Supplementary File 1

Projected health and economic impacts of improving attendance to the NHS Health Check in England: A sex-disaggregated analysis with implications for men's health

Contents

Microsimulation framework

Our 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 BMI, smoking, and blood pressure over time

The risk factors (RF) analysed in the model are BMI, smoking, and blood pressure as described in the table below

Risk factor (RF)	Number of categories (N)	Categories
BMI	*3*	BMI < 25 kg m^{-2} (normal weight) BMI from 25 to 29.99 kg m^{-2} (overweight) BMI ≥ 30 kg m^{-2} (obesity)
Smoking	*2*	Non-smoker Smoker
Blood pressure	*3*	<120 mmHG 120-140 mmHG >140 mmHG

For the RF, let N be the number of categories for a given risk factor, e.g. $N = 3$ for BMI. 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

$$\ln\left(\frac{p_k(t)}{p_1(t)}\right) = \beta_0^k + \beta_1^k t \qquad (0.1)$$

The prevalence of the first category is obtained by using the normalisation constraint $\sum_{k=1}^{N} p_k(t) = 1$. Solving equation (0.1) for $p_k(t)$, we obtain

$$p_k(t) = \frac{\exp\left(\beta_0^k + \beta_1^k t\right)}{1 + \sum_{k'=1}^{N} \exp\left(\beta_0^{k'} + \beta_1^{k'} t\right)}, \qquad (0.2)$$

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

The mean prevalence across all of the available datasets for each measure was used to approximate the β_0 coefficients for each 5 year age and sex group.

Multinomial logistic regression for risk factors BMI and smoking

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 normal weight, overweight and obese 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 $\{\{t_i, \mu_{ki}, \sigma_{ki} | k \in [0,K\text{-}1]\} | i \in [0,N\text{-}1]\}$. 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 obesity in the three states already mentioned, the $k = 0$ regression function represents the probability of being normal weight over time, $k = 1$ the probability of being overweight and $k = 2$ the probability of being obese.

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).

---

[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.

$$S(\mathbf{a},\mathbf{b}) = \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k(\mathbf{a},\mathbf{b};t_i)-\mu_{ki}\right)^2}{\sigma_{ki}^2} \tag{0.3}$$

$$\begin{gathered} p_k(\mathbf{a},\mathbf{b},t) \equiv \frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}} \\ \mathbf{a} \equiv (a_0,a_1,..,a_{K-1}), \quad \mathbf{b} \equiv (b_0,b_1,..,b_{K\text{-}1}) \\ A_0 \equiv 0, \quad A_k \equiv a_k + b_k t \end{gathered} \tag{0.4}$$

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*-1) regression parameters to be determined.
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

$$\frac{\partial S}{\partial a_j} = \frac{\partial S}{\partial b_j} = 0 \qquad \text{for j=1,2,....,k-1} \tag{0.5}$$

noting the relations

$$\begin{gathered} \frac{\partial p_k}{\partial A_j} = \frac{\partial}{\partial A_j}\left(\frac{e^{A_k}}{1+e^{A_1}+..+e^{A_{K-1}}}\right) = p_k\delta_{kj} - p_k p_j \\ \frac{\partial}{\partial a_j} = \frac{\partial}{\partial A_j} \\ \frac{\partial}{\partial b_j} = t\frac{\partial}{\partial A_j} \end{gathered} \tag{0.6}$$

The values of the vectors **a**, **b** that satisfy these equations are denoted $\hat{\mathbf{a}},\hat{\mathbf{b}}$. They provide the trend lines $p_k(\hat{\mathbf{a}},\hat{\mathbf{b}};t)$, 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 2*K*-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 (2*K*-2)-dimensional probability distribution function are obtained by first approximating it as a (2*K*-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 $\hat{\mathbf{S}} \equiv S(\hat{\mathbf{a}},\hat{\mathbf{b}})$. Hence

$$
\begin{aligned}
S(\mathbf{a},\mathbf{b}) &= \tfrac{1}{2}\sum_{k=0}^{k=K-1}\sum_{i=0}^{i=N-1}\frac{\left(p_k(\mathbf{a},\mathbf{b};t_i)-\mu_{ki}\right)^2}{\sigma_{ki}^2} \\
&\equiv S\left(\hat{a},\hat{b}\right)+\tfrac{1}{2}\left(a-\hat{a},b-\hat{b}\right)P^{-1}\left(a-\hat{a},b-\hat{b}\right)+\ldots \\
&\approx S\left(\hat{a},\hat{b}\right)+\tfrac{1}{2}\sum_{i,j}\left(a_i-\hat{a}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{a}_j}\left(a_j-\hat{a}_j\right)+\tfrac{1}{2}\sum_{i,j}\left(a_i-\hat{a}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{a}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right)+ \\
&+\tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{a}_j}\left(a_j-\hat{a}_j\right)+\tfrac{1}{2}\sum_{i,j}\left(b_i-\hat{b}_i\right)\frac{\partial^2\hat{S}}{\partial\hat{b}_i\partial\hat{b}_j}\left(b_j-\hat{b}_j\right)
\end{aligned}
\tag{0.7}
$$

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) $\hat{p}_k(t)=p\left(\hat{\mathbf{a}},\hat{\mathbf{b}},t\right)$

$$
\begin{aligned}
p_k(\mathbf{a},\mathbf{b},t) &= p_k\left(\hat{\mathbf{a}}+\mathbf{a}-\hat{\mathbf{a}},\hat{\mathbf{b}}+\mathbf{b}-\hat{\mathbf{b}},t\right) \\
&= \hat{p}_k(t)+\left(\nabla_{\hat{a}},\nabla_{\hat{b}}\right)\hat{p}_k(t)\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}+\ldots
\end{aligned}
\tag{0.8}
$$

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

$$
\begin{aligned}
&\sigma_k^2(t)\equiv\left\langle\left(p_k(\mathbf{a},\mathbf{b},t)-\hat{p}_k(t)\right)^2\right\rangle=\left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)\left\langle\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}\begin{pmatrix}\mathbf{a}-\hat{\mathbf{a}}\\ \mathbf{b}-\hat{\mathbf{b}}\end{pmatrix}^T\right\rangle\times \\
&\left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)^T=\left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)\,P\,\left(\nabla_{\hat{a}}\hat{p}_k(t),\nabla_{\hat{b}}\hat{p}_k(t)\right)^T
\end{aligned}
\tag{0.9}
$$

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

$$
\sigma_k^2(t)=\begin{pmatrix}\frac{\partial\hat{p}_k(t)}{\partial\hat{a}_1} & \frac{\partial\hat{p}_k(t)}{\partial\hat{a}_2} & \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_1} & \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_2}\end{pmatrix}\begin{bmatrix}P_{aa11} & P_{aa12} & P_{ab11} & P_{ab12}\\ P_{aa21} & P_{aa22} & P_{ab21} & P_{ab22}\\ P_{ba11} & P_{ba12} & P_{bb11} & P_{bb12}\\ P_{ba21} & P_{ba22} & P_{bb21} & P_{bb22}\end{bmatrix}\begin{pmatrix}\frac{\partial\hat{p}_k(t)}{\partial\hat{a}_1}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{a}_2}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_1}\\ \frac{\partial\hat{p}_k(t)}{\partial\hat{b}_2}\end{pmatrix}
\tag{0.10}
$$

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). Individuals in the model are simulated from their year of birth (which may be before the start year of the simulation). In the course of their lives, simulated people can die from one of the diseases caused by obesity or 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 BMI/tobacco-related diseases will change as a consequence of the changing distribution of BMI or smoking level in the population.
The microsimulation incorporates an economic module. The module employs Markov-type simulation of long-term health benefits, health care costs and cost-effectiveness of specified interventions. It synthesises and estimates evidence on cost-effectiveness analysis and cost-utility analysis. The model is used to project the differences in quality-adjusted life years (QALYs), direct and indirect lifetime health-care costs and as a consequence of interventions incremental cost effectiveness ratios (ICERs) over a specified time scale. Outputs can be discounted for any specific discount rate.
This section provides an overview of the initialisation of the microsimulation model and will be expanded upon in the next sections.

Population models

The population is created in the start-year and propagated forwards in time by allowing females to give birth.

People within the model can die from specific diseases or from other causes. A disease file is created within the program to represent deaths from other causes. The following distributions are required by the population editor (

Supplementary Table1.1).

Supplementary Table1.1 Summary of the parameters representing the distribution component

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

Birth model

Any female in the childbearing 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,

$$p_b\left(\text{n at a} \mid N\right) = \frac{N!}{n!(N-n)!}\left(p_b(a)\right)^n\left(1-p_{bm}(a)\right)^{N-n} \tag{0.11}$$

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

$$p_b\left(n\,at\,a\right) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{N!}p_b\left(\text{n at a} \mid N\right) = e^{-\lambda}\sum_{N=n}^{\infty}\frac{\lambda^N}{n!(N-n)!}\left(p_b(a)\right)^n\left(1-p_b(a)\right)^{N-n} \tag{0.12}$$

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 $\lambda p_b(a)$,

$$p_b\left(n\,at\,a\right) = e^{-\lambda p_b(a)}\frac{\left(\lambda p_b(a)\right)^n}{n!} = p_{\lambda p_b(a)}(n) \tag{0.13}$$

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 $N_m/(N_m+N_f)$.

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.

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. This 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.

The 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

---

[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.

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 status) – see ***Error! Reference source not found.***.

Continuous risk factors

In the case of a continuous RF, for each discrete distribution D there is a continuous counterpart. Let β denote the RF value in the continuous scale and let $f(\beta|A,S,t)$ be the probability density function of β for age group A and sex S at time t. Then

$$p_k(t|A,S) = \int_{\beta \in k} f(\beta|A,S,t)\, d\beta. \tag{0.14}$$

Equations (0.2) and (0.14) both refer to the same quantity. However, equation (0.14) uses the definition of the probability density function to express the age-and-sex-specific percentage of individuals in RF category *k* at time *t*. Equation (0.2) gives an estimate of this quantity using equation (0.1) for all *k* = 0, …, N. The cumulative distribution function of β is

$$F(\beta|A,S,t) = \int_0^{\beta} f(\beta|A,S,t)\, d\beta. \tag{0.15}$$

At time *t*, a person with sex S belonging to the age group A is said to be on the p–th percentile of this distribution if $F(\beta|A,S,t) = p/100$. Given the cross-sectional information from the set of distributions D, it is possible to simulate longitudinal trajectories by forming pseudo-cohorts within the population. A key requirement for these sets of longitudinal trajectories is that they reproduce the cross-sectional distribution of RF categories for any year with available data. The method adopted here and in the earlier Foresight report (1) is based on the assumption that person's RF value changes throughout their lives in such a way that they always have the same associated percentile rank. As they age, individuals move from one age group to another and their RF value changes so that they have the same percentile rank but of a different RF distribution. In a nutshell, we assume (in accordance with research on the long-term success rate in dieting) that relatively fat people will remain relatively fat and relatively thin people will remain relatively thin. Crucially it meets the important condition that the cross-sectional RF distributions obtained by simulation match the RF distributions of the observed data.
The above procedure can be explained using the example of the BMI distribution. The BMI distributions are known for the population stratified by sex and age for all years of the simulation (by extrapolation of fitted model, see equation (0.1)). A person who is in age group A and who grows ten year older will at some time move into the next age group A' and will have a BMI that was described first by the distribution $f(\beta|A,S,t)$ and then at the later time t' by the distribution $f(\beta|A',S,t')$. If the BMI of that individual is on the p-th percentile of the BMI distribution, their BMI will change from β to β' so that

$$\beta = F^{-1}\left(\frac{p}{100}|A,S,t\right) \tag{0.16}$$

$$\beta' = F^{-1}\left(\frac{p}{100}|A',S,t'\right) \Rightarrow \beta' = F^{-1}\left(F(\beta\,|\,A,S,t)|A',S,t'\right) \tag{0.17}$$

Where F^{-1} is the inverse of the cumulative distribution function of β, which we model with a continuous uniform, normal or lognormal distribution (depending on the risk factor) within the RF categories (see ***Error! Reference source not found.***). Equation (0.17) guarantees that the

transformation taking the random variable β to β' ensures the correct cross-sectional distribution at time t'.
The microsimulation first generates individuals from the RF distributions of the set D and, once generated, grows the individual's RF in a way that is also determined by the set D. It is possible to implement equation (0.17) as a suitably fast algorithm.

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

$$\begin{bmatrix} p_0' \\ p_1' \\ p_2' \end{bmatrix} = \begin{bmatrix} 1-p & 0 & 0 \\ 0 & 1-p & 1-p \\ p & p & p \end{bmatrix} \begin{bmatrix} p_0 \\ p_1 \\ p_2 \end{bmatrix} \tag{0.18}$$

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 consist 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.

BMI

The microsimulation framework is employed to model and examine obesity and other health related outcomes. It has many modes of operation and can analyse the health outcomes for groups of individuals or specified populations consequent upon single or aggregated interventions using real or hypothetical data over any specified time period. At present, the model estimates the future burden of diseases by making evidence-based extrapolations of selected risk factors specific to several obesity-related diseases (e.g. type 2 diabetes, coronary heart disease, stroke).

Smoking

The microsimulation framework applied to smoking enables us to measure the future health impact of changes in rates of tobacco consumption. In the simulation each person is categorised into one of the three smoking groups: smokers, ex-smokers and people who have never smoked.

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 and (2),

$$RR_{\text{ex-smoker}}(A,S,T_{\text{cessation}}) = 1 + (RR_{\text{smoker}}(A,S) - 1)\exp(-\gamma(A)T_{\text{cessation}}) \tag{0.19}$$

$$\gamma(A) = \gamma_0 \exp(-\eta A) \tag{0.20}$$

where γ is the regression coefficient of time dependency. The constants γ0 and η are intercept and regression coefficient of age dependency, respectively, which are related to the specified disease

Supplementary Table1.2.

Supplementary Table1.2 Parameter estimates for γ_0 and η related to each disease (2)

Disease	γ_0	η
AMI	0.24228	0.05822
Stroke	0.31947	0.01648
COPD	0.20333	0.03087
Diabetes	0.024811	0
Lung cancer	0.15637	0.02065
Hypertension (proxy: diabetes)	0.02481	0

However, a minimum exists when the cessation time is equal to η^{-1}. The minimum value was calculated by the method detailed below (equations (0.21), (0.22) and (0.23)). Where, time *t* is equal to the age *A* of an individual.

$$\rho_{Exsmk}(t) = 1 + (\rho_{smk} - 1) f(t) \tag{0.21}$$

$$\begin{aligned} f(t) &= \exp\left(-\gamma_0 (t - t_0)\exp(-\eta t)\right) \\ &\Rightarrow \\ f'(t) &= -\gamma_0 f(t) e^{-\eta t}\left(-\eta(t - t_0) + 1\right) \end{aligned} \tag{0.22}$$

The function *f(t)* has the following properties:

$$\begin{aligned} &f(t_0) = 1 \\ &f'(t_0) = -\gamma_0 e^{-\eta t_0} \\ &f(t) \text{ has a minimum at } t = t_0 + \eta^{-1} \\ &f(\infty) = A \end{aligned} \tag{0.23}$$

In order to avoid the $RR_{ex\text{-}smoker}$ from increasing the cessation time was set equal to η^{-1} when the cessation time was greater than η^{-1} (see equation (0.24)).

$$RR_{\text{ex-smoker}}(A,S,T_{\text{cessation}}) = \begin{cases} 1+(RR_{\text{smoker}}(A,S)-1)\exp(-\gamma(A)T_{\text{cessation}}) & T_{\text{cessation}} < \eta^{-1} \\ 1+(RR_{\text{smoker}}(A,S)-1)\exp(-\gamma(A)\eta^{-1}) & T_{\text{cessation}} \geq \eta^{-1} \end{cases} \tag{0.24}$$

$$\gamma(A) = \gamma_0 \exp(-\eta A) \tag{0.25}$$

Relative risks

Each RF 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 (0.26).

$$\begin{aligned} p_{\mathrm{A}}(d|\alpha,a,s) &= \rho_{\mathrm{A}|d}(\alpha|a,s)\, p_{\mathrm{A}}(d|\alpha_0,a,s) \\ \rho_{\mathrm{A}|d}(\alpha_0|a,s) &\equiv 1 \end{aligned} \tag{0.26}$$

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,

$$\begin{aligned} p(d|a,s) &= \sum_{\alpha} p_{\mathrm{A}}(d|\alpha,a,s)\pi_A(\alpha|a,s) \\ &= p_A(d|\alpha_0,a,s)\sum_{\alpha}\rho_{\mathrm{A}|d}(\alpha|a,s)\pi_A(\alpha|a,s) \end{aligned} \tag{0.27}$$

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

$$p(d|\alpha,a,s) = \rho_{A|d}(\alpha|a,s)\frac{p(d|a,s)}{\sum_{\beta}\rho_{A|d}(\beta\,|\,a,s)\pi_A(\alpha|a,s)} \tag{0.28}$$

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

The method described above for calibrating incidence statistics has been adapted for multistage diseases. Each individual transition probability between stages of the multistage disease is calculated by calibrating the incidence risk. We have a multistate disease *d* with N states and a single risk factor *B* for this disease which has different states β. The incidence probability for transition from the disease state *i* to *j* given the risk state, β, the person's age, *a*, and gender, *s* is denoted by $p=(d_i d_j|\beta,a,s)$.

$$p(d_i d_j|\beta,a,s) = \rho_{d_i d_j}(\beta,a,s)\, p(d_i d_j\,|\,\beta_0,a,s) \tag{0.29}$$

$$\rho_{d_i d_i}(\beta_0, a, s) = 1 \tag{0.30}$$

Where β_0 is the zero risk state and $d_i d_i$ is the zero risk disease state.

$$p(d_i d_j | a, s) = \sum_{\beta} p_B(d_i d_j | \beta, a, s) \pi_B(\beta | a, s) = p_B(d_i d_j | \beta_0, a, s) \sum_{\beta} \rho_{B|d_i d_j}(\beta | a, s) \pi_B(\beta | a, s) \tag{0.31}$$

For multistate diseases we would have,

$$p(d_i d_j | \beta, a, s) = \rho_{B|d_i d_j}(\beta | a, s) \frac{p(d_i d_j | a, s)}{\sum_{\gamma} \rho_{\gamma|d_i d_j}(\beta | a, s) \pi_{\gamma}(\gamma | a, s)} \tag{0.32}$$

The incidence probabilities would be computed for all the different possible disease state transitions.

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.

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.

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

$$\begin{pmatrix} p_d(a+1) \\ p_{\Omega}(a+1) \\ p_{\bar{\Omega}}(a+1) \end{pmatrix} = \begin{pmatrix} 1 - p_{\bar{\omega}} - p_{\omega} & 0 & 0 \\ p_{\omega} & 1 & 0 \\ p_{\bar{\omega}} & 0 & 1 \end{pmatrix} \begin{pmatrix} p_d(a) \\ p_{\Omega}(a) \\ p_{\bar{\Omega}}(a) \end{pmatrix} \tag{0.33}$$

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

$$p_d(a_0 + N) = \prod_{k=1}^{k=N} \left(1 - p_{\omega}(a_k) - p_{\bar{\omega}}(a_k)\right) \tag{0.34}$$

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

$$p_d\left(a_0+N\right)=\prod_{k=1}^{k=N}\left(1-p_\omega\left(a_k\right)\right) \tag{0.35}$$

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

In equation (0.35) 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 - $p_\omega\left(a_0+1\right)$, the second year - $p_\omega\left(a_0+2\right)$, and so on. These are *disease survival statistics*, closely connected to but not the same as *disease mortality statistics*.

Disease incidence

The probability that a person, who at age 0 does not have the disease, first gets the disease at age a_0 in state K_0 given as

$$p_{inc}\left(a_0,K_0\right)=\prod_{a=0}^{a=a_0-1}\left(1-\sum_{k=1}^{K}p_{0k}\left(a\right)\right)p_{0K_0} \tag{0.36}$$

Disease survival

Once a person has the disease they can possibly change disease stage or they can die from the disease. (This analysis focusses only on the identified disease and does not allow for the possibility that they die from other causes). Suppose they acquire the disease at age a_0 in stage K, then the initial state vector is determined from the initial conditions $p_i\left(a_0\right)=\delta_{iK_0}\quad\left(K_0>0\right),\ p_\omega\left(a_0\right)=0$. At subsequent ages, the state probabilities are given by the recursion equation

$$\begin{pmatrix} p_0(a+1\,|\,a_0,K_0) \\ p_1(a+1\,|\,a_0,K_0) \\ .. \\ p_{N-1}(a+1\,|\,a_0,K_0) \\ p_\omega(a+1\,|\,a_0,K_0) \end{pmatrix} = T(a,a_0) \begin{pmatrix} p_0(a\,|\,a_0,K_0) \\ p_1(a\,|\,a_0,K_0) \\ ... \\ p_{N-1}(a\,|\,a_0,K_0) \\ p_\omega(a)\,|\,a_0,K_0 \end{pmatrix}$$

$$T(a,a_0) \equiv \begin{pmatrix} \left(1-\sum_{k>0} p_{k0}\right) & 0 & ... & 0 & 0 \\ p_{0,1} & \left(1-\sum_{k>1} p_{1,k}\right)\left(1-p_\omega^1(a\,|\,a_0)\right) & ... & 0 & 0 \\ ... & ... & ... & ... & ... \\ p_{0,N-1} & p_{1,N-1}\left(1-p_\omega^1(a\,|\,a_0)\right) & ... & \left(1-p_\omega^{N-1}(a\,|\,a_0)\right) & 0 \\ 0 & p_\omega^1(a\,|\,a_0) & ... & p_\omega^{N-1}(a\,|\,a_0) & 1 \end{pmatrix}$$

(0.37)

Where, for survival model 2,

$$p_\omega^K(a\,|\,a_0) = \begin{cases} p_{\omega 0}^K & (a = a_0) \\ p_{\omega 1}^K & (a_0 < a \le a_0 + 4) \\ p_{\omega 5}^K & (a_0 + 4 < a) \end{cases} \quad (0.38)$$

Costs module

The cost module developed for this study includes both direct and indirect cost calculations.

Direct costs and Indirect Costs

Direct costs are calculated based on a cost per case which is constant throughout the simulation.

$$\text{Direct cost (£) per individual (year)} = \frac{\text{Cost per case (£)} \ * \ \text{Prevalence(year)}}{\text{Alive (year)}} \quad (0.39)$$

95% confidence intervals are calculated from the prevalence rates per individual (P) by the equation below.

$$95\%CI = \text{Direct Costs (year)} * 1.96\sqrt{\frac{P(1-P)}{Trials}} \quad (0.40)$$

Intervention costs

The total intervention cost each year is calculated using the following equation (equation (0.41)).

$$\text{Total Intervention Cost}(y) = \text{Annual Intervention Cost} \times \text{N} \tag{0.41}$$

Where N is the number of people who are undergoing the intervention.

Discounting

For some interest rate r% per annum, an amount C_Y in year Y is worth an amount C_{Y+1} in year Y+1 where

$$C_{Y+1} = \left(1 + \tfrac{r}{100}\right) C_Y \tag{0.42}$$

And hence after N years

$$C_{Y+N} = \left(1 + \tfrac{r}{100}\right)^N C_Y \tag{0.43}$$

Conversely, future amounts are worth less when valued at current prices,

$$C_Y = \left(1 + \tfrac{r}{100}\right)^{-N} C_{Y+N} \tag{0.44}$$

In words: N years from now the cost C_{Y+N}, is discounted (discount rate r% pa) to the cost C_Y now.

References

1. Government Office for Science. Tackling Obesities: Future Choices. Project Report. 2nd ed. 2007.

2. Hoogenveen RT, van Baal PH, Boshuizen HC, Feenstra TL. Dynamic effects of smoking cessation on disease incidence, mortality and quality of life: The role of time since cessation. Cost Eff Resour Alloc [Internet]. 2008 Jan [cited 2014 Apr 29];6:1. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2267164&tool=pmcentrez&rendertype=abstract

3. Hoogendoorn M, Feenstra TL, Schermer TRJ, Hesselink AE, Rutten-van Mölken MPMH. Severity distribution of chronic obstructive pulmonary disease (COPD) in Dutch general practice. Respir Med [Internet]. 2006 Jan;100(1):83–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15894477

4. Craig BA, Sendi PP. Estimation of the transition matrix of a discrete-time Markov chain. Health Econ [Internet]. 2002 Jan [cited 2015 Mar 26];11(1):33–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11788980

5. Heistaro S. Methodology Report: Health 2000 Survey. Helsinki; 2008.

6. Kainu A, Timonen KL, Toikka J, Qaiser B, Pitkäniemi J, Kotaniemi JT, et al. Reference values of spirometry for Finnish adults. Clin Physiol Funct Imaging [Internet]. 2015 Mar 27; Available from: http://www.ncbi.nlm.nih.gov/pubmed/25817817

7. Office for National Statistics. Health and life events guidance and metadata [Internet]. [cited 2015 Nov 13]. Available from: http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/index.html

8. World Health Organization. WHO | Disease and injury country estimates. World Health Organization; [cited 2015 Nov 13]; Available from: http://www.who.int/healthinfo/global_burden_disease/estimates_country/en/

9. Mattila T, Vasankari T, Kanervisto M, Laitinen T, Impivaara O, Rissanen H, et al. Association between all-cause and cause-specific mortality and the GOLD stages 1-4: A 30-year follow-up among Finnish adults. Respir Med [Internet]. 2015 Aug;109(8):1012–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26108990

10. Levey AS, de Jong PE, Coresh J, El Nahas M, Astor BC, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. Kidney Int [Internet]. 2011 Jul;80(1):17–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21150873

11. Tajima R, Kondo M, Kai H, Saito C, Okada M, Takahashi H, et al. Measurement of health-related quality of life in patients with chronic kidney disease in Japan with EuroQol (EQ-5D). Clin Exp Nephrol [Internet]. 2010 Aug;14(4):340–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20567874

12. Saito I, Kobayashi M, Matsushita Y, Mori A, Kawasugi K, Saruta T. Cost-utility analysis of antihypertensive combination therapy in Japan by a Monte Carlo simulation model. Hypertens Res [Internet]. 2008 Jul;31(7):1373–83. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18957808