin{aligned} \frac{\partial p_k}{\partial A_j} &= \frac{\partial}{\partial A_j}\left(\frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}}\right) = p_k\delta_{kj} - p_k p_j \\ \frac{\partial}{\partial a_j} &= \frac{\partial}{\partial A_j} \\ \frac{\partial}{\partial b_j} &= t\frac{\partial}{\partial A_j} \end{aligned} \tag{1.6}$$

The values of the vectors **a**, **b** that satisfy these equations are denoted $\hat{\mathbf{a}}, \hat{\mathbf{b}}$. They provide the trend lines, $p_k\left(\hat{\mathbf{a}}, \hat{\mathbf{b}}; t\right)$, for the separate probabilities. The confidence intervals for the trend lines are derived most easily from the underlying Bayesian analysis of the problem.

Bayesian interpretation

The 2K-2 regression parameters {**a**,**b**} are regarded as random variables whose posterior distribution is proportional to the function exp(-S(**a**,**b**)). The maximum likelihood estimate of this probability distribution function, the minimum of the function S, is obtained at the values $\hat{\mathbf{a}}, \hat{\mathbf{b}}$. Other properties of the (2K-2)-dimensional probability distribution function are obtained by first approximating it as a (2K-2)-dimensional normal distribution whose mean is the maximum likelihood estimate. This amounts to expanding the function S(**a**,**b**) in a Taylor series as far as terms quadratic in the differences $(\mathbf{a}-\hat{\mathbf{a}}), (\mathbf{b}-\hat{\mathbf{b}})$ about the maximum likelihood estimate $\hat{\mathbf{S}} \equiv S\left(\hat{\mathbf{a}}, \hat{\mathbf{b}}\right)$. Hence

$$\begin{aligned} S(\mathbf{a},\mathbf{b}) &= \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k(\mathbf{a},\mathbf{b};t_i)-\mu_{ki}\right)^2}{\sigma_{ki}^2} \\ &\equiv S\left(\hat{a},\hat{b}\right) + \tfrac{1}{2}\left(a-\hat{a}, b-\hat{b}\right)P^{-1}\left(a-\hat{a}, b-\hat{b}\right)+... \\ &\approx S\left(\hat{a},\hat{b}\right) + \tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{a}_j}\left(a_j-\hat{a}_j\right) + \tfrac{1}{2}\sum_{i,j}(a_i-\hat{a}_i)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right) + \\ &+ \tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{a}_j}\left(a_j-\hat{a}_j\right) + \tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right) \end{aligned} \tag{1.7}$$

The (2*K*-2)-dimensional covariance matrix *P* is the inverse of the appropriate expansion coefficients. This matrix is central to the construction of the confidence limits for the trend lines.

Estimation of the confidence intervals

The logistic regression functions $p_k(t)$ can be approximated as a normally distributed time-varying random variable $N\left(\hat{p}_k(t), \sigma_k^2(t)\right)$ by expanding p_k about its maximum likelihood estimate (the trend line) $\hat{p}_k(t) = p\left(\hat{\mathbf{a}}, \hat{\mathbf{b}}, t\right)$

$$\begin{aligned} p_k\left(\mathbf{a},\mathbf{b},t\right) &= p_k\left(\hat{\mathbf{a}}+\mathbf{a}-\hat{\mathbf{a}}, \hat{\mathbf{b}}+\mathbf{b}-\hat{\mathbf{b}}, t\right) \\ &= \hat{p}_k(t) + \left(\nabla_{\hat{a}}, \nabla_{\hat{b}}\right)\hat{p}_k(t)\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix} + \ldots \end{aligned} \tag{1.8}$$

Denoting mean values by angled brackets, the variance of p_k is thereby approximated as

$$\begin{aligned} &\sigma_k^2(t) \equiv \left\langle \left(p_k\left(\mathbf{a},\mathbf{b},t\right) - \hat{p}_k(t)\right)^2 \right\rangle = \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)\left\langle \begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}^T \right\rangle \times \\ &\left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)^T = \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right) P \left(\nabla_{\hat{a}}\hat{p}_k(t), \nabla_{\hat{b}}\hat{p}_k(t)\right)^T \end{aligned} \tag{1.9}$$

When *K*=3 this equation can be written as the 4-dimensional inner product

$$\sigma_k^2(t) = \begin{pmatrix} \frac{\partial \hat{p}_k(t)}{\partial \hat{a}_1} & \frac{\partial \hat{p}_k(t)}{\partial \hat{a}_2} & \frac{\partial \hat{p}_k(t)}{\partial \hat{b}_1} & \frac{\partial \hat{p}_k(t)}{\partial \hat{b}_2} \end{pmatrix} \begin{bmatrix} P_{aa11} & P_{aa12} & P_{ab11} & P_{ab12} \\ P_{aa21} & P_{aa22} & P_{ab21} & P_{ab22} \\ P_{ba11} & P_{ba12} & P_{bb11} & P_{bb12} \\ P_{ba21} & P_{ba22} & P_{bb21} & P_{bb22} \end{bmatrix} \begin{pmatrix} \frac{\partial \hat{p}_k(t)}{\partial \hat{a}_1} \\ \frac{\partial \hat{p}_k(t)}{\partial \hat{a}_2} \\ \frac{\partial \hat{p}_k(t)}{\partial \hat{b}_1} \\ \frac{\partial \hat{p}_k(t)}{\partial \hat{b}_2} \end{pmatrix} \tag{1.10}$$

where $P_{cdij} \equiv \left\langle (c_i - \hat{c}_i)(d_j - \hat{d}_j) \right\rangle$. The 95% confidence interval for $p_k(t)$ is centred given as $\left[\hat{p}_k(t) - 1.96\sigma_k(t), \mathrm{p_k}(t) + 1.96\sigma_k(t)\right]$.

Module two: Microsimulation

Microsimulation initialisation: birth, disease and death models

Simulated people are generated with the correct demographic statistics in the simulation's start-year. In this year women are stochastically allocated the number and years of birth of their children – these are generated from known fertility and mother's age at birth statistics

(valid in the start-year). If a woman has children then those children are generated as members of the simulation in the appropriate birth year.

The microsimulation is provided with a list of relevant diseases. These diseases used the best available incidence, mortality, survival, relative risk and prevalence statistics (by age and gender). At initialisation, the prevalence statistics are used to generate stochastically a simulated person's initial disease state in the simulation start-year. The population of people, so initialised, will stochastically reproduce the national prevalence statistics for each disease. It is assumed that at initialisation the diseases are independent random variables. In the course of their lives, simulated people can die from one of the diseases caused by smoking that they might have acquired or from some other cause. The probability that a person of a given age and gender dies from a cause other than the disease are calculated in terms of known death and disease statistics valid in the start-year. It is constant over the course of the simulation. The survival rates from smoking-related diseases will change as a consequence of the changing distribution of smoking prevalence in the population.

The microsimulation incorporates a sophisticated economic module. The module employs Markov-type simulation of long-term health benefits, health care costs and cost-effectiveness of specified interventions.

Non-health costs were incorporated into the model based on a toolkit in development by analysts in the Department of Health to which we were given early access by Gavin Roberts gavin.roberts@york.ac.uk.

This section provides an overview of the initialisation of the microsimulation model and will be expanded upon in the next sections.

Population models

Populations are implemented as instances of the TPopulation C++ class. The TPopulation class is created from a population (*.ppl) file. Usually a simulation will use only one population but it can simultaneously process multiple populations (for example, different ethnicities within a national population).

Population Editor

The Population Editor Allows editing and testing of TPopulation objects.

The population is created in the start-year and propagated forwards in time by allowing females to give birth. The population in each year follows the population predictions produced by ONS (Office for National Statistics 2012). The ONS population projections for

England have been used as an approximation for the UK. People within the model can die from specific diseases or from other causes. The <deaths by year by sex by age> file is a necessary input to the model – valid in the start year and usually referred to as the deaths from all causes file.

Distributions

Distribution name	symbol	note
MalesByAgeByYear	$p_m(a)$	Input in year$_0$ – probability of a male having age a
FemalesByAgeByYear	$p_f(a)$	Input in year$_0$ – probability of a female having age a
BirthsByAgeofMother	$p_b(a)$	Input in year$_0$ – conditional probability of a birth at age a| the mother gives birth.
NumberOfBirths	$p_\lambda(n)$	λ≡TFR, Poisson distribution, probability of giving birth to n children
MaleDeathByAge	$p_{\Omega m}(a)$	Input in year$_0$, probability of a male dying at age a
FemaleDeathByAge	$p_{\Omega f}(a)$	Input in year$_0$, probability of a female dying at age a

Birth model

Any female in the child bearing years {*AgeAtChild.lo*, *AgeAtChild.hi*} is deemed capable of giving birth. The number of children, n, that she has in her life is dictated by the Poisson distribution $p_\lambda(n)$ where the mean of the Poisson distribution is the Total Fertility Rate (TFR) parameter[2].

The probability that a mother (who does give birth) gives birth to a child at age a is determined from the BirthsByAgeOfMother distribution as $p_b(a)$. For any particular mother the births of multiple children are treated as independent events, so that the probability that a mother who produces N children produces n of them at age a is given as the Binomially distributed variable,

$$p_b\left(n\,at\,a \mid N\right)=\frac{N!}{n!\left(N-n\right)!}\left(p_b\left(a\right)\right)^n\left(1-p_{bm}\left(a\right)\right)^{N-n} \tag{1.11}$$

The probability that the mother gives birth to n children at age a is

$$p_b\left(n\,at\,a\right)=e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{N!}p_b\left(n\,at\,a \mid N\right)=e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{n!\left(N-n\right)!}\left(p_b\left(a\right)\right)^n\left(1-p_b\left(a\right)\right)^{N-n} \tag{1.12}$$

Performing the summation in this equation gives the simplifying result that the probability $p_b(n\ at\ a)$ is itself Poisson distributed with mean parameter $\lambda p_b(a)$,

$$p_b\left(n\,at\,a\right)=e^{-\lambda p_b(a)}\frac{\left(\lambda p_b\left(a\right)\right)^n}{n!}=p_{\lambda p_b(a)}\left(n\right) \tag{1.13}$$

Thus, on average, a mother at age a will produce $\lambda p_b(a)$ children in that year.
The gender of the children[3] is determined by the probability p_{male}=1-p_{female}. In the baseline model this is taken to be the probability $N_m/(N_m+N_f)$.

[2] This could be made to be time dependent; in the baseline model it is constant.
[3] The probability of child gender can be made time dependent.

The Population editor' menu item Population Editor\Tools\Births\show random birthList creates an instance of the TPopulation class and uses it to generate and list a (selectable) sample of mothers and the years in which they give birth.

Time dependent birth rates

The TFR parameter for future years can be input from file if known – or otherwise modelled. In this project the TFR parameter is kept constant overtime (Office for National Statistics 2012). In each year of their simulated life (y at age a), mothers of child bearing age can use the appropriate Poisson parameter $\lambda(a)p_b(a)$ to generate the number of children in that year. Each child is recorded in the mother's Life Event list and processed as part of the current family at the end of the mother's life.

Population dynamics

In some year, Y, the population will consist of N_m males and N_f females with their respective age distributions. In the next year, Y', the numbers will have been depleted by deaths and augmented by the $N_{newborn}$ births. The new, primed, population is determined from the old by the following equation set

$$N_{newborn} = \lambda N_f \sum_{a=AgeAtChild.lo}^{a=AgeAtChild.hi} p_f(a)\left(1-p_f(a)\right)p_b(a) \tag{1.14}$$

$$N'_m = N_m \sum_{a=1}^{a=Age.hi} p_m(a)\left(1-p_m(a)\right) + p_{male}N_{newborn} \tag{1.15}$$

$$N'_f = N_f \sum_{a=1}^{a=Age.hi} p_f(a)\left(1-p_f(a)\right) + p_{female}N_{newborn} \tag{1.16}$$

$$p'_m(a+1) = \frac{N_m}{N'_m} p_m(a)\left(1-p_m(a)\right) \tag{1.17}$$

$$p'_m(a+1) = \frac{N_m}{N'_m} p_m(a)\left(1-p_{\Omega m}(a)\right) \tag{1.18}$$

$$p'_f(a+1) = \frac{N_f}{N'_f} p_f(a)\left(1-p_{\Omega f}(a)\right) \tag{1.19}$$

$$p'_m(0) = \frac{1}{N'_m} p_{male}N_{newborn} \tag{1.20}$$

$$p'_f(0) = \frac{1}{N'_f} p_{female}N_{newborn} \tag{1.21}$$

The Population editor' menu item Population Editor\View\Population dynamics\male implements these equations and draws projected populations year by year.

Deaths from modelled diseases

The simulation models any number of specified diseases some of which may be fatal. In the start year the simulation's death model uses the diseases' own mortality statistics to adjust the

probabilities of death by age and gender. In the start year the net effect is to maintain the same probability of death by age and gender as before; in subsequent years, however, the rates at which people die from modelled diseases will change as modelled risk factors change. The population dynamics sketched above will be only an approximation to the simulated population's dynamics. The latter will be known only on completion of the simulation.

Immigration and emigration

The population module used in this method accounts for time dependent immigration and emigration of individuals. The immigration and emigration rates are based on the differences between the current and predicted populations. The immigrant population are generated sequentially as part of the national population.

Multiple population processing

Multiple populations can be used in a simulation provided they are non-overlapping (people cannot belong to both).

In a simulation, Monte Carlo trials are allocated between current different populations in proportion to their total person count (malesCount+femalesCount). The idea being to provide a representative sample of the combined population.
In a simulation, a population (pop) is current if the simulated year Y satisfies

$$pop \to startYear \le Y \le pop \to stopYear \tag{1.22}$$

Open populations

This model is an *open* population model which allows people to enter and to depart from the population (departure probability $p_\delta(t)$).

Open population, births and deaths

In the year y the number of males and females in the population are denoted as $\{N_m(a,y), N_f(a,y)\}$,

And we suppose that they have departure probabilities $\{p_{m\delta}(a,y), p_{f\delta}(a,y)\}$. The number of new arrivals into each age in the year Y are denoted $\{N_{mArr}(a,y), N_{fArr}(a,y)\}$.

The following analysis applies equally to males and females and we drop the gender suffix. The male and female populations grow according to the recursion relations

$$N(a+1,y+1) = N(a,y)\left(1-p_\Omega(a)\right)\left(1-p_\delta(a,y)\right) + N_{Arr}(a,y) \quad (a>1) \tag{1.23}$$

$$N(1,y+1) = N_{Newborn}(y)\left(1-p_\Omega(0)\right)\left(1-p_\delta(0,y)\right) + N_{Arr}(0,y) \quad (a=0) \tag{1.24}$$

The longitudinal modelling of populations having known cross sectional data

Given a set of X-sectional population projections $\{K_m(a,y), K_f(a,y)|0<=a<=100; Y_0<=y<=Y_1\}$ (the K- population) the question arises of how to model the lives of individuals within the population (the N-population). In the absence of precise arrival (immigration) and departure (emigration) statistics, many solutions exist. The population is constructed iteratively: given the population in year Y the next year' population is calculated from the known birth and death rates; the departure probabilities and arrival numbers are found by matching with the projected K-population.

Minimum arrival and departure model

The minimum arrival and departure model fixes the modelled N-population in the start year and compensates in subsequent years either by having non-zero departure statistics (if N>K) or by importing new people (K>N).
From equation (1.23):

$$if\ N(a,y)\left(1-p_\Omega(a)\right)>K(a+1,y+1)$$

$$\left(1-p_\delta(a,y)\right)=\frac{K(a+1,y+1)}{N(a,y)\left(1-p_\Omega(a)\right)}\quad(a>1)$$

$$\Rightarrow$$

$$N(a+1,y+1)=N(a,y)\left(1-p_\Omega(a)\right)\left(1-p_\delta(a,y)\right)=K(a+1,y+1)\quad(a>1)\qquad(1.25)$$

$$if\ N(a,y)\left(1-p_\Omega(a)\right)<K(a+1,y+1)$$

$$N_{Arr}(a,y)=K(a+1,y+1)-N(a,y)\left(1-p_\Omega(a)\right)\quad(a>1)$$

$$\Rightarrow$$

$$N(a+1,y+1)=N(a,y)\left(1-p_\Omega(a)\right)+N_{Arr}(a,y)=K(a+1,y+1)\qquad(1.26)$$

The implementation of this model can be arranged using multiple populations – one population for each year of the simulation. The first population consists of the base line model that matches the N and K populations in the start year; subsequent populations contain the corrections (the arrivals, if any in that year). When arrivals enter the simulated population they have a start year corresponding to this population's start year. They usually will have been modelled from birth in the appropriate risk and disease environment. Arrivals are ordinary members of the modelled population – they simply enter the population at times after the simulation-start time. Arrivals carry with them a population identifier.

The numbers of males and females and their ages are known for all populations. Within the micro simulation multiple populations are sampled at a rate proportional to their population size.

Risk factors

Risk factor model

The distribution of risk factors (RF) in the population is estimated using regression analysis stratified by both sex S = {male, female} and age group A = {0-9, 10-19, ..., 70-79, 80+}. The fitted trends are extrapolated to forecast the distribution of each RF category in the future. For each sex-and-age-group stratum, the set of cross-sectional, time-dependent, discrete distributions $D = \{p_k(t)|k = 1, \dots N;\ t > 0\}$, is used to manufacture RF trends for individual members of the population.

We model different risk factors, some of which are continuous (such as BMI) and some are categorical (smoking). Only smoking is described here.

Categorical risk factors

Smoking is the categorical risk factor. Each individual in the population may belong to one of the three possible smoking categories {*never smoked*, *ex-smoker*, *smoker*} with their probabilities $\{p_0, p_1, p_2\}$. These states are updated on receipt of the information that the person is either a smoker or a non-smoker. They will be a never-smoker or an ex-smoker depending on their original state (an ex-smoker can never become a never-smoker).
The complete set of longitudinal smoking trajectories and the probabilities of their happening is generated for the simulation years by allowing all possible transitions between smoking categories:

$$\{\textit{never smoked}\} \rightarrow \{\textit{never smoked}, \textit{smoker}\}$$
$$\{\textit{ex-smoker}\} \rightarrow \{\textit{ex-smoker}, \textit{smoker}\}$$
$$\{\textit{smoker}\} \rightarrow \{\textit{ex-smoker}, \textit{smoker}\}$$

When the probability of being a smoker is p the allowed transitions are summarised in the state update equation

$$\begin{bmatrix} p_0' \\ p_1' \\ p_2' \end{bmatrix} = \begin{bmatrix} 1-p & 0 & 0 \\ 0 & 1-p & 1-p \\ p & p & p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} \tag{1.27}$$

After the final simulation year the smoking trajectories are completed until the person's maximum possible age of 110 by supposing that their smoking state stays fixed. The life expectancy calculation will consists in summing over the probability of being alive in each possible year of life.

In the initial year of the simulation, a person may be in one of the three smoking categories; after N updates there will be 3×2^N possible trajectories. These trajectories will each have a calculated probability of occurring; the sum of these probabilities is 1.
In each year the probability of being a smoker or a non-smoker will depend on the forecast smoking scenario which provides exactly that information. Note that these states are two dimensional and cross-sectional {*non-smoking*, *smoking*}, and they are turned into three dimensional states {*never smoked*, *ex-smoker*, *smoker*} as described above. The time

evolution of the three dimensional states are the smoking trajectories necessary for the computation of disease table disease and death probabilities.

Input Data

Population assumptions

Figure 2.1 Population pyramid in 2015 in the UK.

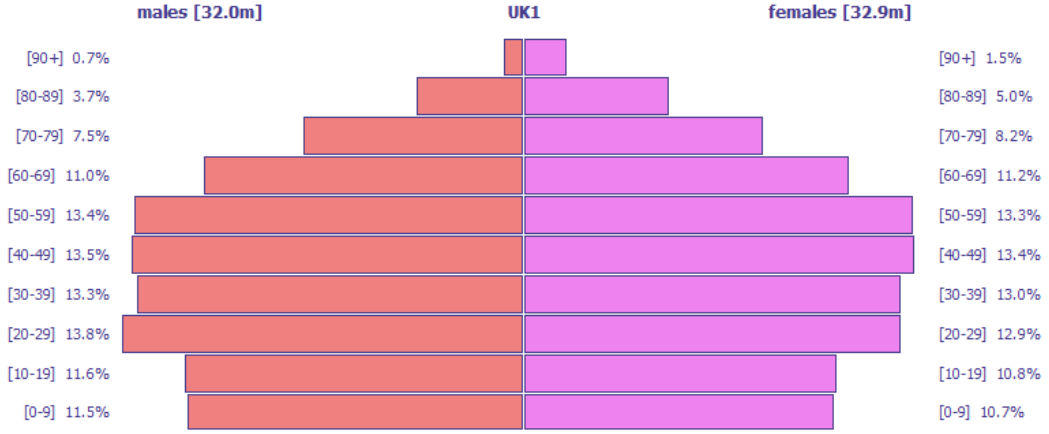


Figure 2.1 Population pyramid in 2015 in the UK

Smoking prevalence

The microsimulation framework applied to smoking enables us to measure the future health impact of changes in smoking prevalence. It includes the impact of giving up and not taking up smoking for the following diseases (chronic obstructive pulmonary disease, coronary heart disease and stroke) and cancers (acute myeloid leukaemia, chronic myeloid leukaemia, bladder, bowel, cervix, liver, lung, kidney, larynx, oesophagus, oral and pharynx, ovary, pancreas, stomach). The model is populated with economic data enabling analysis of future impact on healthcare costs.

In the simulation each person is categorised into one of the three smoking groups: smokers, ex-smokers and people who have never smoked. Projections for these three smoking groups have been created from the 2000 to 2012 General Lifestyle Survey datasets. The most recent dataset is shown in
Table 2. The initial distribution of smokers in the model is based on projected proportions of smokers, ex-smokers and never smokers in the start year of the simulation.

Table 2 Proportions of people who have never smoked, ex-smokers and smokers from the 2012 General Lifestyle Survey dataset

	Male			Female		
Age group	**Proportion never smokers**	**Proportion of ex-smokers**	**Proportion of smokers**	**Proportion never smokers**	**Proportion of ex-smokers**	**Proportion of smokers**
0-4	0.0	0.0	0.0	0.0	0.0	0.0
5-9	0.0	0.0	0.0	0.0	0.0	0.0
10-14	0.0	0.0	0.0	0.0	0.0	0.0

15-19	0.80669	0.0223	0.171	0.77255	0.01961	0.20784
20-24	0.59302	0.07752	0.32946	0.57713	0.09043	0.33245
25-29	0.53276	0.1339	0.33333	0.59319	0.12425	0.28257
30-34	0.53511	0.13801	0.32688	0.58681	0.17188	0.24132
35-39	0.5426	0.19955	0.25785	0.60304	0.16047	0.23649
40-44	0.53157	0.1833	0.28513	0.59304	0.17913	0.22783
45-49	0.54409	0.22326	0.23265	0.61191	0.14982	0.23827
50-54	0.51454	0.22595	0.25951	0.587	0.19503	0.21797
55-59	0.46042	0.30833	0.23125	0.55028	0.26376	0.18596
60-64	0.41586	0.38685	0.19729	0.60204	0.2466	0.15136
65-69	0.40319	0.4511	0.14571	0.51705	0.30303	0.17992
70-74	0.40642	0.47326	0.12032	0.62697	0.24944	0.1236
80+	0.42834	0.50082	0.07084	0.6629	0.25792	0.07919

During the simulation a person may change smoking states and their relative risk will change accordingly. Relative risks associated with smokers and people who have never smoked have been collected from published data. The relative risks associated with ex-smokers ($RR_{\text{ex-smoker}}$) are related to the relative risk of smokers (RR_{smoker}). The ex-smoker relative risks are assumed to decrease over time with the number of years since smoking cessation ($T_{\text{cessation}}$). These relative risks are computed in the model using equations (1.28) and (1.29)(1),

$$RR_{\text{ex-smoker}}(A,S,T_{\text{cessation}}) = 1 + (RR_{\text{smoker}}(A,S) - 1)\exp(-\gamma(A)T_{\text{cessation}}) \quad (1.28)$$

$$\gamma(A) = \gamma_0 \exp(-\eta A) \quad (1.29)$$

where γ is the regression coefficient of time dependency. The constants γ_0 and η are intercept and regression coefficient of age dependency, respectively, which are related to the specified disease (
Table 3).

Table 3 Parameter estimates for γ0 and η related to each disease (Hoogenveen et al. 2008).

Disease	γ_0	η
AMI (CHD)	0.24228	0.05822
Stroke	0.31947	0.01648
COPD	0.20333	0.03087
Lung cancer	0.15637	0.02065
Oesophagus cancer	0.0537424	0
Larynx cancer	0.0279918	0
Kidney cancer	0.0385957	0
Pancreas cancer	0.09279	0
Stomach cancer	0.0264112	0
Bladder cancer	0.05417	0
Oral cavity cancer	0.0493028	0

There are a number of smoking related diseases that will be modelled in this project that were not analysed in the above study (Hoogenveen et al. 2008). For consistency the ex-smoker RR's of the following diseases AML, CML, bowel, cervical, liver and ovarian cancer have been modelled in the same way by using the lung cancer coefficients (γ_0 and η) as a proxy.

Modelling diseases

Disease modelling relies heavily on the sets of incidence, mortality, survival, relative risk and prevalence statistics. The microsimulation uses risk dependent incidence statistics and these are inferred from the relative risk statistics and the distribution of the risk factor within the population. In the simulation, individuals are assigned a risk factor trajectory giving their personal risk factor history for each year of their lives. Their probability of getting a particular risk factor related disease in a particular year will depend on their risk factor state in that year. The necessary equations are given below.

Once a person has a fatal disease (or diseases) their probability of survival will be controlled by a combination of the disease-survival statistics and the probabilities of dying from other causes. Disease survival statistics are modelled as age and gender dependent exponential distributions.

Relative risks

Smoking is treated separately and in an identical fashion. The reported incidence risks for any disease do not make reference to any underlying risk factor. The microsimulation requires this dependence to be made manifest.

The risk factor dependence of disease incidence has to be inferred from the distribution of the risk factor in the population (here denoted as π); it is a disaggregation process:

Suppose that α is a risk factor state of some risk factor A and denote by $p_A(d|\alpha,a,s)$ the incidence probability for the disease d given the risk state, α, the person's age, a, and gender, s. The relative risk ρ_A is defined by equation (1.30).

$$\begin{aligned} p_{\mathrm{A}}\left(d\left|\alpha,a,s\right.\right) &= \rho_{\mathrm{A}|d}\left(\alpha\left|a,s\right.\right) p_{\mathrm{A}}\left(d\left|\alpha_0,a,s\right.\right) \\ \rho_{\mathrm{A}|d}\left(\alpha_0\left|a,s\right.\right) &\equiv 1 \end{aligned} \tag{1.30}$$

Where α_0 is the zero risk state (for example, the moderate state for alcohol consumption). The incidence probabilities, as reported, can be expressed in terms of the equation,

$$\begin{aligned} p\left(d\left|a,s\right.\right) &= \sum_{\alpha} p_{\mathrm{A}}\left(d\left|\alpha,a,s\right.\right)\pi_A\left(\alpha\left|a,s\right.\right) \\ &= p_A\left(d\left|\alpha_0,a,s\right.\right)\sum_{\alpha}\rho_{\mathrm{A}|d}\left(\alpha\left|a,s\right.\right)\pi_A\left(\alpha\left|a,s\right.\right) \end{aligned} \tag{1.31}$$

Combining these equations allows the conditional incidence probabilities to be written in terms of known quantities

$$p(d|\alpha,a,s)=\rho_{A|d}(\alpha|a,s)\frac{p(d|a,s)}{\sum_{\beta}\rho_{A|d}(\beta|a,s)\pi_{A}(\alpha|a,s)} \tag{1.32}$$

Previous to any series of Monte Carlo trials the microsimulation program pre-processes the set of diseases and stores the *calibrated* incidence statistics $p_A(d|\alpha_0, a, s)$.

Acquiring survival and mortality data predictions for a particular disease (*d*)

Published disease statistics are frequently incomplete and occasionally inconsistent. The microsimulation program makes use of a number of supporting methods to check and, as necessary, to supply missing disease statistics.

Approximating survival data from mortality and prevalence

An example is provided here with a standard life-table analysis for a disease *d*.
Consider the 4 following states:

state	Description
0	alive without disease *d*
1	alive with disease *d*
2	dead from disease *d*
3	dead from another disease

p_{ik} is the probability of disease *d* incidence, aged *k*
$p_{\omega k}$ is the probability of dying from the disease *d*, aged *k*
$p_{\bar{\omega}k}$ is the probability of dying other than from disease d, aged *k*

The state transition matrix is constructed as follows

$$\begin{bmatrix} p_0(k+1) \\ p_1(k+1) \\ p_2(k+1) \\ p_3(k+1) \end{bmatrix} = \begin{bmatrix} (1-p_{\bar{\omega}k})(1-p_{ik}) & (1-p_{\bar{\omega}k}-p_{\omega k})\,p_{\alpha k} & 0 & 0 \\ (1-p_{\bar{\omega}k})\,p_{ik} & (1-p_{\bar{\omega}k}-p_{\omega k})(1-p_{\alpha k}) & 0 & 0 \\ 0 & p_{\omega k} & 1 & 0 \\ p_{\bar{\omega}k} & p_{\bar{\omega}k} & 0 & 1 \end{bmatrix} \begin{bmatrix} p_0(k) \\ p_1(k) \\ p_2(k) \\ p_3(k) \end{bmatrix} \tag{1.33}$$

It is worth noting that the separate columns correctly sum to unity.

The disease mortality equation is that for state-2,

$$p_2(k+1)=p_{\omega k}p_1(k)+p_2(k) \tag{1.34}$$

The probability of dying from the disease in the age interval [*k*, *k*+1] is $p_{\omega k}p_1(k)$ - this is otherwise the (cross-sectional) disease mortality, $p_{mor}(k)$. $p_1(k)$ is otherwise known as the disease prevalence, $p_{pre}(k)$. Hence the relation

$$p_{\omega k}=\frac{p_{mor}(k)}{p_{pre}(k)} \tag{1.35}$$

For exponential survival probabilities the probability of dying from the disease in the age-interval [k, k+1] is denoted $p_{\Omega k}$ and is given by the formula

$$p_{\omega k} = 1 - e^{-R_k} \quad \Rightarrow \quad R_k = -\ln\left(1 - p_{\omega k}\right) \tag{1.36}$$

When, as is the case for most cancers, these survival probabilities are known the microsimulation will use them, when they are not known or are too old to be any longer of any use, the microsimulation uses survival statistics inferred from the prevalence and mortality statistics (equation (1.35)).
An alternative derivation equation (1.35) is as follows. Let N_k be the number of people in the population aged k and let n_k be the number of people in the population aged k with the disease. Then, the number of deaths from the disease of people aged k can be given in two ways: as $p_{\omega k} n_k$ and, equivalently, as $p_{mor}(k) N_k$. Observing that the disease prevalence is n_k/N_k leads to the equation

$$\begin{aligned} p_{\Omega k} n_k &= p_{mor}\left(k\right) N_k \\ p_{pre}\left(k\right) &= \frac{n_k}{N_k} \\ &\Rightarrow \\ p_{\Omega k} &= \frac{p_{mor}\left(k\right)}{p_{pre}\left(k\right)} \end{aligned} \tag{1.37}$$

Approximating survival data from mortality, incidence and remission data

We begin with the standard 1 year update equation and by defining some probabilities:

$p_i\left(a,Y\right)$	the incidence probability of the disease at age a
$p_r\left(a,Y\right)$	the remission probability of the disease at age a
$p_\omega\left(a,Y\right)$	the probability of dying from the disease at age a, in year Y
$p_{\bar{\omega}}\left(a,Y\right)$	the probability of dying from other causes at age a, in year Y

And the probabilities of being in a set of states:

S_0	$p_{\bar{d}}\left(a,Y\right)$	the probability of being alive without the disease at age a, in year Y
S_1	$p_d\left(a,Y\right)$	the probability of being alive with the disease at age a, in year Y
S_2	$p_\Omega\left(a,Y\right)$	the probability of being dead as a result of the disease at age a, in year Y
S_3	$p_{\bar{\Omega}}\left(a,Y\right)$	the probability of being dead from other causes at age a, in year Y

The update equation is (the dependence on the year Y is suppressed)

$$\begin{pmatrix} p_{\bar{d}}\left(a+1\right) \\ p_d\left(a+1\right) \\ p_\Omega\left(a+1\right) \\ p_{\bar{\Omega}}\left(a+1\right) \end{pmatrix} = \begin{pmatrix} \left(1-p_{\bar{\omega}}\right)\left(1-p_i\right) & \left(1-p_{\bar{\omega}}-p_\omega\right)p_r & 0 & 0 \\ \left(1-p_{\bar{\omega}}\right)p_i & \left(1-p_{\bar{\omega}}-p_\omega\right)\left(1-p_r\right) & 0 & 0 \\ 0 & p_\omega & 1 & 0 \\ p_{\bar{\omega}} & p_{\bar{\omega}} & 0 & 1 \end{pmatrix} \begin{pmatrix} p_{\bar{d}}\left(a\right) \\ p_d\left(a\right) \\ p_\Omega\left(a\right) \\ p_{\bar{\Omega}}\left(a\right) \end{pmatrix} \tag{1.38}$$

Survival

At some age, a_0, the person is alive and gets the disease – at this age the state vector is, $(0 \quad 1 \quad 0 \quad 0)$.

If we assume the remission probability is zero the person's subsequent life is governed by the equation

$$\begin{pmatrix} p_d(a+1) \\ p_\Omega(a+1) \\ p_{\bar{\Omega}}(a+1) \end{pmatrix} = \begin{pmatrix} 1-p_{\bar{\omega}}-p_\omega & 0 & 0 \\ p_\omega & 1 & 0 \\ p_{\bar{\omega}} & 0 & 1 \end{pmatrix} \begin{pmatrix} p_d(a) \\ p_\Omega(a) \\ p_{\bar{\Omega}}(a) \end{pmatrix} \tag{1.39}$$

At age $a = a_0 + N$ it has the solution

$$p_d(a_0+N) = \prod_{k=1}^{k=N} \left(1-p_\omega(a_k)-p_{\bar{\omega}}(a_k)\right) \tag{1.40}$$

Disease survival probabilities

Disease survival statistics are gathered from those people who do not die from other causes. The probability of surviving N years, given that there is no remission, and that there is no probability of death from other causes is simply

$$p_d(a_0+N) = \prod_{k=1}^{k=N} \left(1-p_\omega(a_k)\right) \tag{1.41}$$

These are longitudinal statistics that, ideally, are gathered by following the life courses of many people who have the disease.

In equation (1.41) it is understood that the disease is contracted at age a_0 and that the death probabilities are the successive probabilities of dying from the disease in the first year - $p_\omega(a_0+1)$, the second year - $p_\omega(a_0+2)$, and so on. These are *disease survival statistics*, closely connected to but not the same as *disease mortality statistics*.

Mortality statistics

In any year, in some population, in a sample of N people who have the disease a subset N_ω will die from the disease.
Mortality statistics record the cross sectional probabilities of death as a result of the disease – possibly stratifying by age

$$p_\omega = \frac{N_\omega}{N} \tag{1.42}$$

Within some such subset N_ω of people that die in that year from the disease, the distribution by year-of-disease is not usually recorded. This distribution would be most useful. Consider two important idealised, special cases

Suppose the true probabilities of dying in the years after some age a_0 are
$\{p_{\omega 0}, p_{\omega 1}, p_{\omega 2}, p_{\omega 3}, p_{\omega 4}\}$

The probability of being alive after N years is simply that you don't die in each year

$$p_{survive}(a_0 + N) = (1 - p_{\omega 0})(1 - p_{\omega 1})(1 - p_{\omega 2})..(1 - p_{\omega N-1}) \tag{1.43}$$

Different survival models

There are three in use (they are easily extended if the data merit):

Survival model 0: a single probability of dying $\{p_{\omega 0}\}$

$p_{\omega 0}$ is valid for all years

Survival model 1: two different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}\}$

$p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ thereafter.

Survival model 2: three different probabilities of dying $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 5}\}$

$p_{\omega 0}$ is valid for the first year; $p_{\omega 1}$ for the second to the fifth year; $p_{\omega 5}$ thereafter

Remember that different probabilities will apply to different age and gender groups. Typically the data might be divided into 10 year age groups.

Calculating survival from incidence and mortality

When a person (of a given gender) dies from a disease they must have contracted it at some earlier age. For Survival model 2, this is expressed

$$\begin{aligned}\hat{p}_{mortality}(a) = {} & p_{inc}(a-1)\,p_{\omega 0} + \\ & + p_{inc}(a-2)(1-p_{\omega 0})\,p_{\omega 1} + \\ & + p_{inc}(a-3)(1-p_{\omega 0})(1-p_{\omega 1})\,p_{\omega 1} + \\ & + p_{inc}(a-4)(1-p_{\omega 0})(1-p_{\omega 1})^2\,p_{\omega 1} + \\ & + p_{inc}(a-5)(1-p_{\omega 0})(1-p_{\omega 1})^3\,p_{\omega 1} + \\ & + p_{inc}(a-6)(1-p_{\omega 0})(1-p_{\omega 1})^4\,p_{\omega 5} + \\ & + p_{inc}(a-7)(1-p_{\omega 0})(1-p_{\omega 1})^4(1-p_{\omega 5})\,p_{\omega 5} + \\ & + \ldots\end{aligned} \tag{1.44}$$

The three probabilities $\{p_{\omega 0}, p_{\omega 1}, p_{\omega 5}\}$ are estimated by minimising

$$S = \sum_{a \in AgeGroup} \frac{\left(\bar{p}_{mortality}(a) - \hat{p}_{mortality}(a)\right)^2}{\bar{\sigma}^2} \tag{1.45}$$

When the longitudinal probability of the disease incidence at age a satisfies the recursion relation

$$p_{inc}(a) = (1 - p_i(0))(1 - p_i(1))..(1 - p_i(a-1))\, p_i(a) \tag{1.46}$$

Survival Statistics CRUK 2010/11

If unavailable, survival data has been approximated from incidence and mortality data. The calculated survival data has been validated by using this data to estimate the mortality data. Figure 2-5 are plots which show the rates of incidence and mortality disease data (collected) along with the rates of mortality calculated from survival (predicted) for increasing age. The figures demonstrate the similarity between the two different mortality datasets for CHD and oral and pharynx cancer. These graphs were used to visually assess how close the fit was between the predicted and collected mortality data.

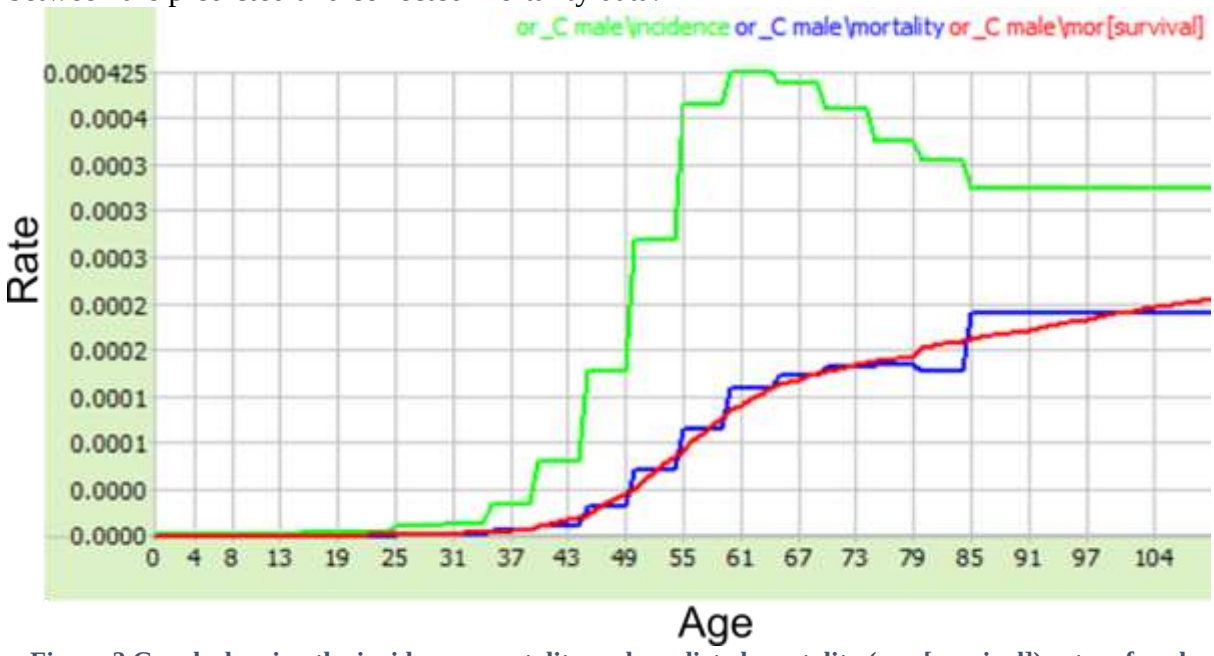


Figure 2 Graph showing the incidence, mortality and predicted mortality (mor[survival]) rates of oral and pharynx cancer for males at different ages

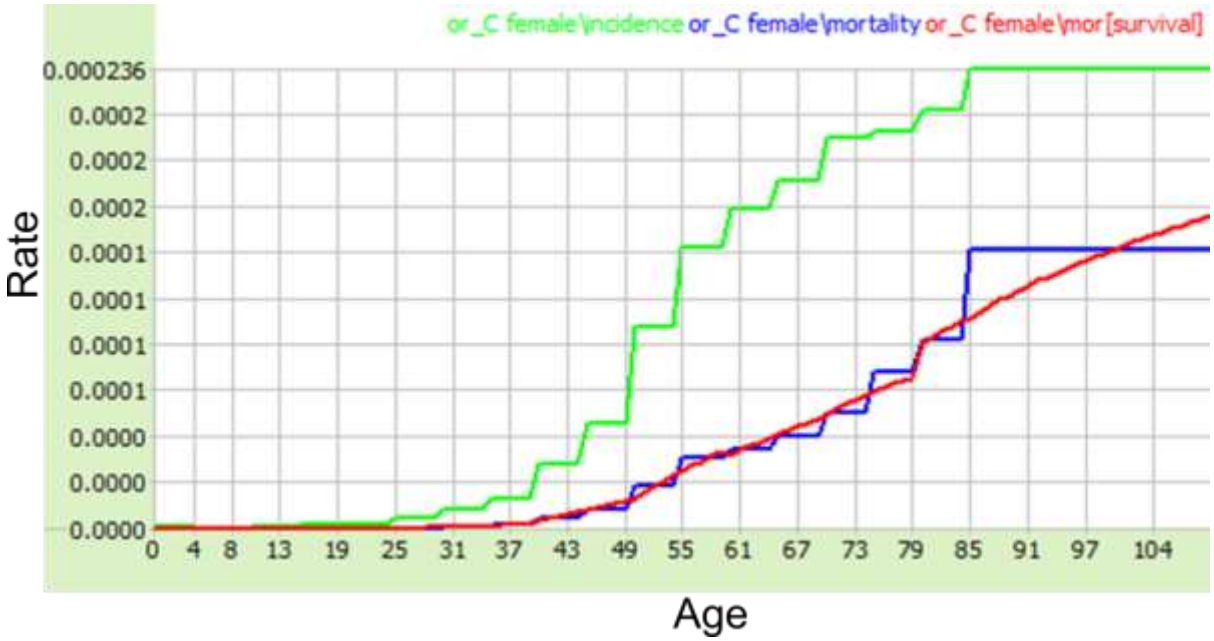


Figure 3 Graph showing the incidence, mortality and predicted mortality (mor[survival]) rates of oral and pharynx cancer for females at different ages

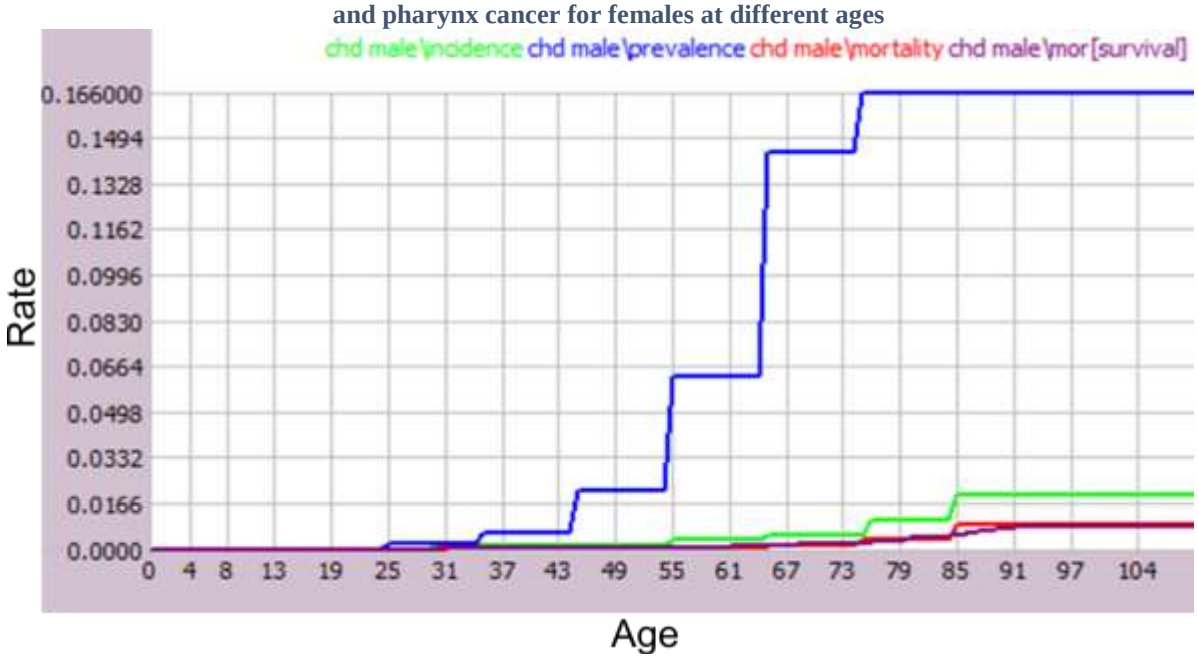


Figure 4 Graph showing the incidence, mortality and predicted mortality (mor[survival]) rates of CHD for males at different ages

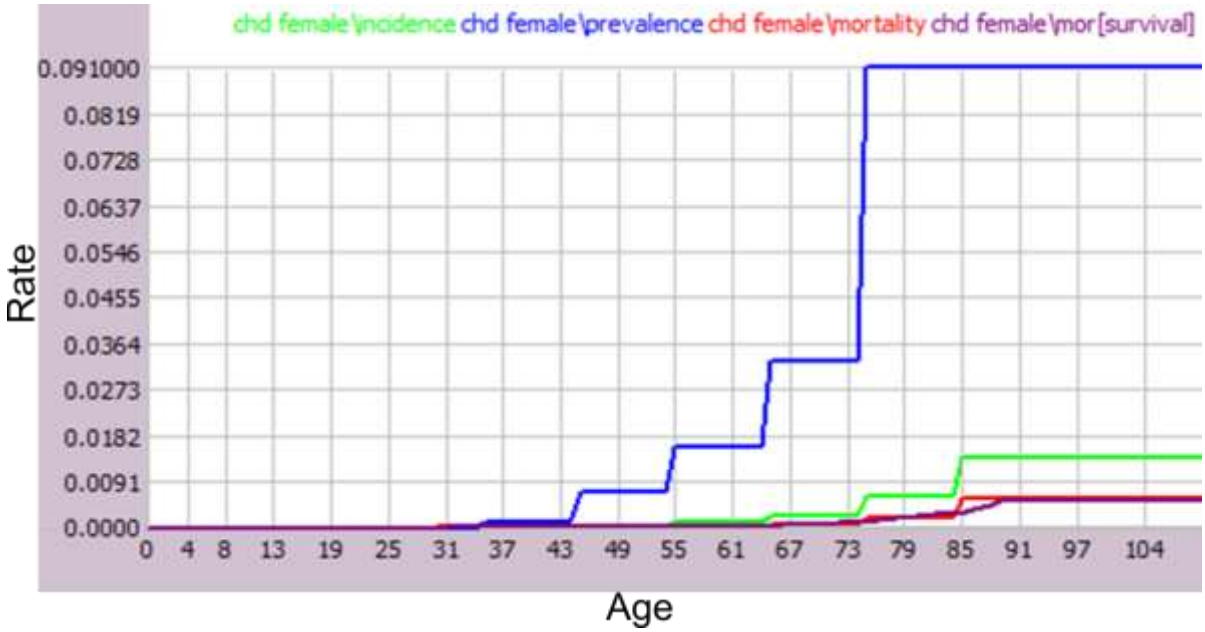


Figure 5 Graph showing the incidence, mortality and predicted mortality (mor[survival]) rates of CHD for females at different ages

The probabilities of being alive after 1, 5 and 10 years are

$$\begin{aligned} p_{survival}(a_0+1) &= (1-p_{\omega 0}) \\ p_{survival}(a_0+5) &= (1-p_{\omega 0})(1-p_{\omega 1})^4 \\ p_{survival}(a_0+10) &= (1-p_{\omega 0})(1-p_{\omega 1})^4(1-p_{\omega 5})^5 \end{aligned} \tag{1.47}$$

Rates

It is common practice to describe survival in terms of a survival rate R, supposing an exponential death-distribution. In this formulation the probability of surviving t years from some time t_0 is given as

$$p_{\text{survival}}(t) = 1 - R^{-1}\int_0^t due^{-Ru} = e^{-Rt} \tag{1.48}$$

For a time period of 1 year

$$\begin{aligned} p_{survival}(1) &= e^{-R} \\ &\Rightarrow \\ R &= -\ln\left(p_{survival}(1)\right) = -\ln(1-p_{\omega}) \end{aligned} \tag{1.49}$$

For a time period of, for example, 4 years,

$$p_{\text{survival}}(t=4) = 1 - R^{-1}\int_0^4 due^{-Ru} = e^{-4R} = (1-p_{\omega})^4 \tag{1.50}$$

In short, the Rate is minus the natural log of the 1-year survival probability.

Survival models 0, 1 and 2

For any potentially terminal disease the model can use any of three survival models, numbered {0, 1, 2}. The parameters describing these models are given below. In this study if the survival rates were unavailable the survival rates were calculated with survival model 2.

Survival model 0

Given the 1-year survival probability $p_{survival}(1)$
The model uses 1 parameter {R}

$$R = -\ln\left(p_{survival}(1)\right) \tag{1.51}$$

Survival model 1

The model uses two parameters {p_1, R}
Given the 1-year survival probability $p_{survival}(1)$ and the 5-year survival probability $p_{survival}(5)$

$$p_1 = 1 - p_{survival}(1)$$
$$R = -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \tag{1.52}$$

Survival model 2

The model uses three parameters {p_1, R, $R_{>5}$}
Given the 1-year survival probability $p_{survival}(1)$ and the 5-year survival probability $p_{survival}(5)$

$$p_1 = 1 - p_{survival}(1)$$
$$R = -\frac{1}{4}\ln\left(\frac{p_{survival}(5)}{p_{survival}(1)}\right) \tag{1.53}$$
$$R_{>5} = -\frac{1}{5}\ln\left(\frac{p_{survival}(10)}{p_{survival}(5)}\right)$$

Modelling costs

Direct costs

The cost model used in the simulation is part of the economics module and, here, simply scales the aggregated individual disease costs according to the relative disease prevalence in years after the start year for which the costs are known.

In any year, the total healthcare cost for the disease *D* is denoted $C_D(year)$. If the prevalence of the disease is denoted $P_D(year)$ we assume a simple relationship between the two of the form

$$C_D(year) = \kappa P_D(year) \tag{1.54}$$

for some constant κ.
For each of the trial years, the microsimulation records the prevalence of each disease call it $P_D(year|trial)$ and the trial population size for that year, $N_{pop}(year|trial)$. Further assume that the prevalence in the whole population $N_{pop}(year)$ is a simple scaling of the trial prevalence, then

$$C_D(year) = \kappa P_D(year) = \lambda \frac{N_{pop}(year) P_D(year|trial)}{N_{pop}(year|trial)} \tag{1.55}$$

for some constant λ.
By comparing any trial year to some initial year, *year0*, the total disease cost in any year is given as

$$\frac{C_D(year)}{C_D(year0)} = \frac{N_{pop}(year)}{N_{pop}(year0)} \frac{N_{pop}(year0|trial)}{N_{pop}(year|trial)} \frac{P_D(year|trial)}{P_D(year0|trial)} \tag{1.56}$$

The same method is applied for total NHS social care costs if they are available for a specific disease.

Non-health costs

Non-health costs were based on a human-capital approach. Two parts, 'mort costs' and 'morb costs', form the basis of non-health cost estimations:

Non-health cost = 'mort cost' + 'morb cost'

<u>Approach to modelling mortality costs</u>:

Mort cost' = summation of gross annual income, from age of death to 65 years old

'Mort cost' is a function of:

- Age
- Sex
- Year of death

'Mort cost' is independent of the cause of death

Data source
Gross income: ONS (2013) Distribution of income, by age and sex

Assumptions

- Individuals are economically active between 16-65
- Social value of paid work is equal (on average) to the total cost of employment
- Data are average figures (i.e. top-down costing approach)

Approach to modelling morbidity costs:
Morbcost represents the potential net income that an individual would have earned had the individual not taken time off of work (sickness absence) due to morbidity. The morbcost of an individual is conservatively estimated by multiplying the net annual income by the productivity rate and the on-cost, which represents overheads associated with employment (constant variable at 30% for the UK).

'Morb cost' = full health gross annual income x (1- productivity) x on-costs
Tax; net income

'Morb cost' is a function of:

- Age
- Sex
- QoL
- Year of disease onset

Data source
Productivity figures: Understanding Society Survey Gross income: Annual Survey of Hours and Earnings On-costs: Eurostat 2012

Paid production is estimated by

- Estimating the *productivity* of the patient – the amount of possible working time they actually spend working – given their age and QoL, using a model based on data from the Understanding Society dataset
- Multiplying this by their gross *wages* if in work, which is estimated using the Annual Survey of Hours and Employment
- Applying an uplift ("*on costs"*) to reflect the overhead costs of their employment

Limitations of the non-health cost method within the microsimulation

- Discounting

The costs are currently not discounted since discounting for mortcost is difficult to calculate since future earnings will need to be discounted prior to summing all the future year earnings. Future work should incorporate this into the model.

- Multi-morbidities

If an individual has multiple diseases it is assumed their quality of life is calculated from the product of the quality of life of each disease estimating the individual's productivity level ('maximum limit approach'). For example, if an individual has CHD (utility weight value of 0.7) and kidney cancer (utility weight value of 0.6), the product of these utility weights are used to estimate the productivity, which in turn is used to calculate the morbcost.

When an individual in the microsimulation has more than one disease, the utility weight of the more debilitating disease is selected for estimating the individual's productivity level ('maximum limit approach') e.g. if an individual has CHD (utility weight value of 0.7) and kidney cancer (utility weight value of 0.6), 0.6 is taken as the overriding utility weight value to estimate the productivity, which in turn is used to calculate the morbcost.

- Productivity impact curve

At this stage, we have included only the productivity impact curve of the full societal costing method, without inclusion of consumption costs. This is because a comprehensive list of the average number of sickness days that patients with various ailments take off in a given year was not available to us. Further, because we modelled many diseases (mostly rare cancers), it was beyond the scope of this study to conduct expert witness studies with specialists/clinicians in order to gain a better idea of the impact that being sick from a particular illness has on absenteeism. Instead, the average utility weight that is associated with a particular disease was used to read off the productivity difference that occurs between a healthy person and that of a person with the illness.

Acknowledgement: Non-health costs were incorporated into the model based on a toolkit in development by analysts in the Department of Health to which we were given early access by Gavin Roberts (2)

SUPPLEMENTARY MATERIAL 2. ICD codes by disease

Diseases	Incidence	Prevalence	Mortality	Survival
CHD	I21-I22	I21-I22	I20-I25	Computed from UKHF model
COPD	Computed from UKHF model	J40-J44	J40-J44	J40-J44
Stroke	I60-I69	N/A	I60-I69	Computed from UKHF model
AML	C92.0, C92.4. C92.5, C93.0, C94.0, C94.2	N/A	C92.0, C92.4. C92.5, C93.0, C94.0, C94.2	Computed from UKHF model
Bladder cancer	C67	N/A	C67	C67
Bowel cancer	C18-20, C21.8	N/A	C18-20, C21.8	Computed from UKHF model
Cervical cancer	C53	N/A	C53	C53
CML	C92.1	N/A	C92.1	Computed from UKHF model
Kidney cancer	C64-C66, C68	N/A	C64-C66, C68	C64 - C66, C68
Larynx cancer	C32	N/A	C32	C32
Liver cancer	C22	N/A	C22	C22
Lung cancer	C33-C34	N/A	C33-C34	C33, C34
Oesophagus cancer	C15	N/A	C15	C15
Oral cancer	C00-C06, C09, C10, C120-C14	N/A	C00-C06, C09, C10, C120-C14	Computed from UKHF model
Ovarian cancer	C56-C57	N/A	C56-C57	C56, C57.0 - C57.7
Pancreatic cancer	C25	N/A	C25	C25
Stomach cancer	C16	N/A	C16	C16

SUPPLEMENTARY MATERIAL 3. Data input reference table

Diseases	Incidence	Prevalence	Survival	Mortality	Costs	Utility weights	Relative risk: smoking
CHD	BHF, 2010 (with BMJ correction)	BHF, 2006	Computed from prevalence and mortality	BHF, 2010 (with BMJ correction)	NHS England	NICE	DYNAMO-HIA, 2014[3]
COPD		PHE	Wildman et al, 2009	ONS, 2010	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Stroke	BHF, 2009	N/A	Computed from prevalence and mortality	BHF, 2010	NHS England	NICE	DYNAMO-HIA, 2014[3]
AML	CRUK, 2011	N/A	Computed from UKHF model	CRUK, 2012	NHS England	Sullivan et al, 2011	Fircanis et al, 2014[4]
Bladder cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Bowel cancer	CRUK, 2011	N/A	computed from UKHF model	CRUK, 2012	NHS England	Sullivan et al, 2011	Botteri et al, 2008(5)
Cervical cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	ICESCC, 2006[6]
CML	CRUK, 2011	N/A	computed from UKHF model	CRUK, 2012	NHS England	Sullivan et al, 2011	Musselman et al, 2013[7]
Kidney cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	Hunt et al, 2005[8]
Larynx cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Liver cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	Lee et al, 2009[9]
Lung cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Oesophagus cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Oral cancer	CRUK, 2011	N/A	computed from UKHF model	CRUK, 2012	NHS England	Sullivan et al, 2011	DYNAMO-HIA, 2014[3]
Ovarian cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	Whiteman et al,

							2006[10]
Pancreatic cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Romanus et al, 2012	DYNAMO-HIA, 2014[3]
Stomach cancer	CRUK, 2011	N/A	ONS, 2012	CRUK, 2012	NHS England	Sullivan et al, 2011	Ladeiras-Lopes et al, 2008[11]

SUPPLEMENTARY FILE 4. Data inputs

Table 1.1 Incidence rates
Table 1.2 Mortality rates
Table 1.3 Survival rates
Table 1.4 Smoking relative risks
Table 1.5 Smoking direct costs
Table 1.6 Smoking utility weights
Table 1.7 UK income distribution

Table 1.1 Incidence rate per 100,000 population

AML																	
Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	1.06	0.67	0.63	0.72	0.94	0.84	1.05	1.45	1.78	2.27	3.86	5.17	8.01	12.99	19.69	25.97	33.97	35.75
females	1.06	0.43	0.65	0.65	1.05	0.99	1.14	1.07	1.79	2.06	3.03	4.2	5.39	8.27	11.47	15.72	19.25	22.83

Bladder cancer																	
Age group	0-4	5-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84	
males	0.2	0	0.1	0.2	0.5	0.9	1.7	4.6	9.5	19.9	37.3	70.5	108.8	153.5	211.3	264.3	
females	0.1	0	0	0.2	0.2	0.6	1	1.7	3.3	7	11.3	18.8	31	44.4	59.7	76.3	

Cervical cancer																	
Age group	0-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84	
males	0	0.1	2.9	19.1	19.7	19.1	15.7	12.1	10.3	10.4	9.1	8.6	9.9	11.8	12.5	12.4	
females	0	0.1	2.9	19.1	19.7	19.1	15.7	12.1	10.3	10.4	9.1	8.6	9.9	11.8	12.5	12.4	

COPD																	
Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	>80
males	1.2	2.6	2.1	0	1.8	3.4	7.9	25	52	106	205.2	298	491.9	437.5	334.8	263.4	0
females	0	3.4	1.6	0	1.3	3.4	9.8	28.2	60.4	121.2	174.5	235.4	326.9	216.8	228	0	0

CML																	
Age group	0-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.2	0.5	0.5	0.8	0.9	1	1.4	1.4	1.7	2.1	2.9	3.6	4.3	5.2	8.3
females	0	0.1	0.3	0.3	0.5	0.3	0.6	0.8	0.6	1.1	1.2	1.3	1.4	2	2.5	3.4	4.3

Colorectal cancer																		
Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0	0.3	0.5	1	2.5	4.7	5.6	11.3	23.1	46.8	85.2	166	252.9	328.3	416.3	491.4	518.6
females	0	0.1	0.3	0.7	1.3	2.6	3.9	5.1	11.2	20.3	36.4	57.9	99.8	147.8	194	257.8	315.1	334.3

CHD							
Age group	0-29	30-54	55-64	65-74	75-75	76-84	>84
males	0	88.1	317	533	533	1017	1987
females	0	21.2	90.3	237	597	597	1395

Kidney cancer																	
Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	1.8	0.3	0.1	0.2	0.3	0.5	1.4	3.8	7.4	11.5	20.4	30.4	46.1	57.1	79.7	95.1	106.2	104.3
females	1.9	0.3	0.1	0.1	0.3	0.5	1.1	2	3.9	5.7	11.2	16.7	23.8	29	41.3	50.7	51.5	51.2

Larynx cancer													
Age group	0-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0	0.5	1.3	3.1	6.8	13.6	17.6	23.1	26.2	25.2	24	23.5
females	0	0.1	0.2	0.4	0.8	1.6	2.4	3.2	4.5	4.7	4.2	4.4	3.2

Table 1.1 (continued)

Liver cancer

Age group	0-4	5-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.5	0.1	0.1	0.3	0.5	0.8	1.7	3.8	7.4	13.8	17.4	27.3	37.7	48	55.5	57.8
females	0.5	0.1	0.2	0.2	0.4	0.5	0.9	1.4	3	4.6	7.4	10.1	17.3	23.4	28.9	33.4

Lung cancer

Age group	0-4	5-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0	0.2	0.3	0.6	1.1	2.7	6.8	15.6	37.8	87.3	148.2	251.5	369.4	463.5	565.2	582.3
females	0.1	0	0.3	0.3	0.5	1	2.2	6	15.9	35.7	70.8	117.5	182.1	239.3	282.2	323.3	274.6

Oesophageal cancer

Age group	0-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.2	0.3	1.1	3.1	6.8	15.8	28	44.1	59.1	75.5	88.9	110.3	116.8
females	0	0	0.1	0.3	0.8	2	4.6	9.1	13.5	19.8	28.8	39.3	53.7	65.6

Oropharyngeal cancer

Age group	0-4	5-9	10-14	15-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.1	0.1	0.1	0.2	0.8	1.1	2.8	6.7	15	27	39.5	42.5	41.5	39	36.1	34.4	31.7
females	0.1	0	0.1	0.2	0.5	0.9	1.5	3.2	5.3	10.3	14.4	16.4	17.8	20	20.3	21.5	23.6

Ovarian cancer

Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.1	0.2	0.6	1.4	3.1	4.3	5.6	7.9	12.5	18.5	26.9	35.3	46.8	57.9	67.5	69	72.6	70.3
females	0.1	0.2	0.6	1.4	3.1	4.3	5.6	7.9	12.5	18.5	26.9	35.3	46.8	57.9	67.5	69	72.6	70.3

Pancreatic cancer

Age group	0-14	15-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.1	0.3	0.9	2.2	5	9.2	18.5	28.6	43.2	58.9	78.4	94.9	109.4
females	0	0.1	0.2	0.3	0.7	1.9	3.3	7.3	12.5	21.6	34.1	49.7	64	82.2	92.6

Stomach cancer

Age group	0-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.1	0.2	0.7	1.3	3.1	5.2	9	15.8	23.4	41.7	69.2	97.5	125.4	136.8
females	0	0.1	0.2	0.4	0.8	0.7	1.9	3.1	4.3	6.1	9.2	14.4	27	40	53.1	61.6

Stroke

Age group	0-44	45-64	65-74	>74
males	7	114	393	794
females	6	69	275	879

Table 1.2 Mortality rate per 100,000 population

AML																		
Age group	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.2	0.1	0.1	0.4	0.3	0.4	0.4	0.7	0.6	1	1.8	3.3	6.6	11.8	19.7	27	35.6	38.6
females	0.3	0.2	0.1	0.2	0.3	0.2	0.4	0.5	0.6	1.2	1.5	2.7	3.9	6.8	11.9	16.4	20	23.8

Bladder cancer													
Age group	0-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.1	0.3	1	2.2	5.5	10.7	21.3	36.9	66	114.5	210.8
females	0	0.1	0.2	0.5	0.7	1.2	2.4	3.8	7.5	13.6	20.9	37.8	66.3

Cervical cancer														
Age group	0-19	20-24	25-29	30-34	35-39	40-44	45-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.3	1.3	1.7	2.5	2.9	3.6	3.7	4.1	5	6.5	8.1	9.3	11.9
females	0	0.3	1.3	1.7	2.5	2.9	3.6	3.7	4.1	5	6.5	8.1	9.3	11.9

COPD											
Age group	0	1-4	5-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	>84
males	0	0	0.1	0	0	0.7	4.5	32.9	121.7	365.2	829.4
females	0.3	0	0	0	0.1	0.5	4	26.4	90.3	257.3	423.1

CML													
Age group	0-24	25-29	30-34	35-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0	0.1	0.1	0.2	0.2	0.5	0.8	1.4	1.6	4.2	6.3
females	0	0	0	0	0.1	0.1	0.2	0.4	0.3	0.7	1	2.3	4.2

Colorectal cancer															
Age group	0-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.2	0.4	1.1	1.6	2.9	6.7	12.8	26.1	43.3	68.1	110.5	163.6	233.2	345
females	0	0.2	0.6	1.1	1.2	2.7	5.3	9.9	16	25.1	39.2	61.2	95.3	147.5	233.6

CHD						
Age group	0-29	30-54	55-64	65-74	75-84	>84
males	0	12.3	57.8	137	347	848
females	0	2.7	14.9	54.6	190	574

Kidney cancer																	
Age group	0-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.1	0	0	0	0	0.2	0.5	1.6	3.2	5.4	10.2	15.6	21.4	32.8	42.6	62.8	80.4
females	0.1	0	0.1	0	0.1	0.1	0.3	0.6	1.3	2.3	4.5	7.5	9.6	16.2	23	29.7	40.6

Larynx cancer											
Age group	0-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.2	0.5	1.3	3	4.5	6.9	8.6	9.8	10.3	17.7
females	0	0.1	0.1	0.3	0.7	0.9	1.4	1.7	2	2.6	3.3

Table 1.2 (continued)

Liver cancer

Age group	0-4	5-14	15-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0.1	0	0	0.2	0.4	0.6	1.1	2.4	5.2	10.3	14.2	22.8	34.8	45.9	58.1	58
females	0.1	0	0.1	0.1	0.2	0.3	0.7	1.3	2.3	5	7.3	10.6	18.3	25.6	31.9	38.2

Lung cancer

Age group	0-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.1	0.5	1.7	4.5	10.7	27.3	62.3	113.7	188.1	292.5	383.9	479.1	559.4
females	0	0	0.1	0.5	1.3	3.4	9.6	23.7	51.4	84.6	134.5	189.7	230.9	287.5	271

Oesophageal cancer

Age group	0-24	25-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.2	0.6	2.4	5.2	12.4	21.7	36.1	50.4	66.8	86.1	115.5	134.5
females	0	0.1	0.2	0.5	1.2	3.3	5.7	9.7	14.2	23.4	35.4	53.3	71.6

Oropharyngeal cancer

Age group	0-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.1	0.4	0.8	2.6	6	9.7	13.4	14.6	15.5	15.6	15	20.3
females	0	0	0.1	0.2	0.5	0.9	2.2	3.6	4	4.7	5.9	8	9.6	14.3

Ovarian cancer

Age group	0-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0.1	0.2	0.3	0.7	1.2	2.3	5.3	9.7	15.6	26	38.4	45.9	58.8	63.8	67
females	0	0.1	0.2	0.3	0.7	1.2	2.3	5.3	9.7	15.6	26	38.4	45.9	58.8	63.8	67

Table 1.2 (continued)

Pancreatic cancer

Age group	0-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0	0.2	0.6	1.7	3.9	7.8	15.6	25.9	39.4	57.2	75.1	93.8	114.8
females	0	0.1	0.1	0.5	1.2	2.5	5.9	10.3	19.9	30.2	46.1	60	80.5	97

Stomach cancer

Age group	0-14	15-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	>84
males	0	0	0.2	0.4	0.8	1.5	2.4	4.2	7.4	11.8	21.1	41.1	63.1	91.8	119.8
females	0	0.1	0.3	0.5	0.4	1.1	1.7	2.2	3.5	5	7.9	16.6	26.2	39.7	55

Stroke

Age group	0	1-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-89	>89
males	1.44	0	0	0.07	0.42	0.55	0.6	2.18	3.45	6.02	8.76	13.72	24.62	35.63	56.42	127.24	257.84	541.58	1009.9	1925.1
females	1.82	0.3	0.06	0.14	0.44	0.34	0.87	1.35	1.9	4.18	6.95	10.91	17.34	23.34	40.99	92.88	221.69	486.07	1024.7	2198.1

Table 1.3 Survival rate (%)

AML						
Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.84	0.86	0.84	0.76	0.63	0.47
female (1 yr survival)	0.85	0.83	0.82	0.75	0.61	0.44
male (5 yr survival)	0.69	0.74	0.69	0.56	0.43	0.28
female (5 yr survival)	0.7	0.7	0.68	0.56	0.43	0.23
Bladder cancer						
Age group	15-49	50-59	60-69	70-79	>79	
male (1 yr survival)	0.86	0.86	0.84	0.78	0.65	
female (1 yr survival)	0.67	0.77	0.75	0.67	0.5	
male (5 yr survival)	0.73	0.69	0.67	0.58	0.42	
female (5 yr survival)	0.48	0.57	0.56	0.48	0.3	
Cervical cancer						
Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0	0	0	0	0	0
female (1 yr survival)	0.97	0.92	0.87	0.77	0.67	0.5
male (5 yr survival)	0	0	0	0	0	0
female (5 yr survival)	0.9	0.79	0.68	0.54	0.42	0.27
Colorectal cancer						
Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.85	0.85	0.84	0.85	0.77	0.62
female (1 yr survival)	0.88	0.87	0.86	0.84	0.76	0.57
male (5 yr survival)	0.67	0.63	0.62	0.68	0.58	0.44
female (5 yr survival)	0.7	0.65	0.65	0.68	0.59	0.42
COPD						
Age group	0-19	20-39	40-59	60-79	>79	
male (1 yr survival)	0.9844	0.9922	0.9961	1	1	
female (1 yr survival)	1	0.9922	1	1	1	
male (5 yr survival)	0.9653	0.9653	0.9922	1	1	
female (5 yr survival)	1	0.9807	1	1	1	
male (10 yr survival)	0.9653	0.9653	0.899	0.874	0.8398	
female (10 yr survival)	1	0.9764	0.9242	0.8919	0.884	
CML						
Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.84	0.86	0.84	0.76	0.63	0.47
female (1 yr survival)	0.85	0.83	0.82	0.75	0.61	0.44
male (5 yr survival)	0.69	0.74	0.69	0.56	0.43	0.28
female (5 yr survival)	0.7	0.7	0.68	0.56	0.43	0.23

Table 1.3 (continued)
CHD

Age group	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	>89
male (1 yr survival)	1	1	1	0.9375	0.9766	0.9375	0.9727	0.8904	0.8125	0.875
female (1 yr survival)	1	1	1	0.9375	0.9687	0.9375	0.9141	0.9062	0.8437	0.875
male (5 yr survival)	1	1	1	0.8515	0.8882	0.8526	0.9726	0.8904	0.8125	0.632
female (5 yr survival)	1	1	1	0.8655	0.9241	0.8642	0.914	0.9062	0.8437	0.5902
male (10 yr survival)	1	1	1	0.8514	0.8739	0.8054	0.8044	0.7176	0.5402	0.5621
female (10 yr survival)	1	1	1	0.8655	0.8986	0.8156	0.7436	0.6245	0.4919	0.5612

Kidney cancer

Age group	15-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.85	0.8	0.78	0.7	0.54
female (1 yr survival)	0.87	0.83	0.78	0.7	0.49
male (5 yr survival)	0.74	0.64	0.59	0.51	0.37
female (5 yr survival)	0.79	0.69	0.62	0.53	0.31

Larynx cancer

Age group	15-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.9	0.89	0.87	0.83	0.76
female (1 yr survival)	0.9	0.89	0.87	0.83	0.76
male (5 yr survival)	0.74	0.72	0.66	0.63	0.63
female (5 yr survival)	0.74	0.72	0.66	0.63	0.63

Liver cancer

Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.41	0.42	0.42	0.35	0.3	0.18
female (1 yr survival)	0.59	0.45	0.4	0.35	0.25	0.13
male (5 yr survival)	0.26	0.21	0.2	0.14	0.07	0.06
female (5 yr survival)	0.32	0.2	0.15	0.13	0.07	0.02

Lung cancer

Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.59	0.4	0.37	0.35	0.3	0.21
female (1 yr survival)	0.68	0.48	0.45	0.42	0.34	0.23
male (5 yr survival)	0.4	0.16	0.13	0.12	0.09	0.06
female (5 yr survival)	0.45	0.21	0.18	0.17	0.12	0.07

Oesophageal cancer

Age group	15-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.52	0.51	0.5	0.43	0.27
female (1 yr survival)	0.56	0.56	0.52	0.41	0.24
male (5 yr survival)	0.17	0.19	0.18	0.15	0.05
female (5 yr survival)	0.27	0.28	0.23	0.15	0.05

Table 1.3 (continued)

Oropharynx cancer

Age group	0-19	20-39	40-59	60-79	>79
male (1 yr survival)	1	0.9453	0.9238	0.9531	0.9687
female (1 yr survival)	1	0.9687	0.9375	0.9453	0.9687
male (5 yr survival)	1	0.9453	0.7136	0.7611	0.7995
female (5 yr survival)	1	0.8532	0.7371	0.8192	0.8193
male (10 yr survival)	1	0.8244	0.7136	0.7189	0.7449
female (10 yr survival)	1	0.8521	0.7208	0.7647	0.7386

Ovarian cancer

Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0	0	0	0	0	0
female (1 yr survival)	0.95	0.91	0.86	0.78	0.64	0.36
male (5 yr survival)	0	0	0	0	0	0
female (5 yr survival)	0.87	0.72	0.58	0.42	0.34	0.17

Pancreatic cancer

Age group	15-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.39	0.29	0.24	0.17	0.08
female (1 yr survival)	0.47	0.34	0.26	0.18	0.09
male (10 yr survival)	0.0602	0.0208	0.0106	0.0066	0.0088
female (10 yr survival)	0.0941	0.0217	0.0096	0.0061	0.0031

Stomach cancer

Age group	15-39	40-49	50-59	60-69	70-79	>79
male (1 yr survival)	0.54	0.55	0.56	0.52	0.44	0.29
female (1 yr survival)	0.57	0.55	0.54	0.53	0.42	0.27
male (5 yr survival)	0.34	0.25	0.26	0.23	0.2	0.08
female (5 yr survival)	0.36	0.3	0.28	0.25	0.2	0.09

Stroke

Age group	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	>89
male (1 yr survival)	1	1	0.9983	0.5	0.9844	0.9824	0.9648	0.9687	1	1
female (1 yr survival)	0.9844	0.998	0.9873	0.5	0.9883	0.95	0.9687	0.918	1	1
male (5 yr survival)	1	1	0.9895	0.0312	0.9824	0.9294	0.9648	0.9687	0.836	1
female (5 yr survival)	0.9798	0.998	0.9758	0.0312	0.982	0.9034	0.9687	0.918	0.844	1
male (10 yr survival)	1	0.9799	0.9709	0.0312	0	0.8509	0.8427	0.6185	0.0561	0.2151
female (10 yr survival)	0.9798	0.9721	0.9546	0.0312	0	0.8259	0.8226	0.5657	0.1125	0.1935

Table 1.4 Smoking relative risks (males)

CHD	male	Age [35-40]	Age [40-45]	Age [45-50]	Age [50-55]	Age [55-60]	Age [60-65]	Age [>65]
	Smoker	3.25	4.71	5.85	3.69	2.71	2.39	1.91
Bowel cancer	male	Age [0-100]						
	Smoker	1.18						
COPD	male	Age [35-40]	Age [40-45]	Age [45-50]	Age [50-55]	Age [55-60]	Age [60-65]	Age [>65]
	Smoker	1	1	1	8.13	9.8	13.21	18.93
Larynx cancer	male	Age [0-100]						
	Smoker	14.6						
Liver cancer	male	Age [0-100]						
	Smoker	1.61						
Oesophageal cancer	male	Age [0-100]						
	Smoker	6.76						
Oropharyngeal cancer	male	Age [0-100]						
	Smoker	10.89						
Stroke	male	Age [35-40]	Age [40-45]	Age [45-50]	Age [50-55]	Age [55-60]	Age [60-65]	Age [>65]
	Smoker	1	1.05	3.75	6.08	3.96	2.55	2.69
Bladder cancer	male	Age [0-100]						
	Smoker	3.27						
Cervical cancer	male	Age [0-100]						
	Smoker	1						
AML	male	Age [0-100]						
	Smoker	1.4						
CML	male	Age [0-100]						
	Smoker	1.31						
Stomach cancer	male	Age [0-100]						
	Smoker	1.62						
Ovarian cancer	male	Age [0-100]						
	Smoker	1						
Kidney cancer	male	Age [0-100]						
	Smoker	2.12						
Lung cancer	male	Age [35-40]	Age [40-45]	Age [45-50]	Age [50-55]	Age [55-60]	Age [60-65]	Age [>65]
	Smoker	1.3	1	5.78	24.97	34.02	31.47	28.4
Pancreatic cancer	male	Age [0-100]						
	Smoker	2.31						

Table 1. 4 Smoking relative risks (females)

CHD	**female**	**Age [35-40]**	**Age [40-45]**	**Age [45-50]**	**Age [50-55]**	**Age [55-60]**	**Age [60-65]**	**Age [>65]**
	Smoker	1.44	2.25	7.71	5.69	1.19	2.56	2.48
Bowel cancer	**female**	**Age [0-100]**						
	Smoker	1.2						
COPD	**female**	**Age [35-40]**	**Age [40-45]**	**Age [45-50]**	**Age [50-55]**	**Age [55-60]**	**Age [60-65]**	**Age [>65]**
	Smoker	1	1	1	12.92	9.47	11.19	14.72
Larynx cancer	**female**	**Age [0-100]**						
	Smoker	13.02						
Liver cancer	**female**	**Age [0-100]**						
	Smoker	1.86						
Oesophageal cancer	**female**	**Age [0-100]**						
	Smoker	7.75						
Oropharyngeal cancer	**female**	**Age [0-100]**						
	Smoker	5.08						
Stroke	**female**	**Age [35-40]**	**Age [40-45]**	**Age [45-50]**	**Age [50-55]**	**Age [55-60]**	**Age [60-65]**	**Age [>65]**
	Smoker	2	5.67	8.22	4.58	5.77	2.76	2.58
Bladder cancer	**female**	**Age [0-100]**						
	Smoker	2.22						
Cervical cancer	**female**	**Age [0-100]**						
	Smoker	1.6						
AML	**female**	**Age [0-100]**						
	Smoker	1.4						
CML	**female**	**Age [0-100]**						
	Smoker	1.31						
Stomach cancer	**female**	**Age [0-100]**						
	Smoker	1.2						
Ovarian cancer	**female**	**Age [0-100]**						
	Smoker	2.4						
Kidney cancer	**female**	**Age [0-100]**						
	Smoker	1.44						
Lung cancer	**female**	**Age [35-40]**	**Age [40-45]**	**Age [45-50]**	**Age [50-55]**	**Age [55-60]**	**Age [60-65]**	**Age [>65]**
	Smoker	2	1	18.08	11.14	17.87	13.32	17.49
Pancreatic cancer	**female**	**Age [0-100]**						
	Smoker	2.55						

Table 1.5 Smoking direct costs

Disease	£ bn	Year
CHD	1.464	2013
Bowel cancer	0.317	2013
COPD	0.800	2013
Larynx cancer	0.022	2013
Liver cancer	0.029	2013
Oesophageal cancer	0.057	2013
Oropharyngeal cancer	0.069	2013
Stroke	0.819	2013
Bladder cancer	0.029	2013
Cervical cancer	0.028	2013
AML	0.314	2013
CML	0.008	2013
Stomach cancer	0.044	2013
Ovarian cancer	0.055	2013
Kidney cancer	0.031	2012
Lung cancer	0.163	2013
Pancreatic cancer	0.058	2013

Table 1.6 Smoking utility weights

Disease	Male	Female
CHD	0.61	0.61
Bowel cancer	0.68	0.68
COPD	0.47	0.47
Larynx cancer	0.85	0.85
Liver cancer	0.62	0.62
Oesophageal cancer	0.90	0.90
Oropharyngeal cancer	0.69	0.69
stroke	0.63	0.63
Bladder cancer	0.71	0.71
Cervical cancer	N/A	0.69
AML	0.65	0.65
CML	0.65	0.65
Stomach cancer	0.71	0.71
Ovarian cancer	N/A	0.85
Kidney cancer	0.66	0.66
Lung cancer	0.56	0.56
Pancreatic cancer	0.79	0.79

Table 1.7 Distribution of mean income and tax, by age and sex, 2012-2013

	Males		
Age	Mean income before tax (£)	Mean tax (£)	Mean income after tax (£)
16-20	13,200	1,010	12,190
20-24	17,400	1,920	15,480
25-29	24,000	3,470	20,530
30-34	31,500	5,580	25,920
35-39	38,300	7,830	30,470
40-44	42,400	9,270	33,130
45-49	44,200	9,910	34,290
50-54	43,500	9,560	33,940
55-59	40,200	8,370	31,830
60-64	33,300	6,220	27,080

	Females		
Age	Mean income before tax (£)	Mean tax (£)	Mean income after tax (£)
16-20	12,000	768	11,232
20-24	15,100	1,390	13,710
25-29	21,200	2,650	18,550
30-34	25,400	3,690	21,710
35-39	27,300	4,260	23,040
40-44	26,900	4,220	22,680
45-49	27,200	4,290	22,910
50-54	26,400	4,030	22,370
55-59	25,000	3,580	21,420
60-64	21,800	2,890	18,910

SUPPLEMENTARY MATERIAL 5: Tobacco duty escalator assumptions

The price of cigarettes was determined from a 2013 overall average retail price of a 20-pack of cigarettes weighted by sales, being £7.13 (12). A typical 25-gram pack of H-RT was priced at £7.89 in 2013 (13). Differences in market share between cigarette smokers and H-RT smokers were incorporated in the simulation (14, 15). Consistent with previous research (16) estimating the prevalence elasticity of tobacco products between 50-75%, the prevalence elasticity for tobacco products in the UK was estimated at 63% of their price elasticity of demand (16). Assuming a price elasticity of -0.5 for cigarettes, and -1.17 for HR-T (17), this results in a prevalence elasticity of -0.315 and -0.74 respectively, where a 10% increase in the price of cigarettes would lead to a 3.15% or 7.4% long-term decline in the prevalence of smokers. The proportion of cigarette smokers versus HR-T smokers were calculated from the GHS dataset and assumed constant over time.

The prices of the two tobacco products were further defined by the rate of consumer price inflation, level of taxation, the 'pass-on' rate, and illicit trade. Based on the existing UK tobacco taxation, it was assumed that from 2015 through to 2035: VAT and ad valorem duty would continue to apply at 20% and 16.5% respectively; and specific duty would increase in the duty escalator from the published rate of £176.20 per 1,000 cigarettes in 2013 (13). All other factors were assumed to be captured by the Consumer Price Index inflation rate, kept constant at 2% per annum. As a result, changes from the 2013 weighted average cigarette price were only affected by the specific duty component of the escalator.

A 'pass-on' rate of tax from the producer to consumer was set at 100%, consistent with evidence from competitive markets (25) and similar previous modelling research (18). Consistent with the most recent available information and existence of activities to contain the illicit market at the time of data collection (19), the illicit tobacco market was estimated to

remain stable at 10% of the total market. The illicit price of tobacco in the UK is estimated to be 50% of the legal price (18), so the illicit price of both cigarettes and H-RT was modelled as such.

SUPPLEMENTARY MATERIAL 6: Sensitivity analysis

Introduction

We carried out a sensitivity analysis on the price elasticity to explore the impact of a higher and lower price elasticity (PE) used to calculate the tobacco duty escalator (TDE) scenario on both smoking prevalence and disease outcomes.

Method

We ran four scenarios as described in Table 4. Table. A PE of -0.5 was chosen in this study since it takes into account the effect of a price increase in all tobacco products without the need for cross-price elasticity figures (because substitution effects are already taken into account). Further, this figure is used by Action on Smoking and Health[4] and based on the literature. We tested the impact of an upper and lower bound PE for cigarettes. The upper bound being a PE of -1.05 derived by HMRC[5], the lower bound being a PE of -0.3. The price elasticity for hand rolled tobacco (H-RT) was held constant. Smoking prevalence from 2015 to 2035 was projected for each of the scenarios, and 20 million individuals were simulated to test the health impacts of each scenario.

Table 4. Table 1. Summary table of scenarios and their definitions

Scenario	Definition
Scenario 0: Baseline	No change in smoking prevalence
Scenario 1: TDE (PE -0.5)	Tobacco duty escalator scenario using a PE of -0.5 for cigarettes
Scenario 2: TDE (PE -0.3)	Tobacco duty escalator with a lower PE of -0.3 for cigarettes
Scenario 3: TDE (PE -1.05)	Tobacco duty escalator with a higher PE of -1.05 for cigarettes

Results

Table 5 shows the impact of the different price elasticities on smoking prevalence by 2035. It was predicted that a PE of -0.3 would result in a change in smoker prevalence from 2015 to 2035 of 33%, and a PE of -1.05 would result in a change in smoker prevalence of 57.8% compared to a change of 39.6% for a PE of -0.5. More than doubling the PE as with the upper bound test decreases the predicted smoking prevalence by around 2% compared with the -0.5 PE by 2035.

[4]ASH reference: http://ash.org.uk/information-and-resources/taxation-illicit-trade/taxation/the-effects-of-increasing-tobacco-taxation/

[5] HMRC reference: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/331580/cig-consumption-uk.pdf

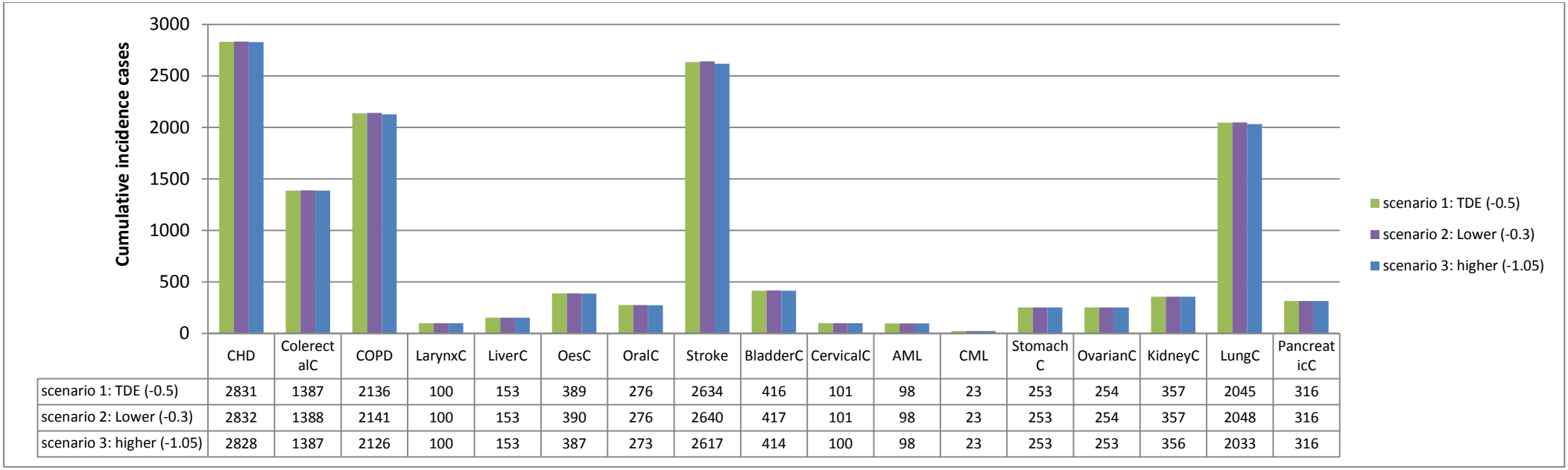
Table 5. Change in smoker prevalence over time and total smoker prevalence in 2035 by PE scenario

Price elasticity	Change in smoker prevalence by 2035 relative to 2015 (holding H-RT PE constant) (%)	Predicted smoker prevalence by 2035 (%)
Scenario 1: TDE (PE -0.5)	39.6	6
Scenario 2: TDE (PE -0.3)	33.0	7
Scenario 3: TDE (PE -1.05)	57.8	4

Table 6 shows the cumulative incidence cases per 100,000 for each scenario by disease in 2015 and 2035. Compared to a PE of -0.5, there are slightly more disease cases when the PE is set to -0.3 and slightly fewer disease cases when the PE is set to -1.05 by 2035. The largest impacts are observed for COPD, stroke, and lung cancer. For many cancers there is no difference across the scenarios. Figure 6 illustrates the differences between each TDE scenario in 2035 by disease.

Table 6. Cumulative incidence cases per 100,000 by scenario and year

	Year	CHD	ColerectalC	COPD	LarynxC	LiverC	OesC	OralC	Stroke	BladderC	CervicalC	AML	CML	StomachC	OvarianC	KidneyC	LungC	PancreaticC
Scenario 0: Baseline	2015	151	71	125	6	8	27	17	162	26	5	5	1	14	13	19	135	19
	2035	2838	1388	2165	101	154	394	280	2662	419	101	99	23	254	255	356	2067	318
Scenario 1: TDE (PE -0.5)	2015	151	71	124	6	8	27	17	162	26	5	5	1	13	13	19	135	19
	2035	2831	1387	2136	100	153	389	276	2634	416	101	98	23	253	254	357	2045	316
Scenario 2: TDE (PE -0.3)	2015	151	71	124	6	8	27	17	162	26	5	5	1	13	13	19	135	19
	2035	2832	1388	2141	100	153	390	276	2640	417	101	98	23	253	254	357	2048	316
Scenario 3: TDE (PE -1.05)	2015	151	71	124	6	8	27	18	162	26	5	5	1	13	13	19	135	19
	2035	2828	1387	2126	100	153	387	273	2617	414	100	98	23	253	253	356	2033	316



	CHD	Colerectal C	COPD	LarynxC	LiverC	OesC	OralC	Stroke	BladderC	CervicalC	AML	CML	Stomach C	OvarianC	KidneyC	LungC	Pancreatic C
scenario 1: TDE (-0.5)	2831	1387	2136	100	153	389	276	2634	416	101	98	23	253	254	357	2045	316
scenario 2: Lower (-0.3)	2832	1388	2141	100	153	390	276	2640	417	101	98	23	253	254	357	2048	316
scenario 3: higher (-1.05)	2828	1387	2126	100	153	387	273	2617	414	100	98	23	253	253	356	2033	316

Figure 6. Cumulatve incidence cases per 100,000 for each TDE scenario by disease in 2035

Discussion

Doubling the PE almost halves the smoking prevalence predicted by 2035. Only small differences are observed in the subsequent disease outcomes by 2035. This is possibly because an ex-smoker's relative risk for many of the diseases, particularly cancers, takes almost two decades to return to that of a never smoker's. Larger differences may be observed if the simulation was run further into the future. The ex-smoker relative risks were computed using a method developed by Hoogenveen and colleagues (20), and as part of the EU-funded DYNAMO project.

References

1. Hoogendoorn M, Feenstra T, Hoogenveen R, Genugten MV, Rutten-van Mölken M. A Health Policy Model for COPD : Effects of Smoking Cessation A health policy model for COPD : effects of smoking cessation.1-104.

SUPPLEMENTARY MATERIAL 7: Projected baseline future trends of smoking prevalence by income quintile in the UK in 2035

As demonstrated in Table 7, the baseline indicates that a socioeconomic gradient in smoking prevalence will remain. In 2035, adult smoking prevalence is estimated to be 14.3% among females and 15.7% among males in income quintile 1 (poorest), compared to 2.6% and 2.4% respectively in income quintile 5.

Table 7. Projected baseline future trends of smoking prevalence by income quintile in the UK in 2035

Baseline (quintile 1 most deprived, quintile 5 least)	**Female prevalence (%)**	**Male prevalence (%)**
Income quintile 1	14.3	15.7
Income quintile 2	11.0	11.7
Income quintile 3	7.8	6.8
Income quintile 4	3.9	4.7
Income quintile 5	2.6	2.4

SUPPLEMENTARY MATERIAL 8. Incidence and cumulative incidence disease outputs by scenario and year

Table 1. Incidence cases in the UK population by scenario and year

Baseline	year		CHD	Colerectal IC	COPD	Larynx C	Liver C	OesC	OralC	Stroke	Bladder C	Cervical C	AML	CML	Stomach C	Ovarian C	Kidney C	LungC	Pancreatic C	Tobacco-related cancers
	2015	Inc	98056	46106	81172	3896	5195	17533	11689	105848	16884	3247	3247	649	9091	8442	12338	87666	12338	238321
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	918
	2020	Inc	91962	44974	77195	3356	5370	14096	10069	89949	14768	3356	3356	671	8055	8726	12083	74510	10740	214131
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	918
	2025	Inc	94092	46354	72645	3459	5535	13145	8994	86482	13837	3459	3459	692	8302	8302	11762	67802	10378	205481
		95% CI	671	671	671	0	0	0	0	671	0	0	0	0	0	0	0	671	0	949
	2030	Inc	97321	47595	68196	2841	4973	12076	8524	85955	13497	3552	3552	710	8524	8524	12076	63223	10656	200326
		95% CI	710	710	710	0	0	0	0	710	0	0	0	0	0	0	0	710	0	1005
	2035	Inc	99552	49413	65399	2907	5087	10900	7993	85745	13080	3633	3633	727	8720	8720	11627	58859	10900	196197
		95% CI	727	727	727	0	0	0	0	727	0	0	0	0	0	0	0	727	0	1028
Scenario 1 (TDE)	**year**		**CHD**	**Colerectal IC**	**COPD**	**Larynx C**	**Liver C**	**OesC**	**OralC**	**Stroke**	**Bladder C**	**Cervical C**	**AML**	**CML**	**Stomach C**	**Ovarian C**	**Kidney C**	**LungC**	**Pancreatic C**	**Tobacco-related cancers**
	2015	Inc	98056	46106	81172	3896	5195	17533	11689	105199	16884	3247	3247	649	9091	8442	12338	88315	12338	238970
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	1299
	2020	Inc	91962	44974	77195	4028	5370	14096	10069	89949	14768	3356	3356	671	8055	8726	12083	73838	10740	214131
		95% CI	671	671	671	0	0	0	0	671	0	0	0	0	0	0	0	671	0	671
	2025	Inc	94092	46354	71953	3459	4843	12453	8994	85790	13837	3459	3459	692	8302	8302	11762	67110	10378	203405
		95% CI	692	692	692	0	0	0	0	692	0	0	0	0	0	0	0	692	0	692
	2030	Inc	96611	47595	66775	2841	4973	11366	8524	83824	13497.1174	3551.873	3551.873	710	8524	8524	12076	61803	10656	198195
		95% CI	710	710	710	0	0	0	0	710	0	0	0	0	0	0	0	710	0	710
	2035	Inc	99552	49413	63219	2907	5087	10900	7993	83566	13080	3633	3633	727	8720	8720	11627	56679	10173	193291
		95% CI	727	727	727	0	0	0	0	727	0	0	0	0	0	0	0	727	0	727

Table 2. Cumulative incidence cases in the UK population by scenario and year

scenario 0	year		CHD	ColerectalC	COPD	LarynxC	LiverC	OesC	OralC	Stroke	BladderC	CervicalC	AML	CML	StomachC	OvarianC	KidneyC	LungC	PancreaticC	Tobacco-related cancers
	2015	Cumu. Inc.	98056	46106	81172	3896	5195	17533	11689	105848	16884	3247	3247	649	9091	8442	12338	87666	12338	238321
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	918
	2020	Cumu. Inc.	566620	272083	472183	23114	31039	93116	65379	579168	93116	20472	19812	4623	50851	51511	72644	479448	68021	1345227
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	918
	2025	Cumu. Inc.	1024633	495206	836750	40261	56365	159030	112059	1009871	161043	36906	35564	8052	92599	93270	130847	822659	119440	2363300
		95% CI	1321	660	1321	0	0	660	660	1321	660	0	0	0	660	660	660	1321	660	2288
	2030	Cumu. Inc.	1489470	723963	1174822	55847	81046	217938	154600	1426131	226111	53122	51760	12259	133487	135530	189334	1136683	170264	3341942
		95% CI	2043	1362	2043	681	681	681	681	2043	681	681	681	0	681	681	681	2043	681	3336
	2035	Cumu. Inc.	1961426	956894	1494025	69730	106322	272708	194002	1837844	289968	69730	67659	15879	176052	176742	246473	1427056	220928	4290145
		95% CI	2071	1381	2071	690	690	690	690	2071	690	690	690	0	690	690	690	2071	690	3382
scenario 1 (TDE)	**year**		**CHD**	**ColerectalC**	**COPD**	**LarynxC**	**LiverC**	**OesC**	**OralC**	**Stroke**	**BladderC**	**CervicalC**	**AML**	**CML**	**StomachC**	**OvarianC**	**KidneyC**	**LungC**	**PancreaticC**	**Tobacco-related cancers**
	2015	Cumu. Inc.	98056	46106	81172	3896	5195	17533	11689	105199	16884	3247	3247	649	9091	8442	12338	88315	12338	238970
		95% CI	649	649	649	0	0	0	0	649	0	0	0	0	0	0	0	649	0	918
	2020	Cumu. Inc.	566620	272083	471523	23114	31039	93116	65379	578507	93116	20472	19812	4623	50851	51511	72644	478787	68021	1344567
		95% CI	1321	660	1321	0	0	660	660	1321	660	0	0	0	660	660	660	1321	660	2288
	2025	Cumu. Inc.	1023962	495206	833395	40261	55694	158359	112059	1005845	161043	36906	35564	8052	91928	93270	130847	819975	119440	2358603
		95% CI	2013	1342	1342	0	671	671	671	1342	671	0	0	0	671	671	671	1342	671	2684
	2030	Cumu. Inc.	1487427	723282	1164606	55166	81046	215895	153238	1416597	225430	53122	51760	12259	133487	134849	188653	1129191	169583	3326959
		95% CI	2043	1362	2043	681	681	681	681	2043	681	681	681	0	681	681	681	2043	681	3336
	2035	Cumu. Inc.	1957284	956204	1474004	69040	105631	269256	190550	1817823	288587	69730	67659	15879	175362	175362	245782	1410487	219547	4259077
		95% CI	2071	1381	2071	690	690	690	690	2071	690	690	690	0	690	690	690	2071	690	3382

References

1. Rudolf T Hoogenveen PHvB, Hendriek C Boshuizen and Talitha L Feenstra. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. Cost Effectiveness and Resource Allocation. 2008.
2. Claxton K, Sculpher M, Palmer S, Culyer AJ. CAUSES FOR CONCERN: IS NICE FAILING TO UPHOLD ITS RESPONSIBILITIES TO ALL NHS PATIENTS? Health Economics. 2015;24(1):1-7.
3. (DYNAMO-HIA) ADMfHIA. United Kingdom vs2.0.6.zip: National Institute for Public Health and the Environment; 2015 [15/11/2014]. Available from: http://www.dynamo-hia.eu/.
4. Sophia Fircanis PM, Naushaba Khan, Jorge J. Castillo. The relation between cigarette smoking and risk of acute myeloid leukemia: An updated meta-analysis of epidemiological studies. American Journal of Hematology. 2014;89(8):E125–E32.
5. Botteri E IS, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA 2008;300(23):2765-78.
6. Cancer ICoESoC. Cervical carcinoma and reproductive factors: collaborative reanalysis of individual data on 16,563 women with cervical carcinoma and 33,542 women without cervical carcinoma from 25 epidemiological studies. Int J Cancer. 2006;119(5):1108-24.
7. Musselman JRBB, Cindy K; Cerhan, James R; Nguyen, Phuong; Hirsch, Betsy. Risk of adult acute and chronic myeloid leukemia with cigarette smoking and cessation. Cancer epidemiology. 2013;37(4):410-6.
8. Hunt JDvdH, Olga L; McMillan, Garnett P; Boffetta, Paolo; Brennan, Paul Renal cell carcinoma in relation to cigarette smoking: meta-analysis of 24 studies. International journal of cancer. 2005;114(1):101-8.
9. Lee YC, C; Yang, YC; Stayner, L; Hashibe, M; Straif K. Meta-analysis of epidemiologic studies on cigarette smoking and liver cancer. Int J Epidemiol. 2009;28:1497-511.
10. Jordan SJW, David C; Purdie, David M; Green, Adèle C; Webb, Penelope M. Does smoking increase risk of ovarian cancer? A systematic review. Gynecologic oncology 2006;103(3): 1122-9.
11. Ladeiras-Lopes RP, AK; Nogueira, A; Pinheiro-Torres, T; Pinto, I; Santos-Pereira, R; Lunet, N. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. Cancer Causes Control 2008 19(7):689-701.
12. HMRC. Tax and Duty Bulletins. . 2015.
13. Info UT. Tobacco Factsheet 2013: HM Revenue & Customs; 2014 [17/02/2015]. Available from: https://www.uktradeinfo.com/Statistics/Pages/TaxAndDutyBulletins.aspx.
14. Office for National Statistics. Smoking (General Lifestyle Survey Overview - a report on the 2011 General Lifestyle Survey). 2013;Chapter 1.
15. Anna B. Gilmore BT, Rosemary Hiscock, Gordon Taylor. Smoking patterns in Great Britain: the rise of cheap cigarette brands and roll your own (RYO) tobacco. Journal of Public Health. 2014:pp. 1–11.
16. WHO. Effectiveness of Tax and Price Policies for Tobacco Control: IARC Handbook of Cancer Prevention Volume 14 2015 [09/02/2015]. Available from: http://www.iarc.fr/en/publications/pdfs-online/prev/handbook14/index.php.
17. Townsend J. Price and consumption of tobacco. British Medical Bulletin. 1996;52(1):132-42.
18. Cobiac LJI, Tak; Nghiem, Nhung; Blakely, Tony; Wilson, Nick. Modelling the implications of regular increases in tobacco taxation in the tobacco endgame. Tobacco Control. 2014;0:1-7.
19. Force HMsRaCaB. Tackling illicit tobacco: from leaf to light. The HMRC and Border Force strategy to tackle tobacco smuggling. . 2015.
20. Hoogendoorn M, Feenstra T, Hoogenveen R, Genugten MV, Rutten-van Mölken M. A Health Policy Model for COPD : Effects of Smoking Cessation A health policy model for COPD : effects of smoking cessation.1-104.