rivm

Detailed description of DYNAMO-HIA calculations

Version 1.0

a Dynamic Model for Health Impact Assessment
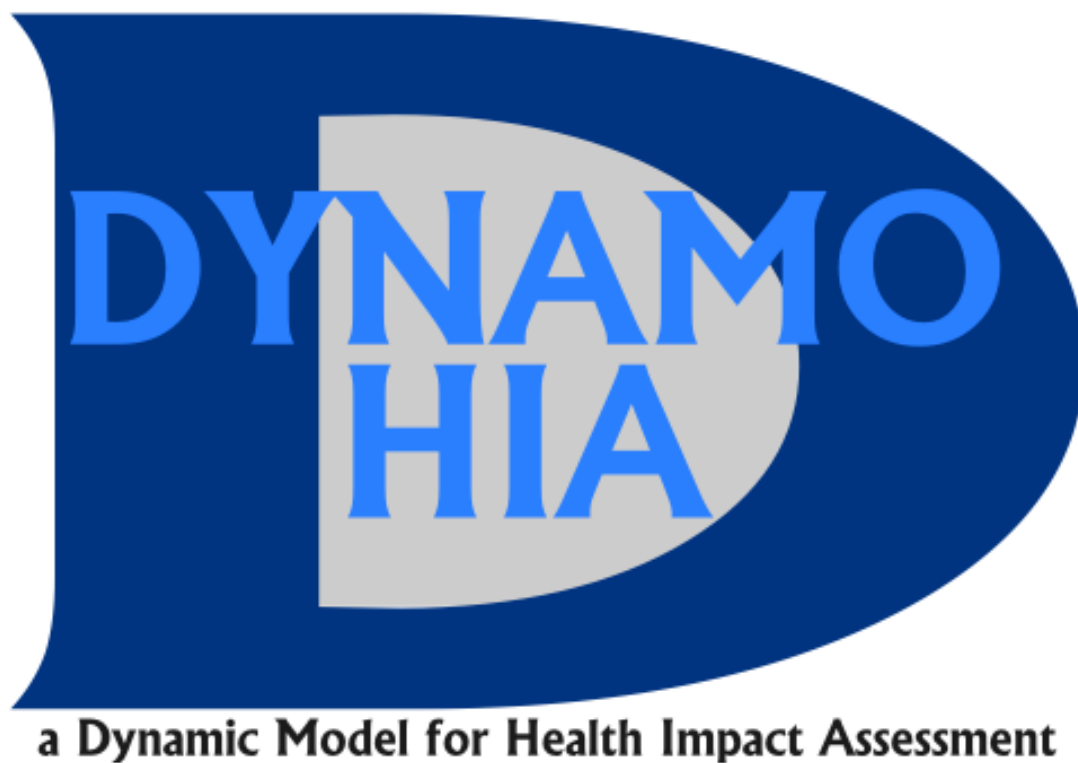
H.C. Boshuizen

October 2010

rivm

1   Table of Contents

2   Introduction and general description of the DYNAMO-HIA Model

This document describes in detail the calculations that are carried out in the DYNAMO-HIA model. This document assumes that the reader already familiar with the general aim and design of the DYNAMO-HIA -model. For a general introduction, we refer to Lhachimi et al.[1] (in preparation). and the DYNAMO-HIA user manual[7], and for a more concise description of the calculation methods to Boshuizen et al.[2] [2](in preparation).

The DYNAMO-HIA model is in the form of a "partial" micro-simulation model. The modelling takes place in the following steps:

1. An initial population of individuals is generated, based on the risk factor exposure data that are given to the model, like percentage of smokers or the distribution of BMI. Also the probability of disease is calculated for each simulated individual in this initial population, based on the data on disease prevalence rates that are given to the model.
2. The risk factor histories of these individuals are simulated over time (simulation module), both in the current situation (business-as-usual scenario, which we will further call the **reference scenario**), and under one or more policy scenarios (further referred to as **alternative scenarios**).
3. From these risk factor histories, and the disease probabilities in the initial state, the probabilities on diseases and mortality for each simulated individual during simulated time are calculated, both in the current situation, and under the policy scenario's.

We call this partial micro-simulation because micro-simulation is used to generate risk factor histories, but not for generating disease histories. The probabilities of disease of each individual are calculated using deterministic methods. This choice is made because incidences on diseases are often small, so large numbers of simulated persons would be

4

needed for a sufficiently accurate simulation. Within the Dynamo model, therefore, using deterministic calculations of disease probabilities is more efficient.

Both the simulation of risk factor histories and the calculation of disease probabilities is based on a Markov Model. A Markov Model is described by states, and transition probabilities between states.

Chapter 2 will describe the states used in the model, while chapter 3 will describe the transition rates that are used and the way the parameters of these transition rates are estimated from the input data. Chapter 4 will describe the generation of the initial population for the simulation, chapter 5 the simulation, and chapter 6 the post-processing calculations, in which the results from the simulated population are used to calculate 1) the number of persons in the real population with and without diseases, and 2) integrated measures as life expectancies, DALEs and life expectancies without disease.

3   Description of the model states

The state of an individual in the DYNAMO-HIA model is determined by the (joint) value of the following characteristics:

1. **<u>Age</u>**

2. **<u>Gender</u>**

3. **<u>Risk factor status</u>**

The risk factor in the model can be one of three forms:

- o A risk factor with a continuous distribution (for instance BMI). In this case the model needs the following input: the shape of the distribution (normal or lognormal), mean of the distribution, the standard deviation of the distribution and in case of the lognormal distribution also the skewness of the distribution. We will further call this **continuous risk factor**.
- o A risk factor in classes (for instance: smokers / former smokers / never smokers). In this case the model needs as input: the percentage in each class. A maximum of 10 classes can be entered in the program. We will further call this **<u>categorical</u> risk factor**.
- o A risk factor in classes (as above), but with the addition that for one class also the duration of being in this class is of interest (for instance in case of smoking the duration of having stopped smoking is of interest). We will further call this **compound risk factor**. At the start of simulation the duration can have the values 0 to 19.

Only one risk factor can be entered in the model. However, the user could use the 10 classes available to define "combination risk factors" like "smoking alcohol drinkers", "non smoking alcohol drinkers" "smoking abstainers" and "non-smoking abstainers" .

4. Disease probabilities

The disease states in the underlying conceptual model for each disease are either 0 (disease not present) or 1 (disease present). In the DYNAMO-HIA model, however, the "state", that is, the value stored during simulation, implemented in the simulation, is not the absence or presence of disease, but the probability of having a disease. This approach, using probabilities, is preferred over the method in which an explicit disease state is simulated for each person, as many diseases are rare, so extremely large numbers of persons need to be simulated to obtain sufficient precision.

We will used the following terminology with regard to diseases:

Independent diseases
With independent we mean that the incidence of these diseases does not depend on the presence or absence of other diseases, given the risk factor history (conditional independence).

Dependent diseases
These diseases can depend on the presence/absence of an independent disease (but in the DYNAMO-HIA model they may not depend on the presence/absence of another dependent disease)

Intermediate diseases or causal diseases
This are independent diseases that are themselves a risk factor for a dependent disease: such a disease has an intermediate role in the causal pathway between risk factors and the

dependent disease. In the DYNAMO_-HIA model a dependent disease can not be an intermediate disease. In other words, the causal network of diseases can only have two layers.

The diseases will be assigned to **disease clusters**. A dependent disease is always in the same cluster as the intermediate diseases that are its risk factors. Independent diseases that are not intermediate diseases form a cluster of one (also referred to as: **single disease**).

For each disease cluster, a person's state will be characterized by the set of probabilities of having each possible combination of diseases within the cluster (conditional on being still alive). For instance, in case of a cluster with diseases A, B, and C, the person will be characterized by the probability of having no disease, of having disease A only, of having disease B only, of having disease C only, of having both disease A and B, of having both disease A and C, of having both disease B and C and of having all three diseases. In the case of an independent disease that is not an intermediate disease (a cluster of one), this will simply be the probability of having this disease (conditional on being alive), or not having this disease. As is shown in Boshuizen et al.[2], the overall probability of all combinations of all diseases can be calculated by multiplying the probabilities from the different clusters.

This can also be seen as constructing a multi-state life table for a group persons that have the same risk factor history as the simulated person.

Furthermore, for diseases for which the user specifies a cured fraction (an option added for better simulation of cancers) the probability of disease is separated in a "cured" fraction and a fraction that in time will die of the cancer. The cured fraction is assumed to be independent of risk factor status or the presence of other diseases.

5. **Survival probability**

Like for diseases, mortality can be a rare event and thus simulation is inefficient. Therefore for each person at each moment in time, the state that is kept track of in the simulation is the survival probability.

4 Estimation of transition rates and occupancy rates of the initial population

4.1 General overview

During simulation, the value of the four types of characteristics (age, risk factor, disease probability, survival probability) is updated at one year intervals. The characteristic sex is constant, so it does not need updating. The basis for the update are the transition rates between the states of the underlying Markov model. With the underlying Markov model we mean the Markov model that has age, gender, risk factor state, disease state and survival state as its states, rather than age, gender, risk factor state, disease probability and survival probability. The underlying Markov Model has the following transitions:

- transitions of age (simply implying that age increases with one year at each update)
- transition rates between different risk factor states
- transition rates from not having disease A to having disease A
- transition rates from being alive to being dead

Note that no transition rate is included from having the disease to not having the disease, so the model does not allow for remission.

4.1.1 Converting transition rates into transition probabilities

The time-step used in the Dynamo model is 1 year, although in principle it could have been any other period. Most input to the DYNAMO-HIA model is in the form of transition rates (for instance, disease incidence and mortality), where a rate is the change in an infinitely small time period. When updating the characteristics of each simulated person, all transition

rivm

rates have to be converted into 1-year transition probabilities. Exceptions are the transitions between risk factor states, where the model expects one-year transition probabilities as input rather than transition rates, and age, where no user input is required.

If only transitions out of a state are possible, conversion from rates to 1 year probabilities can be done by using the formula:

$$P(1 \text{ year})= 1-e^{-rate*1 \text{ year}} \tag{1}$$

where P(1 year) is the 1 year transition probability and rate is the transition rate. However, in cases where there are also transitions into the state, things are more complicated.
A full analytical solution can be calculated by:

$$\mathbf{P}=e^{\mathbf{A}*timestep} \tag{2}$$

where **P** is the matrix of transition probabilities for the states involved and **A** is a matrix of transition rates.

Calculating a matrix exponential is an old numerical challenge[3]. Gallivan and others suggested an algorithm specifically for this type of models[4]. DYNAMO-HIA uses the latter algorithm to solve the equation above.

Furthermore, the computer time required for calculating a matrix exponential increases exponentially with the size of the matrixes involved. In DYNAMO-HIA, we do not include risk factors as a state in this type of calculation, but repeat the calculations for each risk factor state separately, as this makes definition of risk factors and especially of risk factor histories much more flexible. However, the size of the matrix still doubles with every disease added to the model. In order to keep both the dimension of the matrix small, and the calculations transparent, DYNAMO-HIA does not carry out calculations for the entire disease-space, but divides the transition matrix in sections that can be updated independently:

Groups of diseases that are independent of other diseases (conditional on risk factor status) (**disease clusters**) are updated as a group, and calculation of the transition matrix is performed within this group. The updates of the disease states per cluster, and the survival due to mortality from other causes of death (that is, from diseases not included in the model) can be used to calculate every disease state, as well as the overall survival at each moment in time. A more detailed description and the justification of this approach can be found in the technical paper describing DYNAMO-HIA[2].

4.1.2 Overview of transition rates in the model

Table 1 describes the transition rates in the DYNAMO-HIA model. The equations given for the transition rates in this table contain constants such as "baseline incidence". We will further refer to these constants as **(model)parameters**. Some of these constants have to be given by the users (like relative risks on diseases) and are used directly as given, while others are calculated in a parameter estimation module from other data given by the user. In section 3.3 we will describe the way in which we estimate these parameters from the input data.

The element "RR(riskfactor)" in table 1 stands for a function that differs for different types of the risk factors. For the categorical risk factors RR(riskfactor) is equal to the values given in the input. For the other types of risk factors they are calculated as follows:

Continuous risk factors:

$$RR(\text{risk-value}) = RR(\text{per unit})^{\wedge(\text{risk-value-reference value})} \qquad (3)$$

Here the reference value is an arbitrary level of the risk factor for which the RR is set to 1. The user can chose this value. However, choosing a different reference value does not influence results.

Compound risk factors:
For the classes without a duration value: RR(riskfactor) is given directly by the user
For classes with a duration component:

$$RR = RR_{end} + (RR_{begin} - RR_{end}) \exp(- alpha * duration) \tag{4}$$

Here RR_{end}, RR_{begin} and alpha have to be given directly by the user.

Table 1: transition rates in the DYNAMO-HIA model. All transition rate formulas are on the level of the individual, so al probabilities refer to the probability for this individual, not to population probabilities

Type of transition	Quantification of this transition rate
Risk factor state to other risk factor state	Categorical/Compound: Given by user[1] or calculated from input prevalence of the risk factor assuming that risk factor prevalence over age groups does not change in time (see 4.3.2). For compound variables, the duration in the duration class is increase by 1 year at each time step
	Continuous (normally distributed[2]): new value = old value + drift + stdDrift*ε , with Drift =: user specified or calculated from input prevalence of the risk factor assuming that risk factor prevalence over age groups does not change in time (see 4.3.2). stdDrift = calculated from input prevalence of the risk factor assuming that the increase in variance of the risk factor prevalence over age groups does not change in time (see 4.3.2). e = randomly drawn value from a standard Normal distribution
Without disease → with disease (independent diseases)	$Baselineincidence_d . RR_d(riskfactor)$. $RR_d(riskfactor)$ = relative risk on diseased due to risk factor status $Baselineincidence_d$ = incidence rate of diseased in those with $RR_d(riskfactor) = 1$ d = index of disease
Without disease → with disease	$Baselineincidence_d RR_d(riskfactor)\sum_{combi} P(combi) RR_d(combi)$. $RR_d(combi)$ = relative risk on diseased of a particular combination of intermediate diseases $P(combi)$ = probability of the combination of intermediate disease $combi$ = index of each possible combination[2] of intermediate diseases

14

(dependent diseases)	
To death	$RR_{OC}(riskfactor).Baseline\,other\,mortality + \sum_d AM(d)P_{nc}(d,i)$ $+ \sum_d \{RR_d(riskfactor)\sum_{combi}RR_d(combi)Baseline F(d)\}$ $RR_{OC}(riskfactor)$ = relative risk on other mortality due to risk factor status Baseline other mortality = Other mortality rate in those with $RR_{oc}(riskfactor)=1$ $AM(d)$ = Chronic mortality attributable to diseased (Attributable Mortality) $P_{nc}(d)$ = probability of having disease d (excluding cured fraction) Baseline F(d) = Acutely fatal incidence rate of diseased (Fatal) for those with RR(riskfactor)=1

1) Here the input asked or calculated is not transition **rates**, but one-year **transition probabilities**. This means that the conversion from rates to one-year probabilities is not necessary

2) For lognormal distributed risk factors, see see 4.3.2

3) including the combination "having none of the diseases"

The mortality as defined in table 1 can accommodate three disease-mortality processes:

1. For a chronic disease, for which the excess mortality (defined as the difference in mortality rate between those with the disease and those without the disease) only depends on age and gender, but not on how long one has the disease (baselineF(d)=0).

2. For a partly acutely fatal disease: This are diseases (like myocardial infarction or stroke) that occur as a distinct event, that has a very high mortality rate immediately after the event, while those who survive this critical period have a lower mortality, although still higher than the general population. The excess mortality rate after the critical period, like the excess mortality in the first disease process, depends only on age and gender, and not on the duration of the disease (baselineF(d)>0)..

3. For a disease with a cured fraction in which the excess mortality is zero in a part of the patients (cured fraction), while the excess mortality does not depend on duration of disease in the others (Which is equivalent to assuming an exponential survival model with a cured fraction). As the part of the patients that are "cured" can only be identified

in retrospect (after all patients that have not been cured have died), it is not realistic to model "cured" with remission, as also the "cured" patients need the same medical attention as the "non-cured" patients, as it is not known that they are cured. These diseases are modelled by regarding them as two separate diseases, and putting Am(d) to zero for the cured disease. Combinations of fatal fractions and cured fractions are not allowed in the DYNAMO model.

4.2 Estimation of disease probabilities in the initial population

Before giving the methods used to estimate the parameters of the transition rates between states, we first describe the methods used to estimate the disease prevalence rates in the initial population, as these are used to estimate several of these parameters.

The input of the model consists of disease prevalence on the population level. These should be translated into prevalence rates at the individual level. For instance, the model input consists of the prevalence of diabetes, and the prevalence of coronary heart disease on the population level. To generate an initial population, one has to know also the prevalence of having diabetes and coronary heart disease simultaneous, and also how the prevalence rates of diseases are related to the risk factors in the model. This needs to be known before starting the estimation of the modelparameters (the parameters of the transition rates), because several estimation procedures assume that the prevalence rate of diseases on the individual level is known.

The individual prevalence rates can not be inferred from the input data as given by the user to the DYNAMO-HIA model without further assumptions. Here we will throughout make the assumption that the prevalence odds ratio of a disease in two groups is equal to the ratio of the incidence rates of this disease in those groups[5, 6]. This is only true in the case that incidence relative risks and excess mortality are constant (do not depend on age within the period that a disease lasts), which clearly will not always be the case. However, we believe

16

this assumption to be superior to the assumption of independence, which is implicitly used in many other models.

Secondly, for cancers the prevalence needs to be split in the prevalence of "cured" patients, and non-cured patients.

4.2.1   Splitting the cancer prevalence into "cured" and "non-cured" prevalence

For cancers the prevalence needs to be split in the prevalence of "cured" patients, and non-cured patients. In a period with constant incidence and a disease without excess mortality (as cured cancer), the relation between the prevalence at age t_2 and at age t_1 is:

$$p(t_2) = p(t_1)(1 - e^{-incidence*(t_2 - t_1)}) \qquad (5)$$

One option is to use this formula to calculate the prevalence of cured patients at age t+1 from the prevalence at age t, using as incidence at that age the incidence times the cured fraction, starting at age 0 with a prevalence of 0.

This method assumes that there are no cohort effects. In reality, however, it might be the case, due to medical advances, that the current "cured fraction" is larger as it was in the past. In that case this method implies that a percentage of cured cases is used for the past that is too large. This then means that a too large part of the current prevalent cases is considered cured, and as cured patients will remain alive (apart from being subject to general mortality) during simulation, the projected prevalence rates will be too high.

The second option is to calculate the prevalence of non-cured patients at age t+1 (*p(t+1)*)from the prevalence of non-cured patients at age t (*p(t)*) and the incidence of non-cured cancer inc(t) (= incidence * (1-cured fraction)) using:

$$p(t+1) = \frac{(p(t)M(t) - Inc(t))e^{(Inc(t) - M(t))} + Inc(t)(1 - p(t))}{(p(t)M(t) - Inc(t))e^{(Inc(t) - M(t))} + M(t)(1 - p(t))} \qquad (6)$$

where *M(t)* is the excess mortality rate at age t in those with (non-cured) cancer and *Inc(t)* the incidence at age t. M(t) is given by the user, or it is calculated from the median survival.

This approach is less sensitive to cohort effects as the first option, as only cohort effects during the "dying period" after incidence are relevant. For most cancers the median survival time in those not cured is relatively short, so cohort effects will play only a limited role. We therefore implemented the second method in the DYNAMO-HIA model.

4.2.2   Calculating excess mortality rate from median survival

To calculate the excess mortality rate *M(t)* from the median survival, we assume that for the highest age class in the model (95+) M(t) can be calculated as:

$$M(t) = \frac{\ln(2)}{medianSurvival} \quad (7)$$

Once M(t) is know for a range of ages up to the highest age, it can be found for the next lower age by solving the equation:

$$e^{-M(t+\text{int}(median\ Survival(t))*\text{mod}(median\ Survival(t))} \prod_{s=t}^{s=t+\text{int}(MedianSurvival(t))-1} e^{-M(s)} = 0.50 \quad (9)$$

where int(T) is the integer part of T, and mod(T) the non-integer part. Example: int(2.34)= 2 and
mod(2.34)=0.34.
In cases where *M(t)* changes very fast with age, it is possible that this equation has no solution.

18

In that case the following error message is given: "median survival rates for age X and sex 0/1 for disease "diseaseName" are inconsistent with the median survival at the next higher age groups. Please check the data and make sure that survival does not decrease abruptly over age".

4.2.3 Calculating the probability of disease in an individual

As all input in the DYNAMO-HIA model is specified by age and gender, all calculations are carried out separately for each combination of age and gender. For clarity, however, we will leave out age and gender indexes in all the following descriptions.
The general approach here is that the prevalence odds on disease in an individual can be calculated as a relative risk times a baseline prevalence odds for that disease, making the assumption that the prevalence odds ratio of a disease in two groups is equal to the ratio of incidences rates of that disease in those groups[5, 6]. The ratio of the incidence rates are the relative risks that are given by the user. The baseline prevalence odds can be calculated by constraining the overall prevalence rate in the population (calculated by using this baseline odds for each individual and than averaging over the population), to be equal to the input prevalence rate (as given by the user). This approach is base on that used in the RIVM chronic diseases model[7]. However, in that case the assumption is that the ratio of probabilities rather than the odds is equal to the ratio of incidence ratios.

Calculation of prevalence odds for independent diseases:

Step 1:finding an initial estimate

Find an initial estimate by assuming that the prevalence relative risk in each risk factor group is equal to the incidence relative risk, and then calculate a baseline prevalence rate which is used as an initial estimate for an iterative procedure to find the baseline prevalence odds. To

do so, we calculate the mean relative risk in the population, and from that the baseline prevalence rate.

The mean relative risk in the population is calculated as:

For categorical risk factors:

For each disease:

$$RRmean = \sum_c RR_c P(c) \tag{9}$$

where P(c) is the proportion of the population having class c of the risk factor and RR_c is the relative risk on the disease for this class.

For compound riskfactor

Generate a sample of $19+N_r$ persons (20 persons in the class which has 20 duration categories (<1 year, 1-2 years,…. ,19 years and more) and N_r-1 in the other classes). Calculate the proportion of the population in each of these categories P(duration) using the risk factor prevalence and the distribution of the duration part of the compound risk factor as given by the user. Calculate the relative risk in each duration class:

$RR_i(duration)= RR_{end}+(RR_{begin}-RR_{end})exp(-$ beta $*duration)$

where i is an index for each simulated person, duration is the value of the duration (0 for <1 year, 1 for 1-2 year, 19 for 19 years or more).

Then

$$RRmean = \sum_{c \neq durationClass} RR_c P(c) + \sum_i P(duration)RR_i(duration) \tag{10}$$

20

where P(c) is the proportion in class c, P(duration) is the probability of being in the duration class with the particular duration, RR_c the relative risk in class c and RRi(duration) is the relative risk for subject i based on the duration given above.

For continuous risk factors:

Generate a sample of Nsim persons following the distribution of the risk factor. Nsim is equal to the number of simulated persons in the simulated population (as given by the user). However, in the parameter-estimation module this is increased to 100 in case a lower number was given.

To generate such a sample for a continuous risk factor, we use the mean, standard deviation and skewness of the risk factor given as input values.

For the <u>normal</u> distribution, mean and standard deviation define the normal distribution and skewness is zero. If skewness is non zero, and a normal distribution has be specified, the distribution is assumed to be log-normal. In that case a warning is written to the log file stating:

"normal distribution asked, but as skewness is not equal to zero, lognormal distribution is used".

In the reverse case, where a log-normal distribution is specified, but for all ages the skewness is zero, a normal distribution is used, and a the following warning is written to the log file:

"log-normal distribution asked, but as all skewness are zero, normal distribution is used"

For the <u>lognormal</u> distribution we have:

$$mean = \exp(\mu + 0.5\sigma^2)$$
$$variance = [\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2) \qquad (11\text{-}13)$$
$$skewness = [\exp(\sigma^2) + 2]\sqrt{\exp(\sigma^2) - 1}$$

From the skewness we can calculate the parameter σ, then from σ and the input variance we can calculate μ and from that the mean. This is the mean for a log-normal distribution starting at value 0. However, in our case we might want to use a lognormal distribution starting at a higher value (for instance, BMI will never have value 0, but will be at least 10). So the difference between the mean as given by the user and the mean calculated from the variance and skewness will be used as an offset value for the mean, that is the value at which the log-normal distribution will start.

Generating the sample of nSim persons with a continuous risk factor value R(i) will be done using the following algorithm:

1. For i=1 to Nsim generate CP(i)= (i-0.5)/Nsim
2. Calculate R(i)=F^{-1} (CP(i)), where F^{-1} be the inverse of the cumulative probability function of the risk factor (normal or log-normal)
3. Calculate RR_i as RR_i=RR(per unit)$^{\wedge(R(i)\text{-reference value})}$

And then

$$RRmean = \frac{1}{Nsim}\sum_i RR_i \qquad\qquad (14)$$

After estimation of RRmean a first estimate of the baseline odds is:

$$Baselineodds(d\ ) = RRmean(d\ ) / odds(d) \qquad\qquad (15)$$

Step 2: calculate probability of the disease given the baseline odds

From the trial baseline odds value calculate for each class or simulated person:

$$P_c(d \mid c) = \frac{RR_c(d) * Baselineodds(d)}{RR_c(d) * Baselineodds(d) + 1} \qquad (16)$$

or

$$P_i(d \mid i) = \frac{RR_i(d) * Baselineodds(d)}{RR_i(d) * Baselineodds(d) + 1} \qquad (17)$$

Step 3: make overall prevalence equal to the sum of the simulated probabilities

Find the value of the baseline odds that gives (for categorical, compound and continuous risk factors, respectively):

$$P(d) = \sum_c P_c(d)p(c)$$
$$P(d) = \sum_c P_c(d)p(c) + \sum_{duration} P_{duration}(d)p(duration) \qquad (18\text{-}20)$$
$$P(d) = \frac{1}{Nsim} \sum_i P_i(d)$$

using the Newton-Raphson algorithm, and where P(d) is the population prevalence of disease d as given by the user.

Calculation of baseline odds for dependent diseases:

The procedure to calculate the baseline odds for these diseases is largely the same as for independent diseases, only the relative risks for a simulated group of person are calculated differently, as they must include also the dependency on the intermediate disease.

To include this dependency, we need the prevalence of the intermediate disease as a function of risk factor levels. For these, we will use the same approximation as above, namely that for each disease the prevalence odds ratio's are equal to the incidence rate ratio's.

The odds of an intermediate diseases d_{int} can be calculated from the baseline odds of d_{int} and the relative risks.

For a class risk factor or a compound risk factor (where c here and further on will include each different length of stay in the duration class as a separate class) this is:

$$P(d_{int} \mid c) = Odds(d_{int}, c)/(Odds(d_{int} \mid c) + 1) =$$
$$baselineOdds(d_{int})RR_c(d_{int})/(baselineOdds(d_{int})RR_c(d_{int}) + 1)$$

(21)

and for a continuous risk factor this is:

$$P(d_{int} \mid i) = Odds(d_{int} \mid i)/(Odds(d_{int} \mid i) + 1) =$$
$$baselineOdds(d_{int})RR_i(d_{int})/(baselineOdds(d_{int})RR_i(d_{int}) + 1)$$

(22)

For each disease calculate a trial value of the prevalence odds by calculating RRmean as:

For categorical and compound risk factors:

$$RRmean = \sum_c RR_c(dep)P(c) \prod_{d\,int} [\, P(d_{int} \mid c)(RR_{d\,int}(dep) - 1) + 1]$$

(23)

where P(c) is the proportion of the population having class c of the risk factor and $RR_c(dep)$ is the relative risk on the particular dependent disease for this class, d_{int} are the indexes for the relevant intermediate diseases, $RR_c(d_{int})$ is the relative risk on the intermediate disease d_{int} for this class, and $RR_{dint}(dep)$ is the relative risk of the intermediate disease on the dependent disease.

For continuous risk factors:

24

Use the sample of nSim persons as generate above.

Then calculate for each disease:

$$RRmean = \frac{1}{nSim}\sum_i RR_i \prod_{d\,\text{int}} [P(d_{\text{int}} \mid i)(RR_{d\,\text{int}}(dep) - 1) + 1] \qquad (24)$$

The trial version of the baseline odds can be calculated from this RRmean again using

$$Baselineodds(d\ ) = RRmean(d\ )/P(d)$$

where P(d) is the population prevalence of the disease as given by the user.

Once a baseline odds is known, we can calculate for each class or simulated person the probability on d   (P(d|c) or P(d|i) depending on the type of risk factor) by averaging the probability on d given each particular combination of intermediate diseases $j_1 \ldots j_n$ (thus weighting by the probability of the combination of diseases):

$$P(d \mid c) = \sum_{j_1=0}^{j_1=1}\sum_{j_2=0}^{j_2=1}\cdots\sum_{j_n=0}^{j_n=1} \frac{RR_{dis}^* * RR_c(d) * baselineodds(d)}{RR_{dis}^* * RR_c(d) * baselineodds(d) + 1}\prod_{m=1}^{m=n} P(d_m = j_m)RR_{dis}(d \mid d_m = j_m) \qquad (25)$$

where

$$P(d_m = 1) = \frac{RR_c(d_m) * Baselineodds(d_m)}{RR_c(d_m) * Baselineodds(d_m) + 1} \qquad (26)$$

$$P(d_m = 0) = 1 - \frac{RR_c(d_m) * Baselineodds(d_m)}{RR_c(d_m) * Baselineodds(d_m) + 1} \qquad (27)$$

$$RR_{dis}^* = \prod_{m=1}^{m=n} RR(d \mid d_m = j_m) \qquad (28)$$

$RR_{dis}(d|d_i=1)$ is the relative risk on disease d for someone with disease d_i relative to someone without d_i,

$RR_{dis}(d|d_i=0)=1$.

For a continuous risk factor the same applies, changing the index c to i.

Then the Newton-Raphson algorithm is again applied to estimate the baseline odds ratio that yields in the simulated population exactly the input population prevalence.

4.3   Estimation of the parameters of the transition rates.

In the equations in table 1, the probability terms P are values for an in an individual at a particular moment in time (note that in this table they are not the population prevalences, as in the previous paragraph): these probabilities are stored during simulation as the "DYNAMO state values" of the simulated individuals. The RR(risk factor) values are calculated from the risk factor state, as given on page 12. All other terms, as well as the RR values necessary to calculate RR(risk factor) are **model parameters.** The model assumes that the model parameters are constant over time, but can differ by (attained) age and gender. The model parameters have to be supplied to the central core of the model, that uses them to calculate the transition rates (using the equations given in table 1).

Table 2 show the parameters of the transition rates. The model parameters "relative risks on diseases" and (if user-specified) "transition rates of the risk factor" are directly derived from input given by the user. The other model parameters (Baseline incidence rates, Baseline other mortality rates and Baseline fatal disease rates, the relative risks for other mortality, the attributable mortality rates, non user-specified risk factor transition rates, drift of the risk factor variation and offset, and non-user-specified risk factor drift) need to be estimated first. Below we will describe the estimation of each of those parameter.

rivm

All data are specified by age and gender. Therefore the estimation is also performed by age and gender. We will omit these in the description to simplify notation.

Table 2: Parameters of the transition rates

Type of transition rate	Parameter	Calculated from
Risk factor	Transition probability (given)	Given as input
	Transition probability (calculated)	Risk factor prevalence rates, RR for all cause mortality Baseline all cause mortality
	meanDrift, stdDrift offSetDrift (continuous risk factor)	User specified drift (optional), Risk factor mean, standard deviation and skewness, RR for all cause mortality Baseline all cause mortality
Without disease – with disease (independent diseases)	RR(riskfactor)	Given in input (categorical risk factors) or calculated from input RR information and current risk factor status. See page 12
	Baseline incidence	Prevalence of risk factor Relative risks for risk factor, see page 12
Without disease – with disease (dependent diseases)	RR(riskfactor)	Given in input (categorical risk factors) or calculated from input RR information and current risk factor status, see page 12
	Baseline incidence	Prevalence of risk factor Relative risks for risk factor (page 12) Prevalence of intermediate disease Relative risks on intermediate disease for risk factor
	RR_{id}	Given as input

t

where Fatalinc$n$ in the average fatal incidence of disease n in the population, and M the overall mortality rate in the population.

This gives us a linear system of n+1 equations (n= number of diseases), containing n am(d)-parameters and Moc as the parameters that need to be estimated. Gathering all am and the Moc terms at one side of the equation, we get a linear system of equations, in which only the am(d) and Moc values are unknown, and which are solved using standard linear algebra, after calculating the fatal incidence terms Fatalinc*n,* by averaging the terms Fatalinc*n(r)* (calculated in the previous section) over the population.

Estimation in case excess mortality is zero.

If attributable mortality is zero, a small positive excess mortality will result, because those with disease have a worse risk factor profile as those without the disease (unless the disease does not depend on the risk factor, but then it would not be needed in the model).
If we reverse this reasoning, then if an excess mortality of zero is specified, a negative attributable mortality will be calculated. DYNAMO-HIA assumes that in this case the user means to request a zero attributable mortality instead of a zero excess mortality, and sets the attributable mortality to zero. In that case, for every disease for which the attributable mortality is put to zero, the system of equations will have one less equation. Estimation of the other attributable mortalities and of other cause mortality remains the same. Also in cases where the user specifies a excess mortality that is lower than the excess mortality resulting from a zero AM, the AM is set to zero.

4.3.1   Estimation of baseline other mortality and the relative risks for other mortality

In the previous section, we estimated the other mortality for each risk factor class.

For <u>categorical risk factors,</u> the other mortality in the reference category (where the relative risk is equal to 1), equals the baseline other mortality. The RR for other mortality then can be calculated simply as:

$$RR\, other\, mortality(c) = \frac{Moc(c)}{Baseline\, other\, mortality} \qquad (46)$$

For <u>continuous risk factors</u>, we fit a regression model through the estimated values of Moc(r) in order to estimate a single relative risk value on other mortality per unit of exposure (= per unit of the risk factor) as needed by the model.

Hereto we regress the log of the estimated other mortality values Moc(r) on the risk factor values R(i). The exponent of the intercept from this regression then gives the baseline other mortality, the exponent of the regression coefficients of the risk factors give the relative risks.

As this procedure forces a particular shape on the data, the average relative risk estimated by the regression, will be slightly different from the average relative risk estimated in the first part of this procedure, and because of that also the overall mortality calculated from these relative risks will not be exactly equal to the overall mortality as given by the user. Therefore the baseline other cause mortality is calibrated so that the total mortality in the simulated population becomes equal to the input total mortality.

For <u>compound risk factors</u> we first calculate the baseline other mortality rate and the relative risks in the same way as for the categorical risk factors, ignoring the duration information. Second we use non-linear regression to estimate the relative risks (at the begin and at the end) and the alpha parameter for the category with the duration component by fitting the following model:

$$RR_{om} = RR_{om\text{-}end} + (RR_{om\text{-}begin} - RR_{om\text{-}end})\exp(-\,alpha_{om}\,*duration) \qquad (46)$$

As for the continuous risk factor, the average relative risk estimated for the category with the duration component by the regression will be slightly different from the relative risk estimated in the first part of this procedure, so also here the baseline other cause mortality is recalibrated so that the total mortality in the simulated population becomes equal to the input total mortality.

In this procedure we also apply some restrictions on RRend and RRbeing: When RR-end for total mortality (as given by the user) is equal to the RR of total mortality in one of the other classes (without a duration component), RR-end for other mortality is restricted to be equal to RR-other mortality for this class(es).
Similarly, when RR-begin for total mortality is equal to the RR for total mortality of one of the other classes (without a duration component) RR-begin other mortality is restricted to be equal to the RR for other mortality for this class. In case both RR-end and RR-begin for total mortality are equal to an RR from other classes, both are restricted to be equal to those RRs, and only alpha is fitted.
In some cases, however, such a restriction is not possible. For example, when the RR-begin is fixed to 0.85, and the RR-end to 1, while the RRs in the duration category on average are below 0.85 or above 1. In that case the restriction on RR-begin is removed. In the case that regression does not return an estimate even after lifting this restriction, also the restriction on the RR-end is removed. In case these restrictions are lifted for these reasons, messages are written to the log-file.

Another possible complication is that the user puts all persons in the starting population in a single duration class. In that case no estimation of time dependence is possible, and RR-end is made equal to RR-begin, implying no time dependence. In that case a warning message is given:
"100% of the initial population has the same duration. Therefore it is not possible to estimate a time dependent other mortality or other disability. In case other mortality/disability is requested, those relative risks will be made constant over time."

A possible complication for all types of risk factors is that other mortality(i) could become negative in case of high am(d) or Fatalinc(d) values in combination with simulated subjects with high P(d|i) values or $RR_d(i)$ values. If this only occurs sporadically this is handled by setting other mortality(i) to zero for categorical risk factors, and ignoring this class when fitting the other mortality in case of a continuous risk factor or a time in the risk factor class. The baseline other mortality then also is recalibrated, assuring that the total other cause mortality in the population remains valid.

However, if negative other cause mortality values occur frequently, this indicates that the combination of excess mortality rates and disease prevalence rates entered by the user leads to more mortality in the population than is physically possible based on all cause mortality rates.

The following warning will be given if negative mortality occurs in more then 30% of the simulated cases: "Warning: Disease related mortality is too large relative to all cause mortality".

4.3.2   Estimation of Non user-specified risk factor transition rates (nett transition rates)

4.3.2.1 For a categorical risk factor

The program can estimate 1 year risk factor transition probabilities from the risk factor prevalence rates. This option is called "nett transition rates". The procedure used is described in more detail in van de Kassteele et al.[8]. The general idea behind this estimation is that it estimates transition probabilities that will keep the age specific prevalence rate of the risk factor stable over time. That is, in future years the age distribution of the risk factor is assumed to be the same as the current distribution by age. In that case we can say that if prevalence rates in the next higher age group differs from prevalence rates in the current age group, people must have switched from one class to another. It is however impossible to

estimate the transition probabilities back and forth from cross sectional data, we can only estimate the nett transition probabilities (with 2 risk factor classes A and B, the nett effect is the difference between the flow from A to B and that from B to A, thus the minimum number of persons that have to change risk factor level in order to obtain the new prevalence rate).

The estimation of transitions for such cases can be formulated as a transportation problem, a topic solved in operations research. A transportation problem basically attempts to find best way to fulfil the demands of a number of demand points using the supplies from a number of supply points, under the consideration of the costs involved in shipping the product from a supply point to a demand point.

A transportation problem is specified by the supply, demand, shipping costs, a decision variable, and an objective function. In this case the supplies are the predicted class prevalence rates $\widetilde{\mu}_0$ based on the prevalence rates in the current year. It is a vector of length n. The demands are the class prevalence rates $\widetilde{\mu}_1$, also a vector of length n, in the next year. The costs $\mathbf{C}$, an $n \times n$ matrix, are the costs involved in moving from one class to another. The decision variable $\mathbf{x}$, also an $n \times n$ matrix, contains the transitions between classes from row i to column j, and is the variable of interest. The objective function J is the total shipping cost from classes in the current year to classes in the next year. This function has to be minimized:

$$\min \ J = \sum \mathbf{C} * \mathbf{x}, \tag{47}$$

subject to

$$\sum_{j=1}^{n} \mathbf{x}_{\bullet,j} = \widetilde{\mu}_0, \ \sum_{i=1}^{n} \mathbf{x}_{i,\bullet} = \widetilde{\mu}_1, \text{ and } x_{i,j} \geq 0, \text{ for } i, j = 1 \ldots n. \tag{48}$$

The above constrains entail that the row sums of $\mathbf{x}$, i.e. people from a specific classes, must be equal to the supply $\widetilde{\mu}_0$. The column sums of $\mathbf{x}$, i.e. people to specific classes, must be equal to the demand $\widetilde{\mu}_1$. This also means that the total supply must be equal to the total demand. The last constrain entails that no transition may be negative.

Some assumptions have to be made about the cost matrix **C**, since no direct "costs" can be assigned to people moving from one class to another. It is assumed that most people tend to stay in their class. Some will switch one class up or down, some will switch two or even more classes, however we assume this becomes very unlikely. We therefore assign zero costs to transitions to the same class, a cost of one unit for one class up or down, a cost of three units for two classes up or down, and so forth. This choice of costs will assure that it is more likely that two persons change one class, then one person changing two classes, as the latter costs 3 units, and the first 2 times 1 unit.

The transportation problem can now be easily solved by a linear programming algorithm, for which we use the simplex method. Once the transitions **x** have been found, the transition probability matrix **M** is then given by

$$\mathbf{M} = (\mathbf{x} / \tilde{\mathbf{\mu}}_0)^{\mathrm{T}}. \tag{49}$$

so that a new state is given by $\tilde{\mathbf{\mu}}_1 = \mathbf{M}\tilde{\mathbf{\mu}}_0$. A transition probability matrix is calculated for each age.

In the method given above selective mortality is included by not using $\tilde{\mathbf{\mu}}_0$, the prevalence of the current age, in the algoritm, but $\tilde{\mathbf{\mu}}_0*$, the hypothetical prevalence after one year, taking mortality into account, but not letting persons change state:

$$\tilde{\mathbf{\mu}}_0* = \frac{(\exp(-\mathbf{RR}*baselineMort))\tilde{\mu}_0}{\sum_j \exp(-RR_j*baselineMort)\tilde{\mu}_{0j}} \tag{50}$$

For RR_j the user specified relative risks are used.

34

For a compound risk factor

For a compound risk factor the same method is used as given for a classified variable. However, here we first need the RR for all cause mortality for the class with duration. This RR is calculated as:

$$RR_{duration} = \sum_d RR_d P(d) \hspace{4cm} (51)$$

were P(d) is the probability of being in duration class d, given that one is in the class with duration.

In principle the duration information itself could also be used to calculate nett transition probabilities. However, information on duration is often much more difficult to obtain, and expert opinion might be used instead of real data. Therefore this road was not taken.

For a continuous risk factor

For a continuous risk factor a similar method is used. For a continuous risk factor we do not have separate states, therefore it is more intuitive to look at it from the point of the update rule applied to an individual.

The update rule for a **normally distributed** variable with mean μ_0 at the current age and μ_1 at the next age and matching standard deviations σ_0 and σ_1 is:

$$X(1) = X(0) + (-\mu_0) + stdDrift*\varepsilon \quad if \; \sigma_1 > max(\sigma_0) \hspace{2cm} (52)$$
$$X(1) = X(0) + (\mu_1 - \mu_0) \hspace{2cm} if \; \sigma_1 <= max(\sigma_0)$$

Where ε is a randomly generated draw from the standardized Normal distribution, and the parameters of this update rule are the "drift" in mean $(\mu_1 - \mu_0)$ and stdDrift is a constant with which the randomly drawn value of ε is multiplied. The latter is put to zero if the standard deviation at a particular age is equal or smaller to that on the previous age. The effect of this way to update the continuous risk factor is that in case the standard deviation of the risk factor distribution is increasing over age, this increase is reproduced by the random

component. Note that this procedure assumes that the standard deviation has been smoothed over age. Using unsmoothed figures will result in stdDrift values that are too large.

The drift in the mean causes all individuals to increase their risk factor value by the same amount: that means there is, based on only the mean drift (and without the random component), perfect tracking over age.

Values of the mean drift are calculated directly from the mean of the risk factor by age: The mean drift can be calculated as the difference of the risk factor mean at the particular age and the next age. Like with the nett transitions for the categorical risk factor, we first adjust the mean at the age for selective mortality by calculting μ_0*, the mean of the distribution of the risk factor x after applying the risk factor specific mortality rate for one year:

$$\tilde{\mu}_0 * = \frac{\int x \exp(-RR(x)baselineMort)P(x)dx}{\int \exp(-RR(x)baselineMort)P(x)dx} \tag{53}$$

Similarly, we also calculate the standard deviation σ_0*, the standard deviation of the distribution of risk factor x after applying the risk factor specific mortality rate for one year:

$$(\tilde{\sigma}_0 *)^2 = \frac{\int (x - \tilde{\mu}_0 *)^2 \exp(-RR(x)baselineMort)P(x)dx}{\int \exp(-RR(x)baselineMort)P(x)dx} \tag{54}$$

The stdDrift then in calculated as:

$$stdDrift = \sqrt{\sigma_1^2 - (\sigma_0 *)^2} \tag{55}$$

The standard error drift is made the same for all scenario's, the user can only change the amount of the meandrift.

The numerical integration procedure used in Dynamo uses 100 equidistant intervals between μ_0-4σ_0 to μ_0+4σ_0. Preliminary simulations showed that this range yields the most accurate results given the use of 100 intervals, and is superior to using equal probability intervals.

For a **lognormally** distributed risk factor, the risk factor value is first logtransformed into a normally distributed variable. The update rule on the transformed variable is identical to that given above for a normally distributed risk factor. After applying this part of the update rule it then is back transformed. Drift and change in variance on the transformed can be calculated from the input data in exactly the same way as for normally distributed variables, but after logtransformation. If there is an offset, a possible "offset drift" is added after back-transformation. The offset drift can be calculated as the difference of the offsets of two consecutive years.

Both the offset drift and the standard error drift are the same for all scenario's, the user can only change the amount of the meandrift.

The <u>meandrift</u> should always be given by the user as the <u>absolute</u> change (increase/decrease) in the mean value, also in the case that a log-normal distribution is specified for the data. In case of a log-normal distribution, the program calculates the amount of change needed on the logscale.

This way to parametrize the update rule means that during the first step of simulation, the risk factor value changes in the way the user will expect, that is, it increases with the amount given. In case of the log-normal distribution, the model uses the relative change of the amount (above the offset) to update the risk factor. So when the mean value of the risk factor at a particular age changes during simulation (for instance due to extra selective mortality in the scenario), the absolute drift will change, as a relative change is applied by the model.

Therefore, using the lognormal distribution could lead to behaviour that is not always intuitive for the user.

Selective mortality is taken into account in the same way as for the normally distributed risk factors, that is by calculating the parameters of the distribution of the risk factor on the log scale after taking one year survival into account. For this calculation we assume that the offset is not affected by mortality.

rivm

5 Generation of Initial Population

5.1 Population characteristics

Each person in the simulation population has the following characteristics:

A   age

G   gender

C   value $\{1,2,\ldots,n_r\}$ of classified risk factor: present if risk factor type = classified or compound

R   value of continuous risk factor, or the length in the duration class: present if risk factor type = continuous or compound

DM$_1$ to DM$_{nc}$ Disease probability tables for each cluster of diseases (1 to n_c). For non-cancers each table has size $2^{\wedge}(\text{NinCluster}(c))$. Each value will be indicated by $P(i,j)(j<=1)$, giving the probability of the disease combination conditional on being alive. Therefore the sum of the values of a single table is always 1, and therefore only $2^{\wedge}(\text{NinCluster}(c))-1$ values need to be stored.

For diseases with a cured fraction (non-lethal cancers) the table has size 2. For independent diseases (not related to other diseases) the table has size 1.

S   Probability of survival.

Input needed for the generation a an initial population

	Input needed
Input on structure	Number of disease clusters, number of diseases in each cluster, whether a cluster

	concerns a disease with a cured fraction, type of risk factor (compound, continuous, categorical) nSim: numbers in the initial population per age and gender stratum, minimal and maximum age of individuals at the start of simulation
Input Parameters (directly given by the user)	Relative risks on diseases RR(r) , RRdis(di,dd)
Intermediate Parameters	Baseline prevalence odds(d) (see 3.2.3),Cured prevalence fraction (d) (see 3.2.1)

The initial population is generated by:

1 constructing a distribution of risk factors that is close to that of the population being simulated.

2.calculating the probability of disease in each simulated individial based on the risk factor state of this person.

At the start of the simulation all persons are alive, so the survival probability is set to 100%.

Below we will describe the algorithms used for the generation of the initial population in detail.

5.2   Generation of characteristics for the reference population

The reference population is the population to which no intervention is applied. For this population the characteristics are generated as described below.

Generation of A and G:

For each combination of A and G in the run, we create Nsim persons.

Generation of C:

We will generate a number of subjects in each risk factor class close to the real prevalence rates using the following algorithm:

1. For each category: Calculate the number of subjects in this category from nSim, rounded downwards: $N_c = $ Int(P(c)*nSim), but make this minimally one person
2. Calculate the number of missing subjects in the simulation: n_todo=nSim- Sum(N_c)
3. Calculate for each c: P_todo_c=(P(c)*nSim - Int(P(c)*nSim))/n_todo ; This is the proportion of "todo" subjects that should go to class c;
4. Randomly draw the n_todo persons using P_todo_c.

Generation of R:

Generate a sample of Nsim persons following the distribution of the risk factor using the following algorithms:

For continuous risk factors:

1. For i=1 to Nsim generate CP(i)= (i-0.5)/Nsim
2. Calculate R(i)=F^{-1} (CP(i)), where F^{-1} be the inverse of the cumulative probability function of the risk factor

For compound risk factors:

Generate C as given above for categorical risk factors, but include the 20 duration classes as part of P(c).

Make R = 0 for all classes without duration, and R=0 to 19 for the duration class.

Generation of DM

Once we have simulated a risk factor value for each person in the simulation, we use this value to generate the initial probability of disease. Below we give the detailed algoritms used for this

Generation of DM for a single independent disease:

1. Calculate RR_i as :

$$RR_i = RR(\text{per unit})^{\wedge(R(i)-\text{reference value})} \text{ (continuous)}$$

$$RR_i = RR(c) \text{ (categories or non-duration categories)}$$

$$RR_i = RR_{end} + (RR_{begin} - RR_{end})\exp(-\text{ beta }*R(i)) \text{ (duration categories)}$$

2. Calculate Odds(d, i) as BaselineOdds(d)*RR_i,
3. Calculate P(d,i) as Odds(d,i)/(Odds(d,i)+1)

Generation of DM for diseases with a cured fraction:

1. Check whether this cancer is independent of other diseases
2. If not, ask the user to change this
3. Calculate RR_i as

$$RR_i = RR(\text{per unit})^{\wedge(R(i)-\text{reference value})} \text{ (continuous)}$$

$$RR_i = RR(c) \text{ (categories or non-duration categories)}$$

42

$$RR_i = RR_{end} + (RR_{begin} - RR_{end})\exp(- \text{ beta } *R(i)) \text{ (duration categories)}$$

4. Calculate P(d-total,i) as above
5. Calculate P(d-notcured) as P(d-total,i)(1-CuredPrevalenceFraction(d))
6. Calculate P(d-cured, i) as P(d-total,i)*CuredPrevalenceFraction(d)

<u>Generation of DM for a cluster of dependent diseases</u>**:**

1. With n diseases in the cluster (clustersize=n), there will be $2^\wedge d - 1$ characteristics in the simulation to characterize the disease state of the cluster. The combination: no diseases is 1-sum(all other combinations), so it is not explicitly stored.
2. Calculate RRi for each d as

    $RR_i(d) = RR(d,\text{per unit})^\wedge(R(i)-\text{reference value})$ (continuous)

    $RR_i(d) = RR(c,d)$ (categories or non-duration categories)

    $RR_i(d) = RRend(d) + (RRbegin(d) - RRend(d))\exp(- \text{ beta } *R(i))$ (duration categories)
3. Loop through all combinations of diseases in the cluster and calculate the probability of each combination, which is :

$$P(combi) = \prod_{dep} P(dependent\,disease\,|\,all\,independent\,diseases) \prod_{indep} P(independent\,disease)$$

where P((in)dependent disease) is the probability that the disease has the value it has in the combination.

P(dependent disease | all independent diseases) and P(independent disease) can both be calculated in with the following equation, given that RRdis(d1=1| d2=1) (the relative risk on disease 1 given disease 2) is always 1 in cases of independent diseases:

P(d=1|other diseases in cluster) =

$$For\ d=1:\frac{BaselineOdds(d)*RR_i(d)\prod\limits_{d1\in cluster}(d1=1\ in\ combi)RRdis(d=1\,|\,d1=1)}{1+BaselineOdds(d)*RR_i(d)\prod\limits_{d1\in cluster}(d1=1\ in\ combi)RRdis(d=1\,|\,d1=1)}$$

$$For\ d=0:1-\frac{BaselineOdds(d)*RR_i(d)\prod\limits_{d1\in cluster}(d1=1\ in\ combi)RRdis(d=1\,|\,d1=1)}{1+BaselineOdds(d)*RR_i(d)\prod\limits_{d1\in cluster}(d1=1\ in\ combi)RRdis(d=1\,|\,d1=1)}$$

Generation of S:

Value S will be 1 for every person in the initial population

5.3 Generation of characteristics for scenario populations

Dynamo offers three possibilities for scenarios:

1. Using the same initial population as for the reference population, but changing the transition probabilities for the risk factor
2. Using the same transition probabilities for the risk factor, but changing the risk factor distribution in the initial population
3. Changing both the risk factor distribution in the initial population, and the the transition probabilities for the risk factor

In the first situation, the initial population of the reference scenario is simply copied.

In the second situation, for <u>continuous </u>risk factors, the model assumes that the ranking of persons within the population remains the same. So the individual with the xth rank

in the old distribution (that is, the distribution in the reference population) gets the value of the xth rank in the scenario distribution.

For <u>categorical and compound risk factors</u>, subjects need to change risk factor status. In this case, we calculate the minimum number of persons that need to change risk factor status in order to reach the new risk factor distribution, using the same algorithm as is used to calculate net transition rates (see section 0).

For each non-zero risk class transition, we generate scenario subjects by applying each transition to every individual in the "old" risk factor class to which that transition applies. An example might clarify. Say there are 100 individuals in the reference simulation population, and the old prevalence is 50% in class 1, 40% in class 2 and 10% in class 3, and the new prevalence is 20% in class 1, 20% in class 2 and 60% in class 3. Than we assume that 20 individuals originally in class 1 have to go to class 2, 10 of those originally in class 1 have to go class 3 and all 20 subjects in class 2 have to go to class 3. So the non zero transitions are : 1 $\rightarrow$ 2, 1$\rightarrow$ 3 and 2 $\rightarrow$ 3. In that case the simulation population exists of:

- 50 individuals originally in class 1, with risk factor changed to 2
- the same 50 individuals originally in class 1, with risk factor changed to 3
- 40 individuals originally in class 2, with risk factor changed to 3.

During post-processing these individuals receive weights of 20/50, 10/50 and 40/40 respectively, also adding unchanged individuals (from the reference population) from classes 1, 2 and 3 with weights 20/50, 0/50 and 10/10.

This procedure makes that for these scenario's stable results can be obtained with a relatively low number of simulated persons, especially in the case that only a moderately number of persons changes risk class, which is often the case in Health Impact Assessments.

In the third case, individuals from the reference population can not be reused in this way as part of the scenario population, as also transition rates differ between reference and alternative scenario. In this case, for a <u>categorical or compound</u> risk factor, the

transitions from old to new prevalence are used to randomly draw a new value for every individual in the reference population. This procedure means that the new prevalence is not fixed, but affected by a random component. In these scenario's one needs to make Nsim sufficiently large.

For continuous risk factors, a new value is assigned in the same way as before, retaining the ranking of subjects in the population.

In all cases, all scenario-individuals are directly matched to a particular individual in the reference population. Matched individuals will share the same random number during simulation.

The initial disease state of the individual does not change by changing the exposure, so that the scenario-individual at the start of simulation has the same disease state as the matched reference individual. The only exception are the newborns (age 0 at the start of simulation), that get immediately the prevalence belonging to the new risk factor state. In practise, disease prevalence is low at birth, so this choice will hardly influence the outcomes.

rivm

6   **Description of simulation module**

During the simulation, there will be simulated persons with the following characteristics

A   age

G   Gender

C   value $\{1,2,\ldots,n_r\}$ of classified risk factor: present if risk factor type = classified or compound

R   value of continuous risk factor or the length of stay in the duration class for compound risk factors

DM_1 to DM_{nc} Disease probability tables for each cluster of diseases (1 to n_c). For non-cancers each table has size $2^{\wedge}(NinCluster(c))-1$. Each value will be indicated by P(s), giving the probability of the disease combination conditional on being alive.

            For diseases with a cured fraction the table has size 2. For independent diseases (not related to other diseases) the table has size 1.

S   Probability of survival

Update rules are used to update these values in steps of one years. Section 6.1.3 describes these update rules.

In the DYNAMO model, simulation starts with a population aged 0-95 (unless the user restricts the age to a smaller range of simulation). All simulated persons in the starting population are simulated until at the age of 105. This makes it possible to calculate cohort life-expectancies for this cohort. (see chapter 7). Simulated persons born during simulation are only simulated for the number of time steps needed to reach the last simulation year that was given by the user.

6.1.1   Synchronising of scenarios

It is essential for a an efficiently working program that all scenarios that need to be compared are run with the same random numbers for the same draws ("parallel universe approach")[9]. Therefore random draws need to be synchronized between scenarios. This is done by matching each individual in the simulation of an alternative scenario to an individual of the reference scenario, and then starting these matched individuals with the same random seed in all scenarios. We referred to this as the "parallel universe approach, as we simulate what would happen to an individual in a parallel universe, where all random happenings are identical to the real universe, but another policy has been being implemented. In making the initial populations for the scenario's, individuals are copied from the reference population to a scenario population, including their random seed.

In the implementation the use of the random-generator is not conditional on characteristics of an individual, so that random number generation can not get out of sync between matched reference and scenario individuals.

6.1.2   Adding newborns to the population

Generating newborns is handled by including individuals with negative age in the initial population.

With the exception of age itself, which is increased by one year during each time step, states of the newborns are kept constant until they reach age 0. From that moment on they are considered as born, and risk factor, disease and survival state are updated using the update rules described below.

6.1.3   Update rules for characteristics of the simulated persons.

6.1.3.1 Update rule for A

Input:

Parameters	Characteristics
none	A

Rule:

A updated = A + 1

6.1.3.2 Update rule for G

Input:

Parameters	Characteristics
	G

Rule:

G updated = G

6.1.3.3 Update rule for C

Input:

Parameters	Characteristics
Transition matrix for age A, gender G	C (A en G needed to select parameter but not in rule itself)

Rule:

Draw a random C from the vector of probability Q. Q is found by taking the row of the transition matrix(A,G) belonging to current risk factor value C.

When zero transition are specified, C stays constant.

6.1.3.4 Update rule for R　　　　　　(if type=continuous risk factor)

<u>Input</u>:

Parameters	Characteristics
meandrift (A,G) , stddrift(ratio)(A,G), offsetdrift(A,G) , type (normal or lognormal), offset(A,G)	R (A en G needed to select parameter but not in rule itself)

<u>Rule</u>:

If R has a normal distribution
R updated=R +drift(A,G) +Normal*stddrift(A,G)
Where Normal is a draw from the standardized normal distribution

If R has a log-normal distribution:
R updated=offset+offsetdrift+exp(log(R-offset) +drift(A,G) + Normal*stddrift(A,G))

NB: for the lognormally distributed R, meandrift(A,G) is not the drift of the mean as entered into the program, but the drift of the parameter of the lognormal distribution.

Update rule for R　　　　　　(if type = compound risk factor)

<u>Input</u>:

Parameters	Characteristics
Number of class of which this is the duration	R ,C

<u>Rule</u>:

If C= duration class then R updated = R+1

Else R=0

Update rule for DM(d)   (for independent disease)

This rule needs to be repeated for each disease d

<u>Input:</u>

Parameters	Characteristics
Baseline incidence: i_0 (d,A,G), Attributable mortality AM(d,A,G), Relative risks for risk factor on the disease (RR),	DM(d), R, C (A en G needed to select parameter but not in rule itself)

<u>Rule:</u>

Calculate RR as:

$RR=RR(per unit)^{\wedge(R(i)\text{-reference value})}$ (in case of a continuous riskfactor)

$RR=RR(c)$ (in case of a categorical risk factor, or the non-duration classes of a compound risk factor)

$RR= RR_{end}+(RR_{begin}-RR_{end})exp(- beta *R(i))$ (in case of the duration category of a compound risk factor)

Calculate the incidence Inc as

$Inc=i_0(d,A,G)*RR$

Solving the exponential of the transition rate matrix

$\quad\quad\quad$ - Inc $\quad\quad\quad$ 0

+ Inc   - AM(d,A,G)

yields the following formulae for DM(d)

DM(d) updated =

$$\frac{(DM(d)AM(d,A,G)-Inc)e^{(Inc-AM(d,A,G))}+Inc(1-DM(d))}{(DM(d)AM(d,A,G)-Inc)e^{(Inc-AM(d,A,G))}+AM(d,A,G)(1-DM(d))}$$

In case AM(d,A,G)=I(d,A,G) the denominator becomes zero. In this case we solve the system of differential equations for the transition matrix

          - Inc   0

          + Inc   - Inc

which yields:

$$1-\frac{1-(DM(d))}{1+Inc(1-DM(d))}$$

Update rule for DM(d)   (for diseases with cured fraction)

Input:

Parameters	Characteristics
Baseline incidence: i_0 (d,A,G), Attributable mortality AM(d,A,G), Relative risks for risk factor on the disease (RR), percentage cured (CuF(d,A,G)	DM(d), R, C (A en G needed to select parameter but not in rule itself)

DM has two values:

DM(c): the probability of having cured disease

DM(nc): the probability of having a lethal form of the disease.

In terms of a transition matrix, there are 3 states: without the cancer, with cured cancer, with non-cured cancer. The vector D_{old} of initial probabities (before the update) of these 3 states are:

1-DM(c)-DM(nc); DM(c),DM(nc).

The transition matrix T is (first row gives the rate of change of the first state as a function of the states) :

$$
\begin{array}{ccc}
- \text{Inc}_{tot} & 0 & 0 \\
+ \text{Inc}_{tot}f_{cured} & 0 & 0 \\
- \text{Inc}_{tot}(1-f_{cured}) & 0 & -am
\end{array}
$$

Here inc_{tot} is the incidence of the disease, calculated in the same way as in the update rule for a single independent disease. The matrix exponential of this matrix also has an analytical solution, which is implement in the software.

The general analytic solution for DM(c) (cured prevalence) after a timestep t is:

$$
\left\{ \left(e^{at}(a - itot)\left(fcured(-1 + ptot) + e^{itot\,t}(fcured + pcured - fcured\,ptot)\right)\right) \right/
$$
$$
\left(e^{itot\,t} itot \right.
$$
$$
\left(-1 + fcured + pcured - fcured\,ptot - e^{at}(fcured + pcured - fcured\,ptot)\right) +
$$
$$
a\left(e^{at}(-1 + fcured)(-1 + ptot) + e^{itot\,t}(-pcured + ptot) + \right.
$$
$$
\left.\left. e^{(a+itot)t}(fcured + pcured - fcured\,ptot)\right)\right)\right\}
$$

and for DM(nc) (non cured prevalence):

$$
\left\{ \left(e^{at} \, (-1 + \text{fcured}) \, \text{itot} \, (-1 + \text{ptot}) + \right. \right.
$$
$$
e^{\text{itot}\,t} \, (a \, (-\text{pcured} + \text{ptot}) + \text{itot} \, (-1 + \text{fcured} + \text{pcured} - \text{fcured}\,\text{ptot})) \Big) \Big/
$$
$$
\left(e^{\text{itot}\,t} \, \text{itot} \, \left(-1 + \text{fcured} + \text{pcured} - \text{fcured}\,\text{ptot} - \right. \right.
$$
$$
e^{at} \, (\text{fcured} + \text{pcured} - \text{fcured}\,\text{ptot}) \Big) + a \, \Big(e^{at} \, (-1 + \text{fcured}) \, (-1 + \text{ptot}) +
$$
$$
\left. \left. e^{\text{itot}\,t} \, (-\text{pcured} + \text{ptot}) + e^{(a + \text{itot})\,t} \, (\text{fcured} + \text{pcured} - \text{fcured}\,\text{ptot}) \Big) \Big) \right\}
$$

where pcured=DM(c) at the start of the time step;

ptot = the sum of DM(c) and DM(nc) at the start of the time step and

a= attributable mortality. In the DYNAMO model with implemented this solution for t=1.

The incidence itot is calculated in the same way as for independent diseases.

Update rule for DM(s)   (for clusters of mutually dependent diseases)

Input:

Parameters	Characteristics
Baseline incidence: i_0 (d,A,G), Attributable mortality AM(d,A,G), Relative risks for risk factor on the disease (RR),Relative risks for diseases on diseases (RRdis)	DM(s), R, C (A en G needed to select parameter but not in rule itself)

DM(s) gives the disease probabilities (conditional on being alive) for each possible combination of the diseases (state s) in the cluster. We use s as argument here in stead of d to indicate that this characteristic indicates a state, and not a single disease.

The state: "none of the diseases" is not stored in the simulation as it can be calculated as 1-sum(other entries). Thus for n diseases in the cluster there are 2^n -1 values of DM(s), where s

is a number, having values 1 to 2^n, which written in binary form indicates the presence of each disease. For instance, s= 100100^1 indicates that the first and fourth disease are present, but diseases 2,3,5 and 6 are not present.

For this rule there are the following steps:

1. add state s=0 and calculate DM(0) as 1 minus the sum of all DM(s).
2. calculate the **rate** of **going** from state s1 to each other state s2 T(s1→s2). (see later for how to calculate these rates), and construct the matrix **T** that contains these values.
3. calculate the matrix of **probabilities** of **going** from state s1 to each other state s2 **P** (containing all probabilities P(s1→s2)) from
   $$\mathbf{P} = \exp(\mathbf{T}.t)$$
   This is carried out by applying the algorithm described by Gallivan et al.[4].
4. calculate the probability of being in state s2 after update, conditional on being alive at the **beginning** of the interval:
   $$DMU(s2) = \sum_s DM(s)P(s \to s2)$$
5. calculate the probability of being in state d2 after update, conditional on being alive at the **end** of the interval :
   $$DM(s2)updated = DMU(s2) / \sum_s DMU(s)$$

Calculation of T(s1→s2)

To calculate the matrix with transition rates, T(s1→s2) we first determine which bits (diseases) in s2 differ from those in s1, as those are the diseases that need to change during the transition. For this we apply the following rules:

---

[1] In the software implementation the ordering is the other way round: the last digit represents disease number 0, the one before that

1. The model does not include remission, so in case there is a disease present in s1 that is not present in s2, T(s1→s2)=0

2 For the other transition rate between two different states s1 and s2: Let O-ac be the set of diseases which are 0 in s1, but 1 in s2 (the set of diseases that is to be acquired) and O-1 the set of diseases which are 1 in s1 (irrespective of their value in s2). As the rate gives the change in an infinitely small period of time, there can only be one change in disease between state. So if O-ac has more than one member, T(s1→s2)=0.

If O-ac has one member, d, then let $inc_0(d)$ be the incidence of disease d, calculated in de same way as of the update of a single independent disease (including the dependence on risk factor values, but not yet the dependence on the presence on other diseases). The rate of acquiring disease d during the update is given by:

$$T(s1 \rightarrow s2) = inc_0(d) \prod_{e \in O-1} RR_{dis}(e,d)$$

3 Mortality rates (attributable mortality and case fatality) are entered as (negative) terms to the diagonal of the matrix. In case a disease in the combination is zero, also the incidence to this disease is a negative contribution to the diagonal term. So the transition rate T(s1→s1) for a diagonal term is:

$$T(s1 \rightarrow s1) = -\sum_{s} T(s1 \rightarrow s) - \sum_{d} I(d=1)am_d - \sum_{d} fatal_d(s1)$$

where am_d is the attributable mortality for disease d (added for those disease that are 1 in s1) and $fatal_d(s1)$ the fatal incidence due to disease d in those with state s1 (added for all states). This fatal incidence is calculated in the same way as given for the incidence above.

Update rule for S

disease 1 etc.

Input:

Parameters	Characteristics
Baseline incidence: i_0 (d,A,G), Attributable mortality AM(d,A,G), Relative risks for risk factor on the disease (RR),Relative risks for diseases on diseases (RRdis) Relative risks for risk factor on other cause mortality(RRom) Baseline other cause mortality (baselineOM)	R, C, D, all DM (A en G needed to select parameter but not in rule itself)

Rule:

Total survival is given by the multiplication of survival fractions due to the different clusters of diseases, and due to other causes.

For independent diseases the survival fraction due to the disease (that is the fraction with which the disease decreases survival in a period t) is given by:

$$S_d = \frac{am(1 - DM(d))\ e^{-inc} + (am.DM(d) - inc)e^{-am}}{am - inc}$$

With:
Am: the attributable mortality for the disease
Inc: the incidence of the disease, calculated as given in the update rule for a single independent disease
DM(d)): the prevalence of the disease (=DM(d))

The survival fraction for other cause mortality is given by:

$$S_{OM} = e^{-om}$$

where om is other cause mortality. This is calculated as follows:

First calculate RR_{om} as :

$RR_{om} = RR_{om}$ (per unit)$^{\wedge (R(i)\text{-reference value})}$ (in case of a continuous riskfactor)

$RR_{om} = RR_{om}$ (c) (in case of a categorical risk factor, or the non-duration classes of a compound risk factor)

$RR_{om} = RR_{om\text{-}end} + (RR_{om\text{-}begin} - RR_{om\text{-}end}) \exp(- beta *R(i))$ (in case of the duration category of a compound risk factor)

Then om= RR_{om}. Baseline-OM.

For diseases with a cured fraction the survival fraction due to these diseases is calculated from the values DM(cured) and DM(not cured) at the start of the time step as:

$$S_d = DM(cured) + e^{-am.} DM(notcured) +$$
$$(1 - DM(cured) - DM(notcured)).$$
$$\{e^{-inc\_cured} + (1 - e^{-inc\_cured}) \frac{(inc\_cured - inc\_noncured)}{inc\_cured} + \frac{(e^{-inc\_cured} - e^{-am})inc\_noncured}{am - inc\_cured}\}$$

For clusters of diseases, the survival is calculated as:

$$S_c = \sum_s DMU(s)$$

where DMU is calculated as given in the update rule for clusters of diseases.

Finally S is calculated by multiplying all the survival fractions as calculated above.

rivm

7 Post-processing module and output generated

The DYNAMO simulation module delivers a population of simulated persons. As output in this form can be very large, it is not stored or made available to the user. This output contains the actually simulated persons, and is not always representative of the population that is simulated. In the post-processing stage the simulated persons are upweighted to the real population numbers as given by the user in the input module. After this weighting, detailed group specific data can be outputted in the form of excel readable XML files. Also DYNAMO-HIA can be run in batch-mode, in which case as csv file is generated, containing similar information, but in one flat file rather than in a workbook with multiple worksheets. In batch mode also output objects are created and stored, which can be viewed later using the graphical user interface. In the batch mode, only the data generated over the user specified running time are included, despite the fact that the model simulates all subjects alive at the start of simulation until age 105.

7.1 Weighting procedure

Weighting is applied to upweight the simulated population to numbers in the real population. First, for categorical risk factors, weights are applied to make the simulated population representative in terms of risk factor classes and changes between scenarios in risk factor classes (see section 5.3). Due to this weighting process, after weighting the risk factor prevalence of categorical risk factors at the start of simulation will be exactly equal to the prevalence entered into the model, with the exception of scenarios populations where both the initial prevalence of the risk factor and the risk factor transitions are changed. The latter will contain a random component.

Second, the number per age and gender are up-weighted to become representative of the population, using the demographic data as supplied by the user.

This procedure delivers population data for scenario populations where the intervention is 100% successful and is applied to every age and gender group. From these data, effects of interventions with lower success rates or effects of an intervention reaching only particular age or sex groups are calculated by combining "successful" scenario individuals and "unsuccessful" reference individuals in the correct proportion.

7.2   Content of DYNAMO-output files

excel readable XML files can be read by using "open with" and then choose "excel".

The output in these files are in the form of numbers of persons in the total population.

The excel readable output exists in two forms:

Yearly data:
For each simulated year (as given by the user), the numbers in the simulated population (males, females or both combined) by:
- Age
- Actual risk class in that year (for continuous variables: categorized in groups)

For each age /risk factor class combination, the files contain:

1. For continuous risk factor: the mean value
2. For compound risk factor: the mean value of duration
3. The total number of years lived in that class
4. Either numbers in the class with the different disease states (= combinations of diseases within each cluster), or the numbers with each disease (to choose by the user).
5. The number of DALY-weighted years in that class

6. The number of persons with one or more diseases in that class

Each year is on its own worksheet of the excel workbook.

Cohort data:

For each simulated cohort (group with same age at start of simulation), the numbers by

Year of simulation (always up to the age of 105)

Risk class at the start of follow-up (for continuous variables: categorized in groups)

For each age/gender/risk factor class combination, the files contain:

1. For continuous risk factor: mean value

2. For compound risk factor: mean duration

3. Total number of persons alive

4. Disease state or disease (to choose by the user).

5. The number of DALY-weighted years in that class

6. The number of persons with one or more diseases in that class

Each cohort is on its own worksheet of the excel workbook. The user can produce separate workbooks for men and women, or a single workbook in which numbers from men and women are combined.

Batch output:

For each simulated year (as given by the user), the numbers in the simulated population (males and females separately) by:

- Age

- Actual risk class in that year (for continuous variables: categorized in groups)

For each age /risk factor class combination, the files contain:

1. The year

2. The scenario

3. The gender

4. The risk factor class

5. For continuous risk factor: the mean value

6. For compound risk factor: the mean value of duration

7. The total number of years lived in that class

8. Either numbers in the class with the different disease states (= combinations of diseases within each cluster), or the numbers with each disease (to choose by the user).

9. The number of DALY-weighted years in that class

10. The number of persons with one or more diseases in that class

Correspondence with the use of data in the graphical user interface

Yearly data are in the DYNAMO-HIA graphical user interface - output module used to calculate:

- Prevalence of disease by year, or by age (within each year)

- Risk factors distribution by year, or by age (within each year)

- Mortality numbers (those dying between the year and the next year) by year, or by age (within each year)

- Period life expectancy and Sullivan Health expectancy

Cohort data are are in the DYNAMO-HIA graphical user interface - output module used to calculate:

- Cohort Life expectancy (by age)

- Disease free life expectancy and life expectancy with disease (cohort)

- Survival

- Disability adjusted life expectancy (cohort)

rivm

7.3   Calculation of integral measures of health.

The program calculates the following health expectancy values:
- total life expectancy
- life expectancy free of a particular disease
- life expectancy free of all diseases in the model
- disability free life expectancy or DALE. Which of these is calculated, depends on the data given by the user. If DALY-weights were given, DALEs are calculed. If percentages of disability have been given, disability free life expectancy is calculated.

For life expectancy free of all modelled diseases, the prevalence of having no disease is calculated for each simulated individual. As the clusters of diseases that are calculated in DYNAMO are mutually independent given the risk factor history, this can be calculated by multiplying the probabilities of being disease-free from each cluster of diseases.

Similarly, within each cluster, the DALY-weights can be applied to the different disease-states in this cluster (including the healthy state, that is the state without any of the disease in the model). If a person has two diseases, the DALY weight given to the person is calculated as:

$$1- W_{daly.total}=(1-W_{daly-disease1}).(1-W_{daly-disease2})(1-W_{daly-healthy})$$

For ease of explanation, we will use the term "ability" for 1-DALY-weight, defined as either 1 – the fraction disabled or as a health valuation where 1 is perfect health, and 0 health as bad as being dead.
Then in general the ability due to different diseases is calculate as:

$$ability= baselineability \prod_{d} ability_d$$

where the baseline ability is ability in a person without any of the diseases that are included in the model, and $ability_d =1$ for a person without disease d. The abilities of the different

health states in the cluster can be averaged (weighted for the probability of the health state) to the average "ability" in the cluster. As disease clusters are mutually independent given risk factor history, for a simulated individual the overall ability can then be calculated by multiplying the abilities due to different disease clusters.

Health expectancies are calculated both as cohort health expectancies and cross-sectional ("Sullivan") health expectancies.

The cohort health expectancies are calculated directly from the simulating the cohort (always simulated until age 105), and the calculation consists of simply counting the years lived in the particular health state during this follow-up period by all simulated individuals, and dividing this by the number of individuals at the start of simulation. Implicitly this means that the life expectancy in the age category 105+ is taken as 0.5. However, as individuals of age 105 and older are extremely rare, this choice does not influence the results.

Cross-sectional health expectancy is a synthetic measure, that is calculated for a particular calendar year. Thereto first the mortality in that year is calculated by taking the difference between the survival in the next year and the current year. This mortality is combined with the prevalence of disease prevalence or disability in the particular year in a period life table constructed based on the data from the one calendar year, and a health expectancy calculated from the life table in the usual way. For the cross-sectional health expectancy, only ages upto 95 are included. The life expectancy in the age category 95+ is taken as: $-1/\ln(1-\text{mortality rate (95)})$.

rivm

Reference List

1. Lhachimi, S.K., et al., *Standard tool for quantification in health impact assessment a review.* Am J Prev Med, 2010. **38**(1): p. 78-84.
2. Boshuizen, H.C., et al., *The DYNAMO-HIA model: An efficient implementation of a risk factor / chronic disease Markov model for use in health impact assessment.*
3. Moler, C. and C. Van Loan, *Nineteen Dubious Ways to Compute the Expotential of a Matrix.* SIAM Review, 1978. **20**(4): p. 801-836.
4. Gallivan, S., et al., *A computational algorithm associated with patient progress modelling.* Computational Management Science, 2007. **4**(3): p. 283-299.
5. Neogi, T. and Y. Zhang, *Re: "Easy SAS calculations for risk or prevalence ratios and differences".* Am J Epidemiol, 2006. **163**(12): p. 1157; author reply 1159-61.
6. Alho, J.M., *On prevalence, incidence, and duration in general stable populations.* Biometrics, 1992. **48**(2): p. 587-92.
7. Hoogenveen, R.T., P.H. van Baal, and H.C. Boshuizen, *Chronic disease projections in heterogeneous ageing populations: approximating multi-state models of joint distributions by modelling marginal distributions.* Math Med Biol, 2009.
8. Kassteele J, et al., *Estimating net transition probabilities from cross-sectional data with application to risk factors in chronic disease modelling.*
9. Shechter, S.M., et al., *Increasing the efficiency of Monte Carlo cohort simulations with variance reduction techniques.* Med Decis Making, 2006. **26**(5): p. 550-3.