Prevent User Manual

School of Population Health
University of Queensland
Department of Public Health
Erasmus MC

Production and layout:
Maatoplossing
Meuleberg 47
5954 AS Beesel
the Netherlands
www.Maatoplossing.nl

Revision History

Version	Date	Author	Description
0.1	06-04-07	Bert Janssen	Beta release used for review. Portions are based on 'Prevent Documentation V3' written by Jan Barendregt.
0.2	16-04-07	Bert Janssen	Added all conclusive comments from Esther de Vries-Baquero & Isabelle Soerjomataram
0.3	24-04-07	Bert Janssen	Added output description chapters
0.4	04-05-07	Bert Janssen	Added all conclusive comments from Esther de Vries-Baquero & Isabelle Soerjomataram
0.5	09-05-07	Bert Janssen	Added all conclusive comments from Esther de Vries-Baquero & Isabelle Soerjomataram
0.6	13-05-07	Bert Janssen	Added comments from Jan Barendregt, Esther de Vries-Baquero, Isabelle Soerjomataram, Maryska Janssen-Heijnen & Bert Janssen
0.7	25-10-07	Bert Janssen	Added description of all other tables for the DYNAMO-HIA project
0.8	04-12-07	Bert Janssen	Added comments from Esther de Vries
0.9	03-03-08	Bert Janssen	Added comments from Wilma Nusselder
1.0	13-05-08	Bert Janssen	Final Version. Later versions will be updated by ErasmusMC

Preface

This manual describes version 3.0 beta of the **Prevent** application, written by Jan J. Barendregt. Since the application is still in beta phase, some functionality may not be complete, and some bugs may still be present. Also there is some inconsistency in the naming conventions. Where possible this is indicated.

The manual consist of two parts:

- Description of the definition and data files
- Using the Prevent application

The description of the data files emphasizes on the data files structure and its contents. There are notes wherever the data file content has restrictions. It also explains where the items of the datafile are used in the Prevent application.

Using the Prevent application emphasizes on the use of the program. It also explains where the items of the datafile are used in the Prevent application.

This manual does not describe the logic behind the calculations. Please consult the literature list for more information.

Table of Contents

General information

Data types

Throughout this manual different types of data types are used. The following table gives an explanation about what is meant by each data type:

Data type	Description	Possible values
boolean	Toggle between TRUE of FALSE	TRUE, FALSE
integer	Any whole number either positive or negative and zero	...,-3,-2,-1,0,1,2,3...
text	All letters, numbers and punctuation marks	
binary	All characters are allowed	

If any data type is limited in length, the maximum length is indicated.

Data files location

All data files are stored in **Program Files\Prevent\data** by default. This can be changed by editing the prevent.ini file. See the chapter about the ini file for more information.

Data file description

The next chapters describe the data files. This section is divided into two parts:

- Definition tables
  - Describing tables which define Prevent settings and default values
- Data tables
  - Describing tables which hold the actual data

Definition tables

GeneralTab

DatasetName	BaseYear	HiAge	PopulationProjection	ProjectionMax	IncidenceOnly	Disability	Costs	Costunit	Description	Model
Test	2000	95	TRUE	75	TRUE	FALSE	FALSE	Euro		

The first table to be read by the programme containing basic definitions.
Please note: only the first record is being used, other records are ignored.

Fields:

Field name	Data type	Limitation	Description
DatasetName	text	max 50 chars	Free field for benefit of the user only.
BaseYear	integer		The base year is the year that will be the base of the projection. Base year is always correlated to time=0 The first year of projection: time=1.
HiAge	integer		The highest open-ended age group in the dataset. Allowed values are 0-110 in 5 year increments.
PopulationProjection	boolean		If TRUE, Population projection is provided by user. If FALSE, Population projection is not provided by user, thus needs to be calculated by Prevent.
ProjectionMax	integer		when PopulationProjection is TRUE, this field gives the number of years of available projection input. Prevent will not allow runs longer than this maximum.
IncidenceOnly	boolean		If TRUE, Prevent will calculate the effect of risk factor changes on disease incidence, but not calculate disease prevalence and mortality.
Disability	boolean		If TRUE, the model expects the TotalDisability table to contain data.
Costs	boolean		If TRUE, the model expects the TotalCost table to contain data.
Costunit	text	max 50 chars	The currency of the cost input.
Description	text	no limit	Free field, for notes for the benefit of the user.
Model	binary		Allows one to store the Prevent executable in the database. This way the user is able to create a package containing both the data and the used version of the Prevent program.

DiseasesAndRiskfactors

Name	Disease	PrevalenceOnly	Continuous	Cohortprev	MortalityOnly	Morbidity	Remission	Disability	Costs	RiskfactorToo	Lookback
Colorectal cancer	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0
Lung cancer	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0
Obesity	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	0
Smoking	FALSE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	20

This table lists the diseases and risk factors, and specifies a number of properties.

Fields:

Field name	Data type	Limitation	Description
Name	text	max 50 chars	Unique name, not case sensitive, spaces allowed.
Disease	boolean		If TRUE, the name corresponds to a disease. If FALSE,the name corresponds to a risk factor.
PrevalenceOnly	boolean		Currently diseases cannot be described by (age-specific) prevalence, only risk factor has prevalence.
Continuous	boolean		Only applicable to risk factors. If TRUE the risk factor is described by a continuous distribution, otherwise it is categorical.
Cohortprev	boolean		When false the default risk factor prevalence mode is used, when true Prevent expects additional table to be present with the information needed for the Mendez model.
MortalityOnly	boolean		If TRUE, a disease will be defined as a cause of death only (hence will have no morbidity). Not applicable for risk factors.
Morbidity	boolean		This field is curently not used.
Remission	boolean		Can be TRUE only when MortalityOnly is FALSE. If TRUE, the disease model used includes remission, and input must consist of incidence, case fatality, and remission (see ref 15 for the disease model used). If FALSE, the disease model excludes remission, and input must consist of incidence, disease prevalence and mortality (see ref 10 for the disease model used).
Disability	boolean		Can be TRUE only when MortalityOnly is FALSE and Disability in GeneralTab is TRUE. If TRUE, then disease-specific disability weight input is expected.
Costs	boolean		Can be TRUE only when MortalityOnly is FALSE and Costs in GeneralTab is TRUE. If TRUE, then disease-specific cost input is expected.
RiskfactorToo	boolean		If TRUE, the disease is also a riskfactor. Not applicable for risk factors.
Lookback	integer		The number of years into the past that prevalence data is provided for risk factors. Presently the status of this parameter for diseases and risk factors that have a risk factor themselves is that it should be put to 0 (See Known issues).

Data from this table is visible in:

- The disease selection boxes when the Disease field is TRUE.
- The risk factor selection boxes when the Disease field is FALSE

DiseaseRiskfactorRelation

Diseases	Risk factors	CUM	LAT	LAG
Colorectal cancer	Obesity	0	0	10
Lung cancer	Smoking	0	3	17
Obesity	Smoking	0	1	9

This table links risk factors and diseases, and specifies the lag times.

Fields:

Field name	Datatype	Limitation	Description
Diseases	text	max 50 chars	The name of the disease (or risk factor) influenced. Make sure the spelling is exactly like Name in DiseasesAndRiskfactors, otherwise it will not be recognized.
Risk factors	text	max 50 chars	The name of the risk factor (or disease) influencing the disease. Make sure the spelling is exactly like Name in DiseasesAndRiskfactors, otherwise it will not be recognized.
CUM	integer		Currently not used.
LAT	integer		LAT is the time before ANY change in risk occurs (e.g. 3 years for lung cancer) LAG starts to have an effect after the LAT years (so it starts 3 years after a change in exposure). LAG then gives the number of years (17 years for lung cancer) it takes for the RR to fully reach the RR of the new exposure group (e.g. RR=1 if exposed people become unexposed). Several functions can be modelled in PREVENT for LAG (linear, exponential and logarithmic).
LAG	integer		See above for an explanation of LAG and its relation to LAT
LAGFunction	text	max. 50 chars	Linear, logistic,

Data tables

The next section describes the tables which contain the actual input data.

A note on age categories

These tables have in common that the data is age-specific, and they all use the same flexible system of categorization of ages. An example is given in the table below.

Lage	Hage	Males	Females
0	1	0.00552	0.00506
1	2	0.00049	0.00048
2	49	0.00036	0.00027
49	76	0.00020	0.00016
76	77	0.00018	0.00019
77	95	0.00019	0.00008
95	96	0.00020	0.00008

An age category is defined by two numbers: Lage gives the starting age, and the category then runs up to but excluding the age of Hage. So in the example we have input for age 0 (0-12 months), age 1 (12-24 months etc.), ages 2-48, 49-75, 76, 77-94, and 95 and over (assuming that the highest age as defined in GeneralTab is 95). This system allows any age categorization, as long as the following rules are obeyed:

- The lowest Lage must always start at 0.
- The highest Hage must be the highest age as defined in GeneralTab plus 1.
- No gaps are allowed, thus the Lage of one age category must be equal to the Hage of the previous one.

Together these rules in fact mean that the model will not interpolate or extrapolate missing data. The handling of wide (=wider than 1) age categories depends on the kind of input data. When the input is the total number in that age category (for example population numbers) each single year of age is assigned a value equal to the input number divided by the width of the age category. When it is a rate, probability, weight, or relative risk each single year of age simply gets assigned the input number. No smoothing is performed: when the user thinks the resulting input categories are to coarse, s/he should first do some smoothing, and use the smoothed data as input instead. Of course it is always best to use 1-year age group data whenever available.

Please note: data in the Males and Females fields have different meanings depending on the tables where they are used:
In CatrfExposure: it is the prevalence of certain risk factor for males and for females.
In DiseaseInputVars: it is the rate of certain disease for males and for females.

Scenarios, trends, and interventions

Scenarios, interventions, and trends are implemented in various ways, but there is also a common denominator. The first is that they all have a name, the second that they have a time field. An example is the DiseaseInterventions table below. Other examples are table: CatrfInterventions and ConfrIntervenions.

InterventionName	DiseaseName	Param	Time	Lage	Hage	Males	Females
Lung Cancer	IHD	case fatality	1	0	96	-0.03	-0.02
Lung Cancer	IHD	case fatality	15	0	96	0	0

The Name field (here called 'InterventionName') is a unique name to identify the scenario (or trend, or intervention). As the model steps through time (internally the base year equals 0) from 0,1,2…, it checks whether the time field of the scenario corresponds with the simulation time. If so, the values in the Males and Females fields are applied, and remain in force until the time of the next entry of the scenario.
For example, in the case of the "Lung Cancer" above the case fatality annually declines with 3% (males) and 2% (females) for years 1-14, and becomes constant in year 15. Without this second entry, the trend would be applied until the end of the simulation.
In the case of the DiseaseInterventions table the values are interpreted as annual proportional change, but this may be different in other tables. In the CatrfExposure table, for example, the values are interpreted as the proportion exposed to the risk factor in that particular year. Apart from the interpretation, however, the system works the same.
In several tables fields 'ScenarioName' and 'Time' are present, but currently the model does not offer the possibility to choose anything other than the 'Standard' and '0' values these fields contain. This, however, may change in the future.

TotalMortality

ScenarioName	Time	Lage	Hage	Males	Females
Standard	0	0	1	0.00552	0.00506
Standard	0	1	2	0.00049	0.00048
Standard	0	2	3	0.00036	0.00027
Standard	0	3	4	0.0002	0.00016
Standard	0	4	5	0.00018	0.00019
Standard	0	5	6	0.00019	0.00008

This table contains total mortality probabilities of general population of, the base year of the data set (Time=0). The field 'ScenarioName' is not used currently.

Please note: This table is only used when Population Projection is not provided by user. However Prevent expects this table to be filled with data, otherwise the program will not run. Dummy data for Males and Females must be > 0 and < 1.

Population

ScenarioName	Time	Lage	Hage	Males	Females
BaseYear	0	0	1	103549	98199
BaseYear	0	1	2	102854	98163
BaseYear	0	2	3	99191	95356
BaseYear	0	3	4	99112	93733
BaseYear	0	4	5	98864	94470
BaseYear	0	5	6	101584	96998
.	.	.	.	.	.
Projection	1	0	5	503175	479448
Projection	1	5	10	509812	486971
Projection	1	10	15	494619	473619
Projection	1	15	20	473221	450989
Projection	1	20	25	482533	470996

This table contains population numbers of the base year of the data set and optionally projections of future population numbers. The fields 'ScenarioName' and 'Time' are used to distinguish between the base year population and the projection years. The BaseYear and Projection scenario names are mandatory.

Please note: when ScenarioName=BaseYear, the values of Lage and Hage must be provided in 1 year age categories.
When ScenarioName=Projection, the values of Lage and Hage must be increased by any integer.

Numbers in Males and Females represent (observed or projected) numbers of males/females in that particular age-group and year.

CatrfCats

RiskfactorName	Catno	CatName
Smoking	0	Never
Smoking	1	Current

Categorical risk factors can have any number of categories. Categories are defined in this table by assigning a number and a name. Numbers start at 0 with the reference group, which corresponds to Relative Risk=1.

CatrfExposure

RiskfactorName	ScenarioName	Time	Catno	Lage	Hage	Males	Females
Smoking	Standard	-20	0	0	15	1	1
Smoking	Standard	-20	0	15	20	0.5713	0.5509
Smoking	Standard	-20	0	20	25	0.5158	0.5102
Smoking	Standard	-20	0	25	30	0.419	0.4587
Smoking	Standard	-20	0	30	35	0.3472	0.4441
.	.	.	.	.	.	.	.
Smoking	Standard	-20	1	0	15	0	0
Smoking	Standard	-20	1	15	20	0.4287	0.4491
Smoking	Standard	-20	1	20	25	0.4842	0.4898
Smoking	Standard	-20	1	25	30	0.581	0.5413
Smoking	Standard	-20	1	30	35	0.6528	0.5559

In the default mode this table contains exposure by age for each time from –lookback to as far in the future as desired. When Cohortprev is 'true', Prevent expects only the age specific prevalences at time –lookback.
Proportions exposed by age, sex, category number, and time. Prevent expects the proportions to add up to 1. The model also expects the first entry to be at time '-lookback' (see DiseasesAndRiskfactors). The field 'ScenarioName' is currently not used.

Please note: Numbers in Males and Females represent the proportion exposed to the risk factor in the indicated age group and year. Values should be between 0 (nobody exposed) to 1 (100% exposed). The proportions of the different levels of one risk factor within one year and age-group should add up to 1.

CatrfCohortInitial

RiskfactorName	ScenarioName	Time	Catno	Age	Males	Females
Smoking	Standard	-27	0	18	0,5713	0,5509
Smoking	Standard	-27	1	18	0.4287	0.4491
Smoking	Standard	-26	1	18	0.4287	0.4491
Smoking	Standard	-26	0	18	0.5713	0.5509
Smoking	Standard	-25	1	18	0.4195	0.4379

This table contains the prevalence at the intiation age for all times from –lookback to as far into the future as desired.

CatrfCohortChange

RiskfactorName	ScenarioName	Time	Catno	Lage	Hage	Males	Females
Smoking	Standard	-27	1	18	30	0	0
Smoking	Standard	-27	1	30	50	-0,00718	-0,00718
Smoking	Standard	-27	1	50	96	-0,04528	-0,04528
Smoking	Standard	-7	1	18	30	-0,00209	-0,00209
Smoking	Standard	-7	1	30	50	-0,02147	-0,02147
Smoking	Standard	-7	1	50	96	-0,05958	-0,05989

This table contains the change ‘rates’ by age, sex, and time. For declines the ‘rates must be negative, for increases positive.‘rates’ is used because the numbers are used as probabilities, but negative probabilities don’t exist. Note that change ‘rates’ are given for the exposed categories only (so not for catno 0).

CatrfCohortMortRR

RiskfactorName	ScenarioName	Time	Catno	Lage	Hage	Males	Females
Smoking	Standard	0	1	0	96	2	2

The excess mortality of the exposed I've implemented as a relative risk on total mortality. This table contains the relative risks by age, and sex. The Time field should contain 0. Note that the relative risks are given for the exposed categories only (so not for catno 0).

CatrfRelRisk

Riskfactor	Disease	Scenario	Time	Disvar	Catno	Lage	Hage	Males	Females
Smoking	Lung cancer	Standard	0	Incidence	0	0	96	1	1
Smoking	Lung cancer	Standard	0	Incidence	1	0	96	20	20
Smoking	Obesity	Standard	0	Prevalence	0	0	96	1	1
Smoking	Obesity	Standard	0	Prevalence	1	0	96	0.5	0.5

Relative risks for each combination of categorical risk factor and disease. **Scenario** must be Standard, **Time** must be zero and **DisVar** must be either Incidence or Prevalence.

Please note: Numbers in Males and Females is relative risk. Catno 0=reference
***CatNo** corresponds with **Catno** in the other tables*

CatrfInterventions

InterventionName	RiskfactorName	Param	Time	Catno	Lage	Hage	Males	Females
Taxes	Smoking	Prevalence	1	1	0	96	0.07	0.07
Taxes	Smoking	Prevalence	10	1	0	96	0	0

Categorical risk factor interventions are expressed as the proportion of the currently exposed that are no longer exposed (proportion given for 'Males' and 'Females').
(this is inconsistent with the DiseaseInterventions and OtherMortalityTrend tables, where declines are negative. So this might change in the next release). In the table, 'Taxes' decreases smoking by 7% annually in both males and females for all ages in the time period BaseYear+1 until BaseYear+10, after which the proportions of the currently exposed remain stable.

ConrfExposure

ConrfName	ScenarioName	Time	Lage	Hage	Distribution	MPar1	MPar2	MPar3	MParmin 1	MParmin 2	MParmin 3	FPar1	FPar2	FPar3	FParmin 1	FParmin 2	FParmin 3
Obesity	Standard	0	0	65	normal	26.2	5	0	21	1	0	25.3	6	0	21	1	0
Obesity	Standard	0	65	96	normal	27.4	4	0	21	1	0	28.4	4.5	0	21	1	0
Obesity	Standard	1	0	65	normal	26.462	5.05	0	21	1	0	25.553	6.06	0	21	1	0
Obesity	Standard	1	65	96	normal	27.674	4.04	0	21	1	0	28.684	4.545	0	21	1	0
Obesity	Standard	2	0	65	normal	26.72662	5.1005	0	21	1	0	25.80853	6.1206	0	21	1	0
Obesity	Standard	2	65	96	normal	27.95074	4.0804	0	21	1	0	28.97084	4.59045	0	21	1	0
Obesity	Standard	3	0	65	normal	26.99389	5.151505	0	21	1	0	26.06662	6.181806	0	21	1	0
Obesity	Standard	3	65	96	normal	28.23025	4.121204	0	21	1	0	29.26055	4.636355	0	21	1	0
Obesity	Standard	4	0	65	normal	27.26383	5.20302	0	21	1	0	26.32728	6.243624	0	21	1	0
Obesity	Standard	4	65	96	normal	28.51255	4.162416	0	21	1	0	29.55315	4.682718	0	21	1	0
Obesity	Standard	5	0	65	normal	27.53646	5.25505	0	21	1	0	26.59055	6.30606	0	21	1	0
Obesity	Standard	5	65	96	normal	28.79768	4.20404	0	21	1	0	29.84869	4.729545	0	21	1	0
Obesity	Standard	6	0	65	normal	27.81183	5.307601	0	21	1	0	26.85646	6.369121	0	21	1	0
Obesity	Standard	6	65	96	normal	29.08565	4.246081	0	21	1	0	30.14717	4.776841	0	21	1	0
Obesity	Standard	7	0	65	normal	28.08995	5.360677	0	21	1	0	27.12502	6.432812	0	21	1	0
Obesity	Standard	7	65	96	normal	29.37651	4.288541	0	21	1	0	30.44864	4.824609	0	21	1	0
Obesity	Standard	8	0	65	normal	28.37085	5.414284	0	21	1	0	27.39627	6.49714	0	21	1	0
Obesity	Standard	8	65	96	normal	29.67027	4.331427	0	21	1	0	30.75313	4.872855	0	21	1	0
Obesity	Standard	9	0	65	normal	28.65455	5.468426	0	21	1	0	27.67024	6.562112	0	21	1	0
Obesity	Standard	9	65	96	normal	29.96698	4.374741	0	21	1	0	31.06066	4.921584	0	21	1	0
Obesity	Standard	10	0	65	normal	28.9411	5.523111	0	21	1	0	27.94694	6.627733	0	21	1	0
Obesity	Standard	10	65	96	normal	30.26665	4.418489	0	21	1	0	31.37127	4.9708	0	21	1	0

This table describes the exposure of continuous risk factors, by age, kind of distribution (currently choice of Normal, lognormal, lognormal with offset, and Weibull), sex, and time. There are six parameter fields for each sex: three to describe the current parameters (MPar1-3 for males, Fpar1-3 for females), and three to describe the theoretical minimum risk (MParmin1-3 & FParmin1-3. See reference 16 for the concept of theoretical minimum risk). The lognormal with offset uses the 3rd field for its offset parameter, the other distributions use only the first two fields. The ScenarioName field is currently not used.
Theoretical minimum risk of a risk factor can not be zero. The theoretical minimum value that is taken by prevent is that of the base year.
Mpar1-3 and Fpar1-3 are parameters of the distribution and not the same as the parameters on page 19.

Note:
Par1=mean/mu,
Par2=standard deviation/sigma,
Par3=offset for lognormal (mu and sigma)

ConrfRelRisk

Riskfactor	Disease	Scenario	Time	Disvar	Lage	Hage	RiskFunc	Mpar1	Mpar2	Mpar3	Fpar1	Fpar2	Fpar3
Obesity	Colorectal cancer	Standard	0	incidence	0	65	Linear	-0.3	0.06	0	-0.3	0.06	0
Obesity	Colorectal cancer	Standard	0	incidence	65	96	Linear	-0.3	0.06	0	-0.3	0.06	0

This table describes the risk functions that relate risk to each increase in level of exposure from a continuous risk factor.
Currently there is choice of linear, two-piecewise linear, or logit, by age and sex. The two-piecewise linear function uses all three parameter fields (Mpar1-3 & Fpar1-3), the other, only the first two. The Scenario, Time and Disvar fields are currently not used.
Mpar1-3 and Fpar1-3 are parameters of the distribution and not the same as the parameters on page 18.

Note:
Par1=the intercept
Par2=the change per unit change of the risk factor
Par3=the measure unit of the factor

ConrfInterventions

InterventionName	RiskfactorName	ScenarioName	Time	Lage	Hage	MPar1	MPar2	MPar3	FPar1	FPar2	FPar3
Halfway to the theoretical minimum	Obesity	Standard	1	0	65	0.5	0.5	0	0.5	0.5	0
Halfway to the theoretical minimum	Obesity	Standard	1	65	96	0.5	0.5	0	0.5	0.5	0
Halfway to the theoretical minimum	Obesity	Standard	2	0	65	0	0	0	0	0	0
Halfway to the theoretical minimum	Obesity	Standard	2	65	96	0	0	0	0	0	0

Interventions for continuous risk factor variables for each age group can be defined in this table. Lage and Hage and sex is defined by MParx (males) and FParx (females). In the example the obesity distribution in the intervention scenario abruptly changes at time 1 from reference (=1) to a distribution that is half way (Mpar1=0.5, Fpar=0.5) between the current and the theoretical minimum risk distribution (defined in Table ConrfExposure). The two records with the Time field = 2 ensure that after the abrupt change at time 1 no further intervention changes occur (Mpar1=0, Fpar1=0). The ScenarioName field is not used at present.

Note:

If users include exposure data for more than 1 age groups (in ConrfExposure), then in this worksheet type and size of the interventions should be specified for the similar age-group distribution.

DiseaseInputVars

DiseaseName	ScenarioName	Sex	Disvar	Lage	Hage	Value
Colorectal cancer	Standard	Males	Incidence	0	25	0
Colorectal cancer	Standard	Males	Incidence	25	26	6.37E-07
Colorectal cancer	Standard	Males	Incidence	26	27	1.91E-06
Colorectal cancer	Standard	Males	Incidence	27	28	3.82E-06
Colorectal cancer	Standard	Males	Incidence	28	29	6.37E-06
Colorectal cancer	Standard	Males	Incidence	29	30	9.55E-06
Colorectal cancer	Standard	Males	Incidence	30	31	1.34E-05
Colorectal cancer	Standard	Males	Incidence	31	32	1.78E-05
Colorectal cancer	Standard	Males	Incidence	32	33	2.23E-05

The input into this table depends on the settings in DiseasesAndRiskfactors:

- If Disease=TRUE and the rest=FALSE the model expects Disvar – Incidence with corresponding age-sex specific incidence rates.
- If MortalityOnly= TRUE the model expects the Disvar field to read Mortality, with disease specific mortality rates in the Value field at base time.
- If MortalityOnly= FALSE and Remission= TRUE the model expects the Disvar field to have values Incidence, Case fatality, and Remission, with corresponding hazard rates in the Value field at base time.
- If MortalityOnly= FALSE and Remission= FALSE the model expects the Disvar field to have values Incidence, Prevalence, and Mortality, with corresponding hazard rates, proportions and rates in the Value field at base time. These variables have to be internally consistent (see reference 18). For women the Sex field should read ‘Females’. The ScenarioName field is not used at present.

DiseaseInterventions

InterventionName	DiseaseName	Param	Time	Lage	Hage	Males	Females
Lung cancer trend	Lung cancer	incidence	1	0	96	-0.01	0.01
colorectal cancer trend	colorectal cancer	incidence	1	0	96	0.02	0.01
colorectal cancer trend	colorectal cancer	incidence	20	0	96	0	0

It is possible to define trends in disease variables that are not related to changes in known risk factors, or that may be related but are not sufficiently captured by the current cross-sectional with time lags risk factor model. In the example table a 1% annual negative trend for males (Males=-0.01) in lung cancer incidence is defined, together with a 1% annual increase for females (Females=0.01). As defined here, both trends will start in BaseYear + 1, and remain in effect for the duration of the simulation run as defined in GeneralTab-ProjectionMax. Besides the lung cancer trend, in the example, an increasing trend for colorectal cancer has been specified to be 2% for males and 1% for females, starting both in BaseYear +1, and remain in effect until BaseYear +20, after which they will be stable.
See also the section 'A note on scenarios, trends and interventions' above.

Please note: Numbers in Males and Females represent annual proportional change and can be both positive (increasing trends) or negative (decreasing trend)(0 = no change, 1=100% increase per year, -1=-100% increase per year = 100% decrease).

BirthRates

InterventionName	Time	Lage	Hage	Males	Females
Standard	0	0	15	0	0
Standard	0	15	16	0.000227207	0.000216387
Standard	0	16	17	0.000684163	0.000651584
Standard	0	17	18	0.001699006	0.001618101
Standard	0	18	19	0.00343298	0.003269504
Standard	0	19	20	0.006535654	0.006224433

The BirthRates table specifies the number of boys and girls born to a woman alive, by age. If no scenario is specified, the one called 'Standard' is used, but the model allows to choose a different one in tables CatrfRelRisk and ConrfRelRisk.

DiseaseCost

DiseaseName	ScenarioName	Time	Lage	Hage	Males	Females
CVA	Standard	0	0	25	0.00	0.00
CVA	Standard	0	25	30	0.00	55984.83
CVA	Standard	0	30	35	20426.18	5639.93
CVA	Standard	0	35	40	9729.29	2957.31
CVA	Standard	0	40	45	6447.92	2886.12

If the Costs field in DiseasesAndRiskfactors= TRUE the model expects total costs per prevalent case (single currency units) in this table for that disease mentioned in the column DiseaseName. The costs are given for males and females for each age group defined with Lage and Hage. The fields 'ScenarioName' and 'Time' are currently not used.

DiseaseWeight

DiseaseName	ScenarioName	Time	Lage	Hage	Males	Females
COPD	Standard	0	0	96	0.542	0.542
CVA	Standard	0	0	96	0.306	0.306
Dementia	Standard	0	0	96	0.45431	0.45431
IHD	Standard	0	0	96	0.126	0.126
Lungcancer	Standard	0	0	96	0.1	0.1
Other cancer	Standard	0	0	96	0.1	0.1

If the Disability field in DiseasesAndRiskfactors=TRUE the model expects a disability weight entry in this table for that disease (DiseaseName). The fields 'ScenarioName' and 'Time' are currently not used.

Migration

InterventionName	Time	Lage	Hage	Males	Females
Mean 1999-2003	0	0	1	615	564
Mean 1999-2003	0	1	2	450	603
Mean 1999-2003	0	2	3	203	300
Mean 1999-2003	0	3	4	193	215
Mean 1999-2003	0	4	5	217	177
Mean 1999-2003	0	5	6	169	150
Mean 1999-2003	0	6	7	168	163

By default, Prevent assumes zero migration, but it allows to specify migration scenarios. These are expressed as the net migration in numbers by age and sex.

What is 'Mean 1999-2003' in this table?

OtherMortalityTrend

InterventionName	Time	Lage	Hage	Males	Females
CBS trend 2002-2049	0	0	1	-0.00356091	-0.001446353
CBS trend 2002-2049	0	1	2	-0.003702944	-0.001870027
CBS trend 2002-2049	0	2	3	-0.004146881	-0.002299768
CBS trend 2002-2049	0	3	4	-0.004736509	-0.002502882
CBS trend 2002-2049	0	4	5	-0.005059737	-0.002935053
CBS trend 2002-2049	0	5	6	-0.004566404	-0.002320694
CBS trend 2002-2049	0	6	7	-0.004736474	-0.002320694

Prevent calculates an 'all other causes'-mortality from the total mortality and the disease-specific mortalities. By default, this 'all other causes'-mortality is assumed to be constant, but it is possible to define an annual trend. This is expressed as the proportional annual change in the 'all other causes'-mortality rate (how is this calculated?), and the rate (linear I time?)for any given year is calculated cumulatively from the base year rate.

TotalCost

ScenarioName	Time	Lage	Hage	Males	Females
Standard	0	0	5	1409.18	1164.51
Standard	0	5	10	828.29	654.33
Standard	0	10	15	945.18	770.05
Standard	0	15	20	963.07	1052.96
Standard	0	20	25	1091.52	1385.22
Standard	0	25	30	1159.14	1728.75
Standard	0	30	35	1255.60	1964.42
Standard	0	35	40	1410.76	1779.89
Standard	0	40	45	1523.41	1746.27
Standard	0	45	50	1668.10	1928.23
Standard	0	50	55	1872.95	2080.86
Standard	0	55	60	2127.02	2255.86
Standard	0	60	65	2690.87	2578.35
Standard	0	65	70	3590.27	3489.21
Standard	0	70	75	4750.00	4805.44
Standard	0	75	80	6701.95	7289.36
Standard	0	80	85	9503.21	11192.65
Standard	0	85	90	14098.30	17155.70
Standard	0	90	95	20132.95	23690.50
Standard	0	95	96	30314.66	31296.95

This table contains total health care costs per person in the life span (this data is the cost in euros from the 'Dutch Cost of Illness 1999' study) for males and females in the age groups defined with Lage and Hage. The fields 'ScenarioName' and 'Time' are currently not used.

TotalDisability

ScenarioName	Time	Lage	Hage	Males	Females
Standard	0	0	15	0.018	0.017
Standard	0	15	30	0.047	0.051
Standard	0	30	45	0.063	0.071
Standard	0	45	60	0.093	0.097
Standard	0	60	70	0.165	0.169
Standard	0	70	80	0.283	0.276
Standard	0	80	96	0.426	0.437

This table contains generic total disability weights (this data is based on the 'Global Burden of Disease' study). The fields 'ScenarioName' and 'Time' are currently not used.

Using the Prevent application

The Prevent Plus application has several option screens which might interact with each other. The next chapters will explain each setting in the application.

The ini file

Prevent uses an ini file: **Program Files\Prevent\Prevent.ini**. The contents of this file is similar to the following:

```
[Current set]
current=Test Eurocadet1

[Data sets]
x1=Test Eurocadet
x2=Test Eurocadet1

[Set paden]
xp1=C:\Program Files\Prevent\data\PreventEC.mdb
xp2=C:\Program Files\Prevent\data\PreventEC1.mdb
```

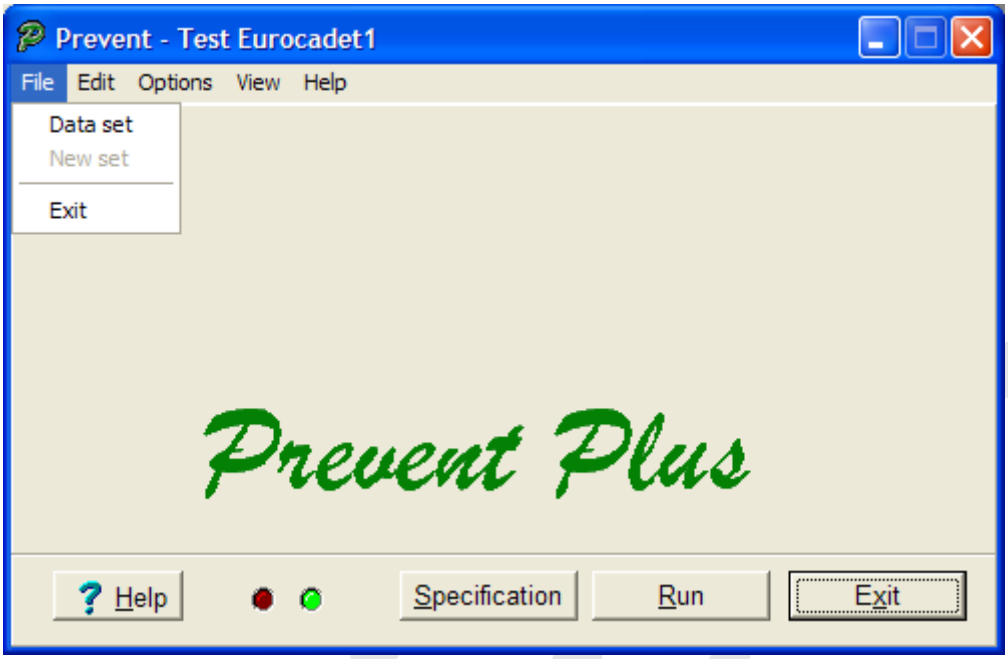
Current set defines the default dataset to load. This set is loaded even if you don't load a set manually.

Data sets defines a list of datasets which will be recognized by the application. If you define a new dataset you need to edit the ini file manually.

Set paden defines where the actual databases are located. This needs to be manually edited also.

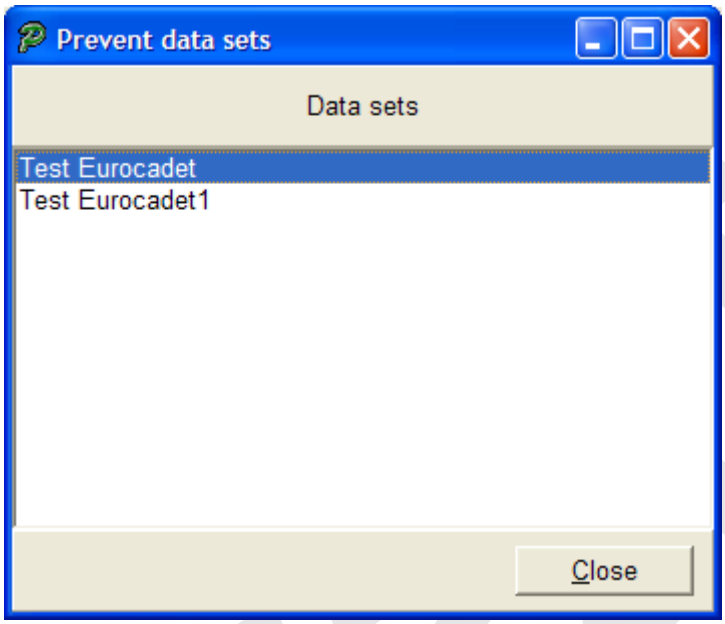
Starting Prevent

Go to **Start - Program Files** – **Prevent - Prevent** to start the application. After start up the following screen appears.



File - Data set

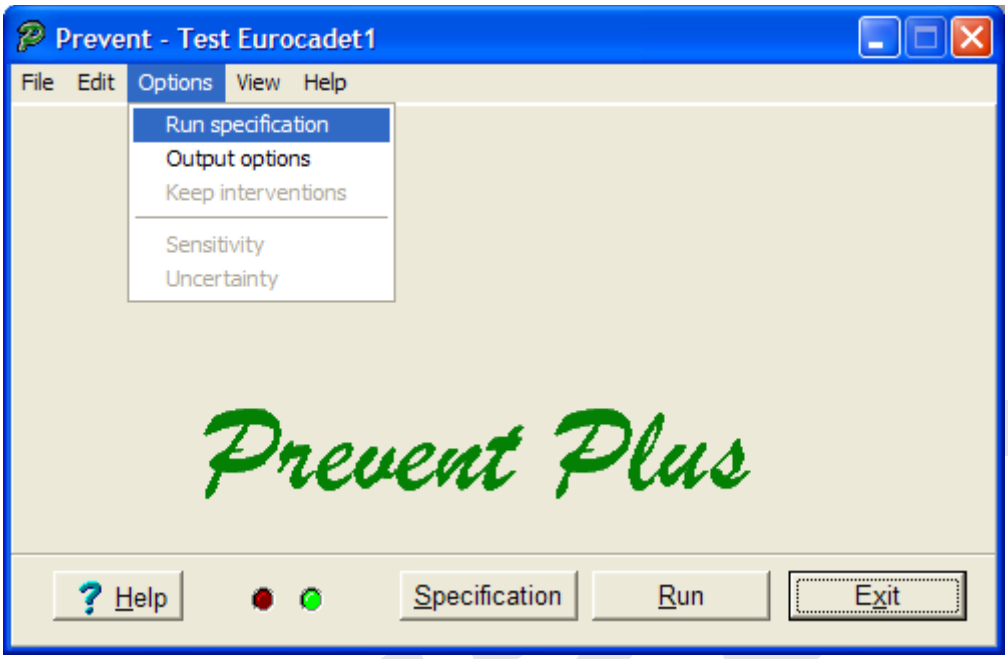
Via **File** – **Data set** the appropriate dataset can be selected. All datasets visible here must be defined in the database and their name and location must be defined in the **Prevent.ini** file.



Select the appropriate dataset and click **Close**.

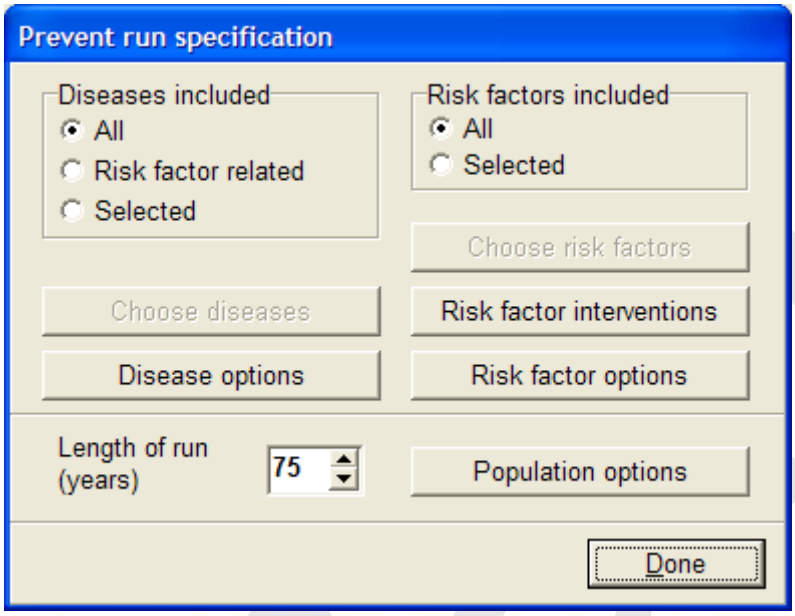
Options

The **Options** menu can be used to define the Prevent options.



Options – Run specification

In this screen you can specify all options for your run. Some items are disabled by default, but will be enabled if you select the appropriate options. For example: **Choose diseases** will be enabled as soon as **Diseases included – Selected** is selected.



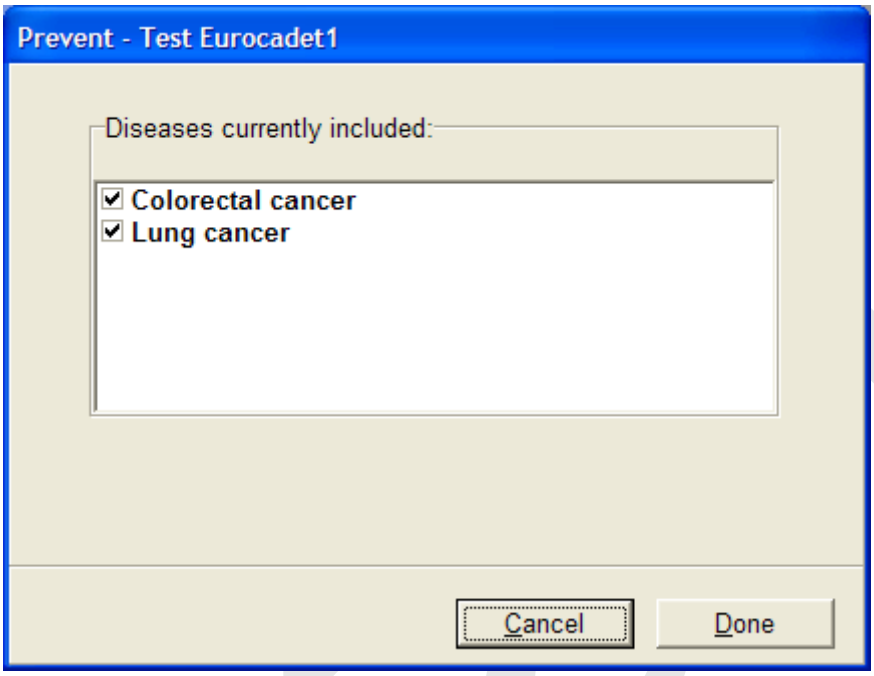
- **Diseases included**
  - Select **All** if you want to include all diseases
  - Select **Risk factor related** to only include specific diseases as defined in **DiseasesAndRiskfactors** table.
  - Select **Selected** to select specific diseases as defined in **DiseasesAndRiskfactors** table.
- **Choose diseases**
  - Select specific diseases. Only enabled if Selected above is selected. See 'Choose diseases' chapter.
- **Disease options**
  - Select options per disease. See the Disease options chapter. See 'Disease options' chapter.
- **Length of run (years)**
  - Select how long the analysis run will be in years, starting with the base year as defined in 'GeneralTab'. Maximum length equals ProjectionMax in GeneralTab.
- **Risk factors included**
  - Select **All** if you want to include all risk factors
  - Select **Selected** to only include specific risk factors
- **Choose risk factors**
  - Select specific risk factors. Only enabled if Selected above is selected. See 'Choose risk factors' chapter.
- **Risk factor interventions**

  - Select one risk factor intervention as defined in **DiseasesAndRiskfactors** table. See 'Risk factor interventions' chapter.
- **Risk factor options**
  - Select one risk factor. See 'Risk factor options' chapter.
- **Population options**
  - See 'Population options' chapter.

Please note: the run specification can also be defined by clicking the 'Specification' button on the main screen.

Options – Run specification – Choose diseases

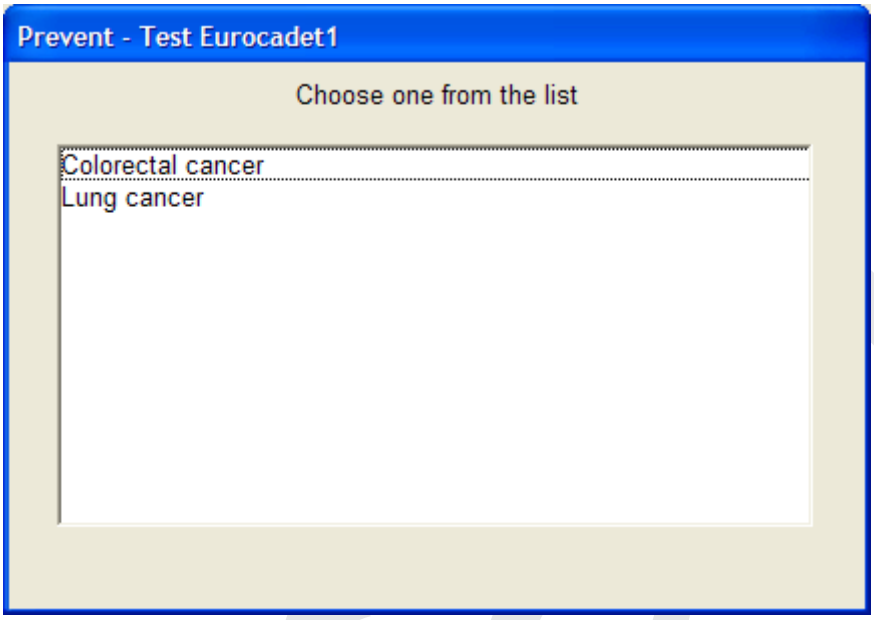
After selecting **Diseases included – Selected** the button **Choose diseases** is enabled. After clicking this button a screen similar to this appears.



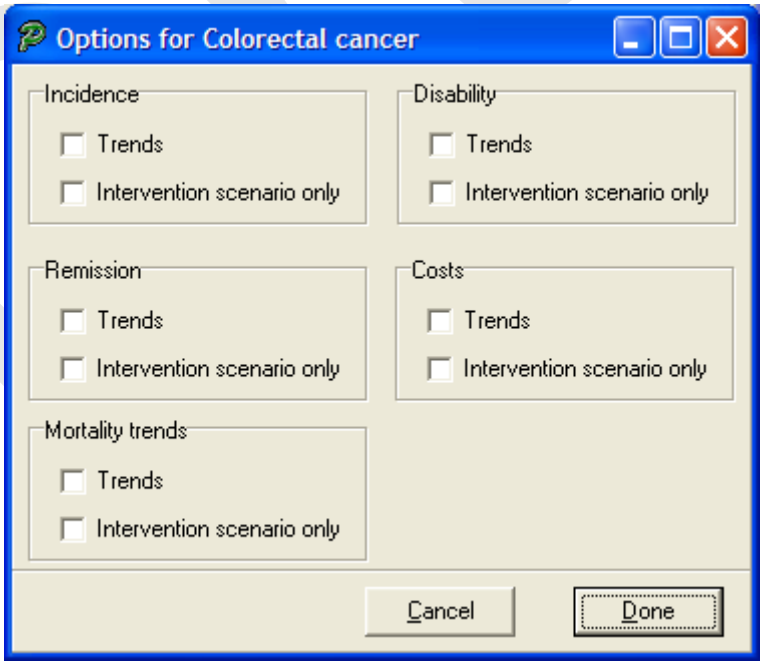
Here you can select the diseases you want to include in you analysis. Select **Done**, to accept your changes. By default, all diseases will be included.

Options – Run specification – Disease options

After clicking **Disease options** a screen appears in which a disease must be selected. If you want to use this option, the desired diseases must be defined in your database first. See description of the DiseasesAndRiskfactors table.



After selecting a disease, several options can be enabled.

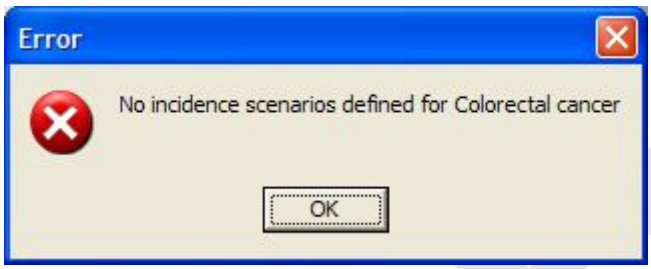


If Trends is selected, changes in incidence defined in DiseaseInterventions is used.

If intervention scenario only is selected, incidence of the intervention group will be modelled according to the given scenario in CatrfInterventions or ConrfIntervention with autonomous trend.

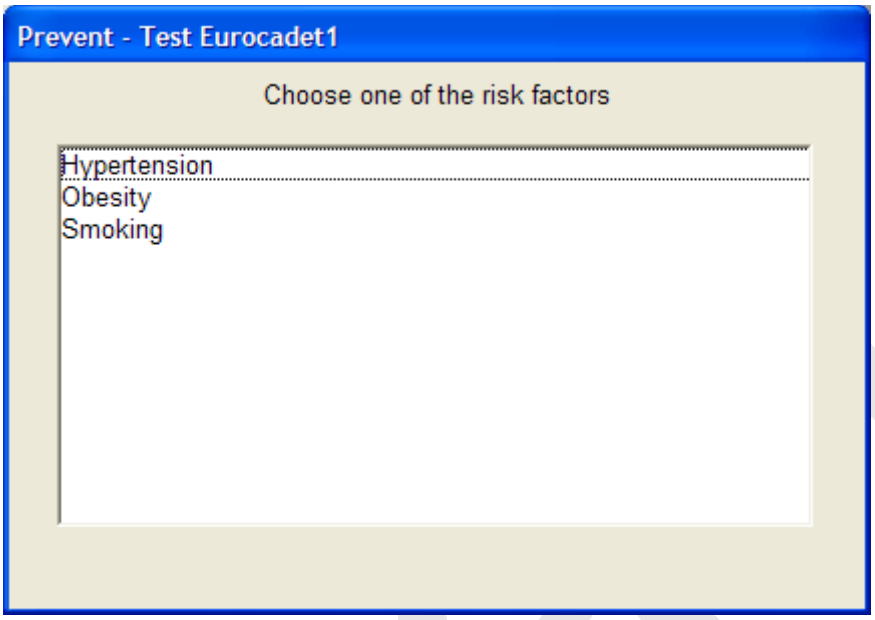
As for the reference group, the incidence will be modelled without autonomous trend described in DiseaseInterventions. Please remember to also choose the intervention scenario that you want to use for the modelling in the Options - Run specification – Risk factor interventions.

If you select an option which has no parameters defined in the DiseasesAndRiskfactors table an error screen appears:

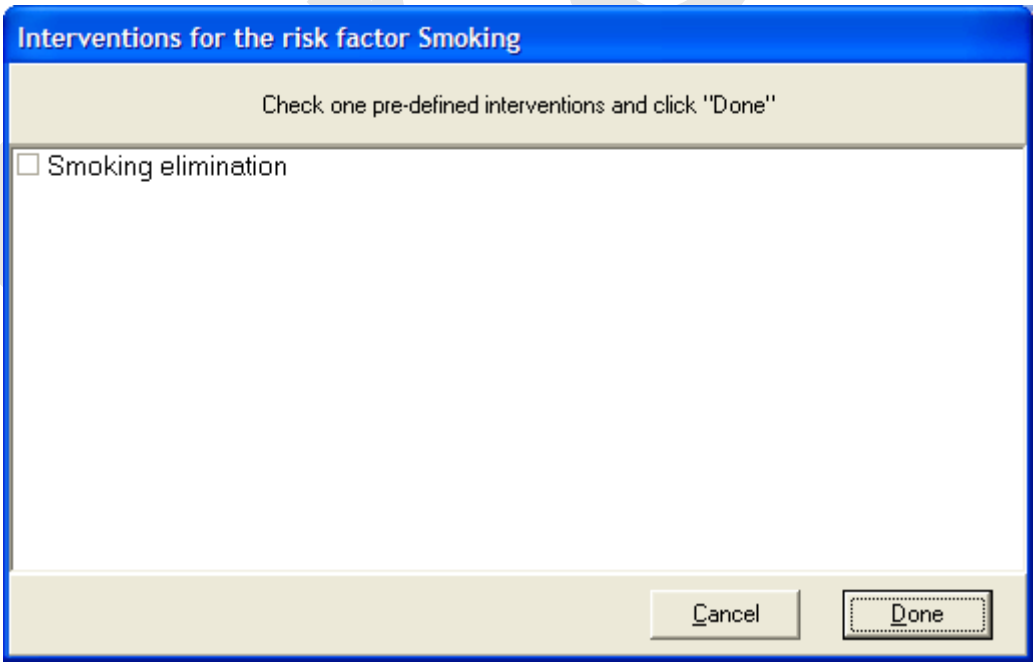


Options – Run specification – Risk factor interventions

After clicking **Risk factor interventions** a screen similar to this appears.



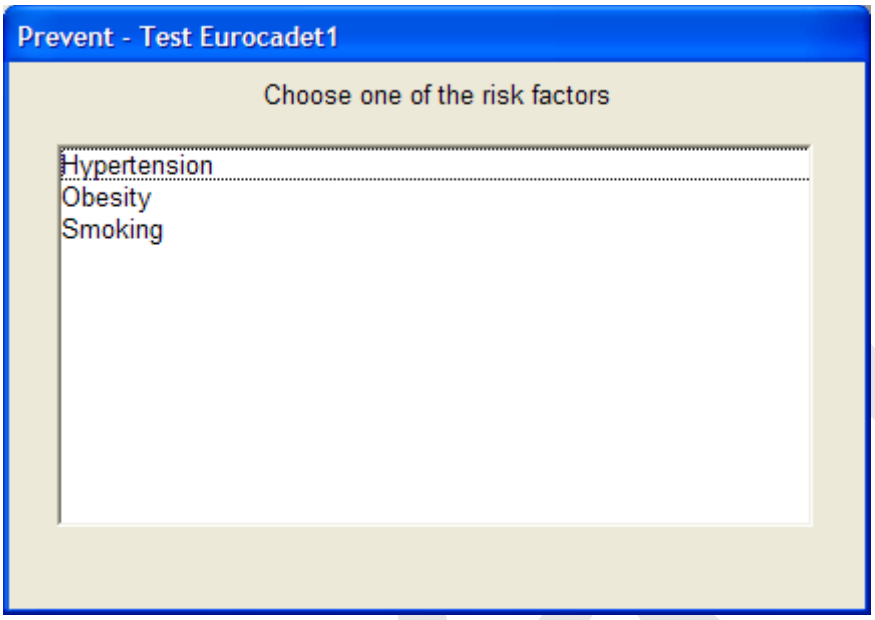
Items listed here must be defined in your DiseasesAndRiskfactors table first. When this risk factor, for example 'smoking', has an intervention defined in your database, you can select one.



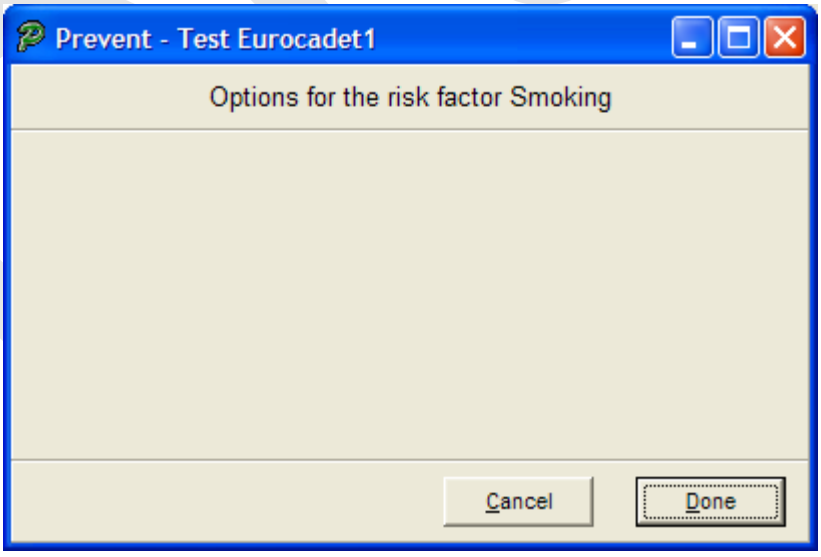
Please note: you must click on the check-mark to toggle selections

Options – Run specification – Risk factor options

After clicking **Risk factor options** a screen similar to this appears:

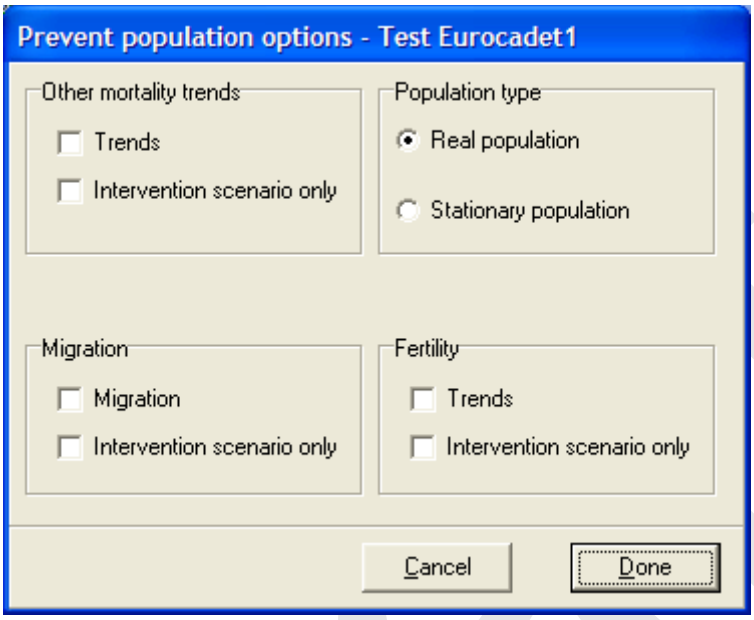


Items listed here must be defined in the DiseasesAndRiskfactors table. When this risk factor has an option defined in your database, you can select one or more.



Options – Run specification – Population options

After clicking **Population options** a screen similar to this appears:

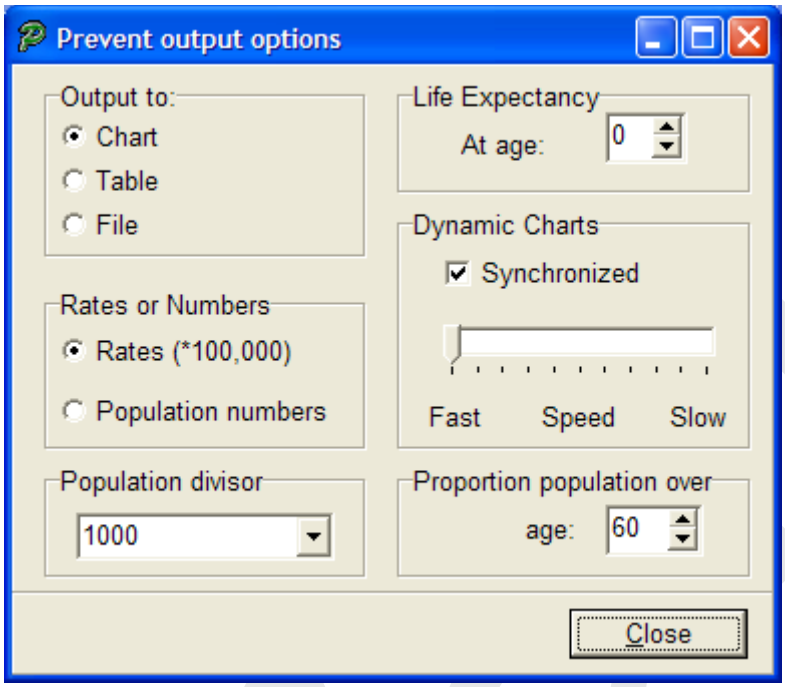


Select the options for your run.

These settings specify the population projection if data on population projection is not fed into the database by the user, thus Prevent needs to calculate the projection by itself (using these specifications). Explanation for the details of these setting is outside the scope of Eurocadet project, because within this project population projection is fed to prevent.

Options – Output options

When you have defined the Run specifications, the next step is to define the Output options.



- Output to:
  - Chart: output will be available as an interactive chart which can be used to visually explain the impact of interventions to future development of the disease.
  - Table: output will be available as a table. The data from this table can be used in other programs by using copy and paste.
  - File: output will be available in a comma separated file (csv). Please note that the file is locked by Prevent until you exit the program. After exiting Prevent the file can be used in other programs.
- Rates or Numbers
  - Rates (*100,000):
  - Population numbers
- Population divisor
  - Divisor of the numbers in the output formats: Chart, Table and File.
- Life Expectancy
  - At age: only used if Prevent needs to calculate population projection. For the Eurocadet project this must always be 0.
- Dynamic charts
  - Synchronized

- Speed: Speed of the Chart run.

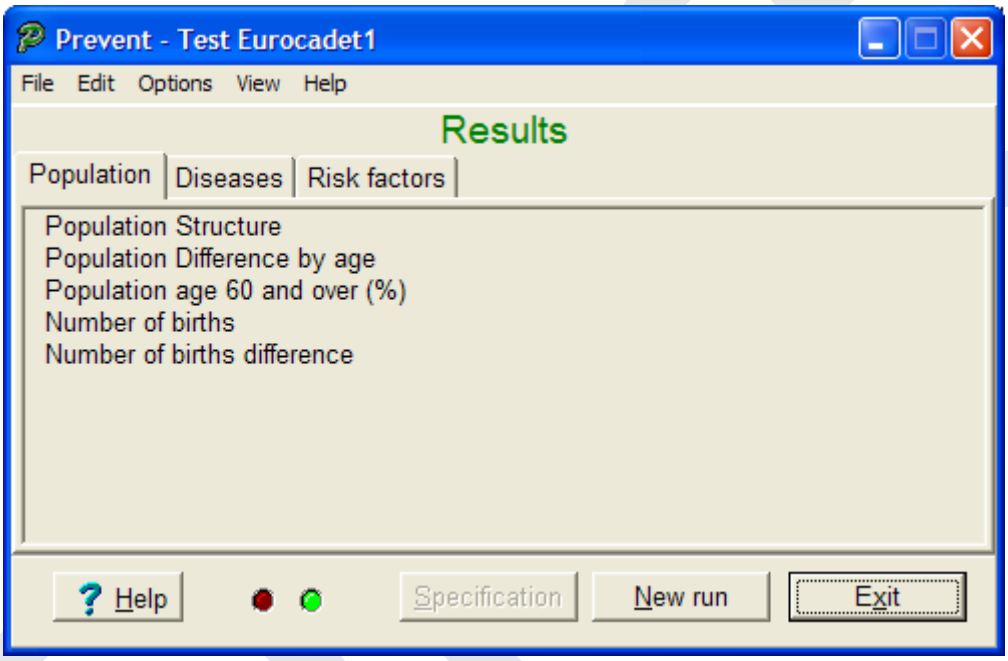
- Proportion population over
  - Specifies the output that will be used on Population – Population age 60 and over. (the number '60' will always be shown but the output will change accordingly)

Results

After setting all the options and pressing the **Run** button in the main screen, the run results are presented in 3 groups:

- Population
- Diseases
- Risk factors

Each group has its own tab:

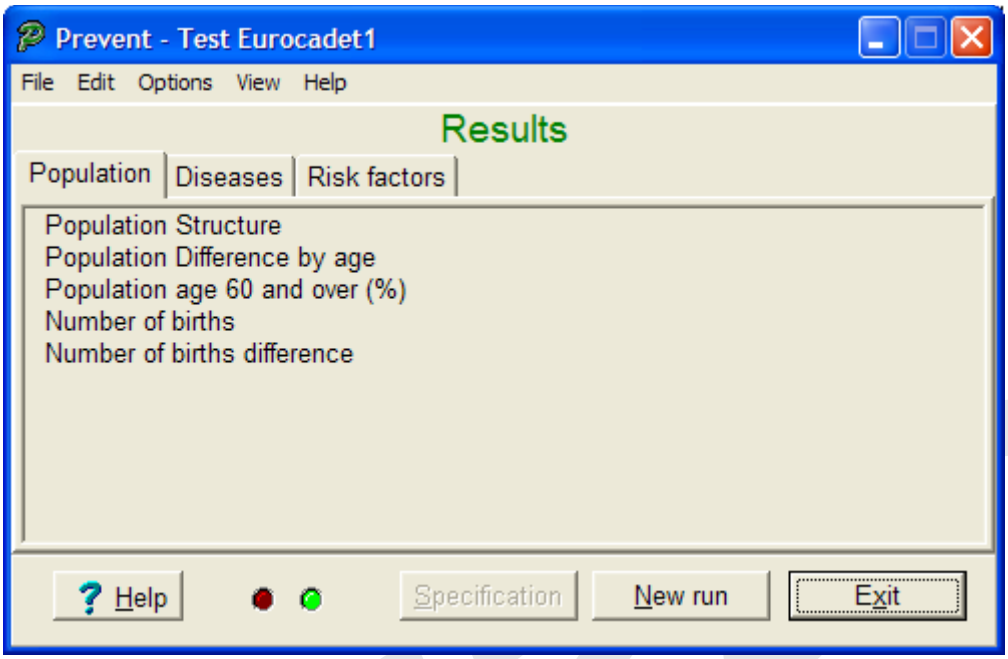


Depending on your **Output options**, results are available as a chart, table or file.

The next chapters describe the output as charts. The output as a table or file contains the same data, only in a different format. Therefore they will not be described in separately.

Results – Population

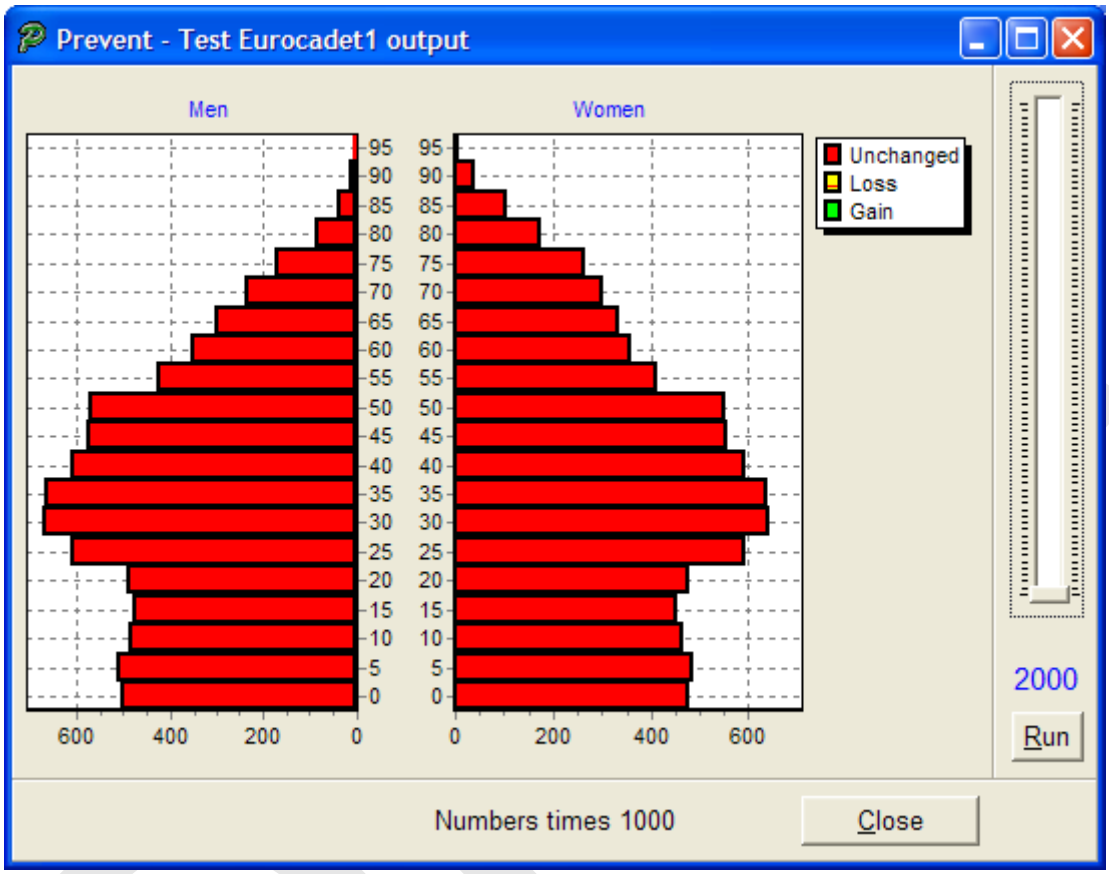
In the **Population** tab, each chart is presented as a singe line:



By clicking on a line the corresponding graph will be displayed. For the output Population data from the Population table is used.

Results – Population – Population structure

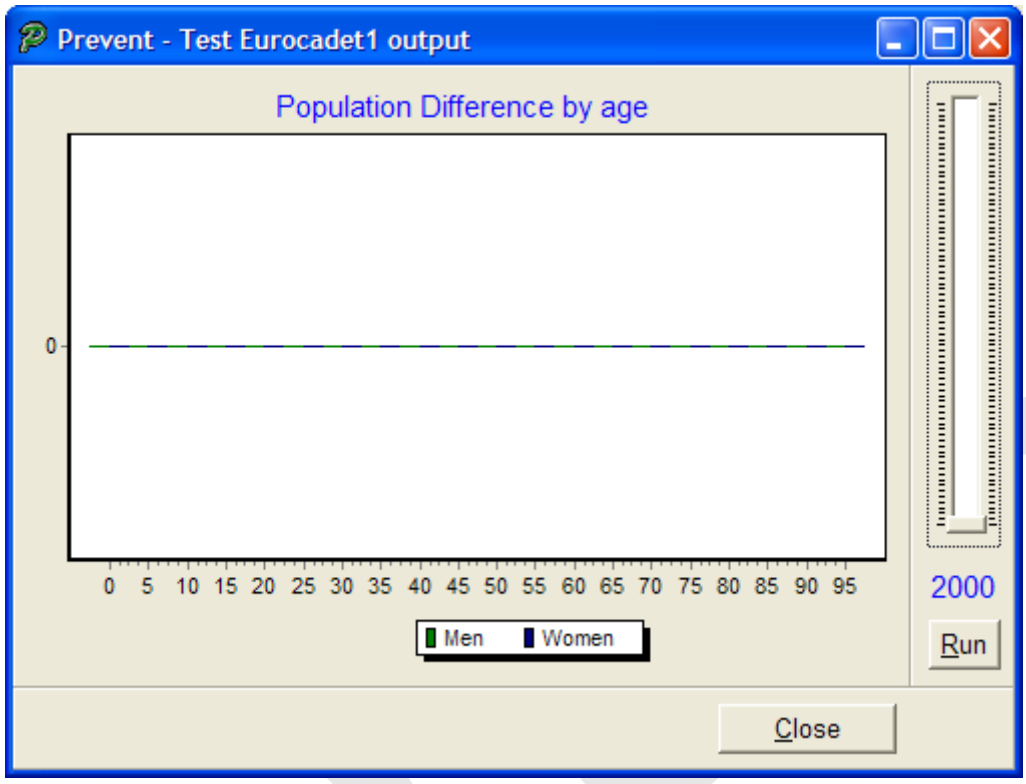
This graph displays the population structure, starting at the base year.



By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Results – Population – Population Difference by age

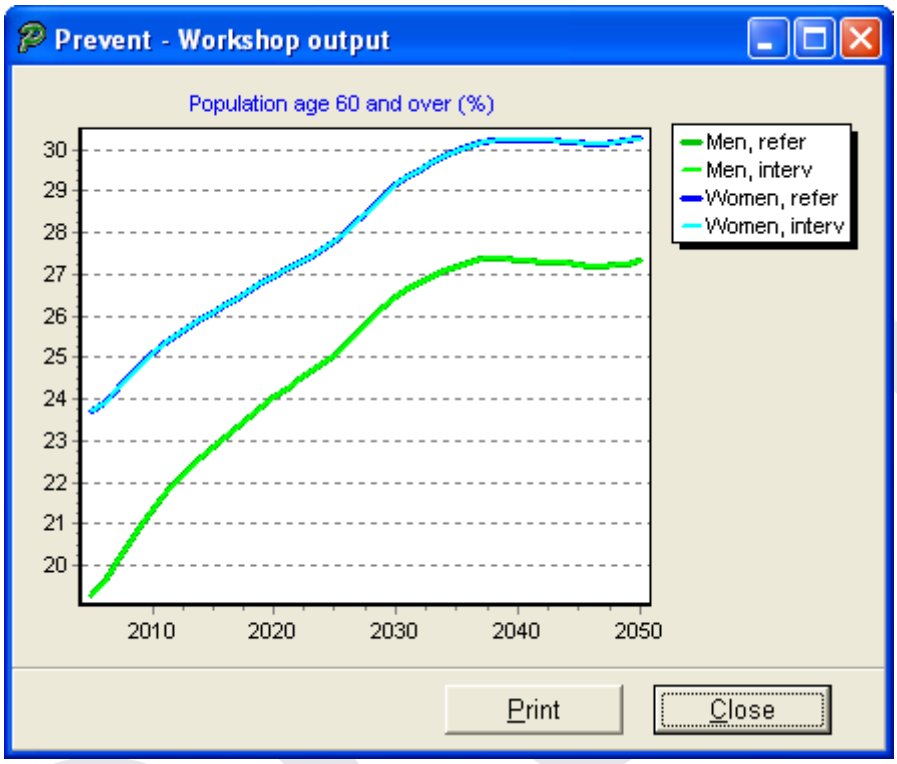
This graph displays the population difference by age, starting at the base year.



By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Results – Population – Population age 60 and over (%)

This graph displays the Population age 60 (as defined in Output Options) and over in percent, starting at the base year.

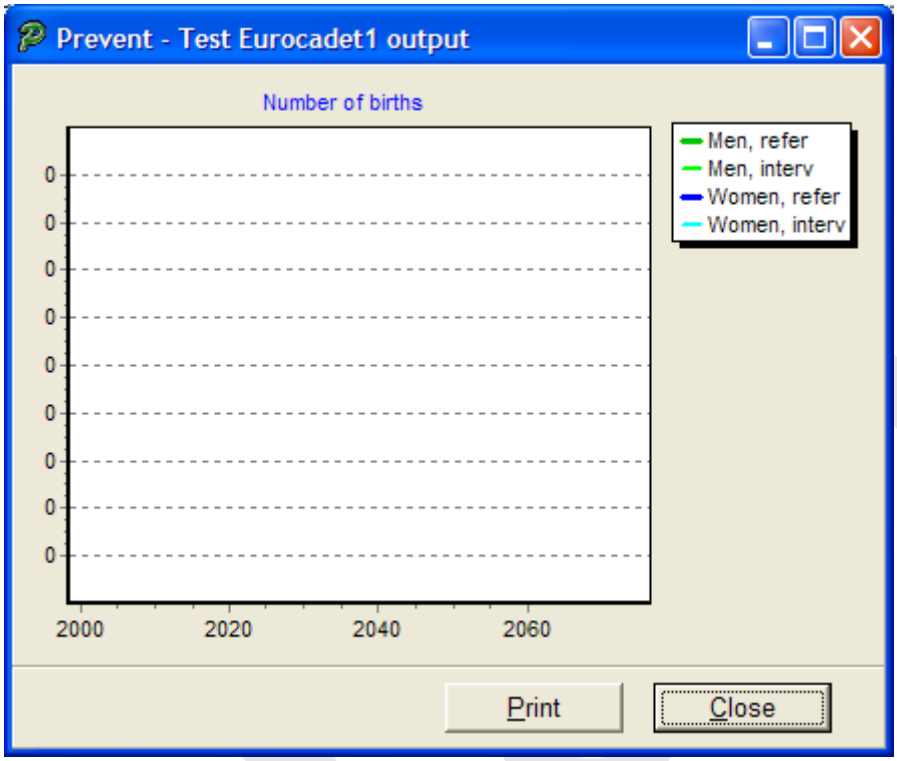


The graph can be printed by clicking the print button.

Please note: refer=reference and interv=intervention.

Results – Population – Number of births

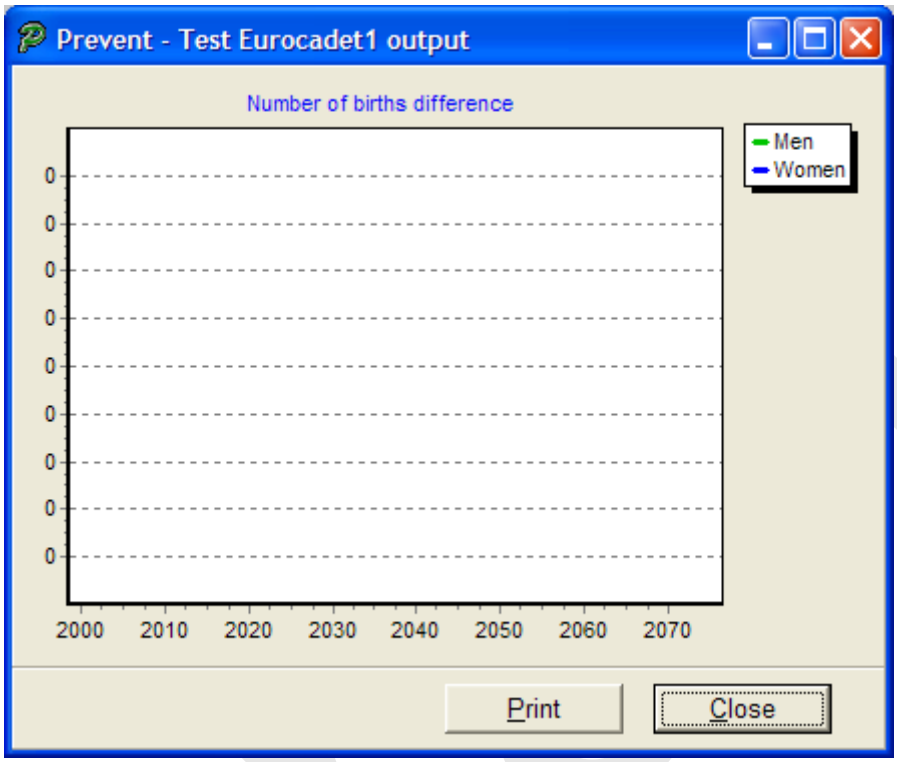
This graph displays the Number of births, starting at the base year.



The graph can be printed by clicking the print button. For the output Population data from the Population table is used.

Results – Population – Number of births difference

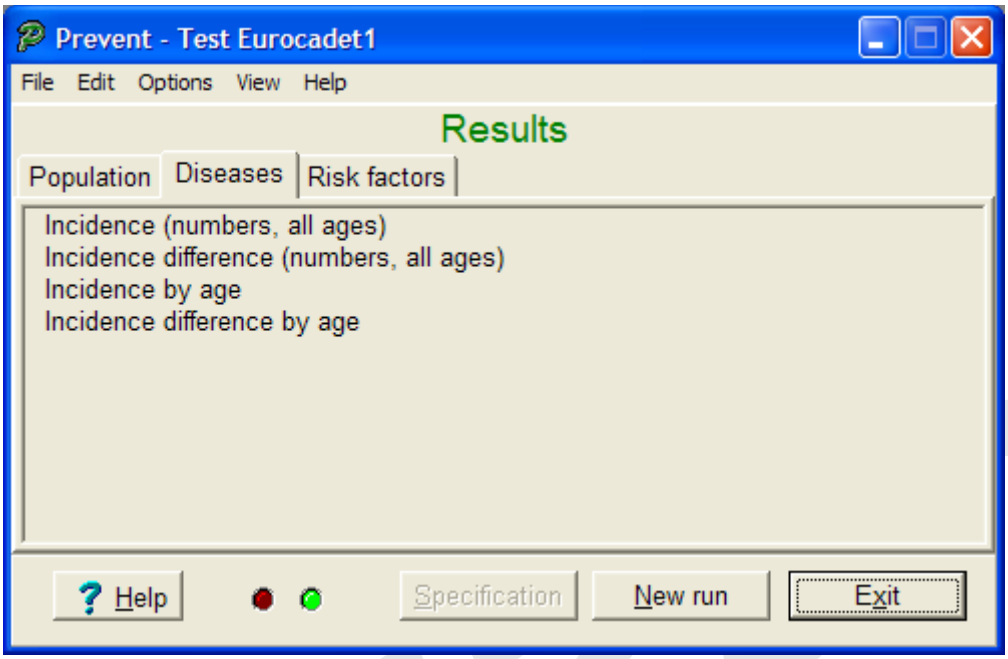
This graph displays the Number of births difference, starting at the base year.



The graph can be printed by clicking the print button.

Results – Diseases

In the **Diseases** tab, each chart is presented as a singe line:

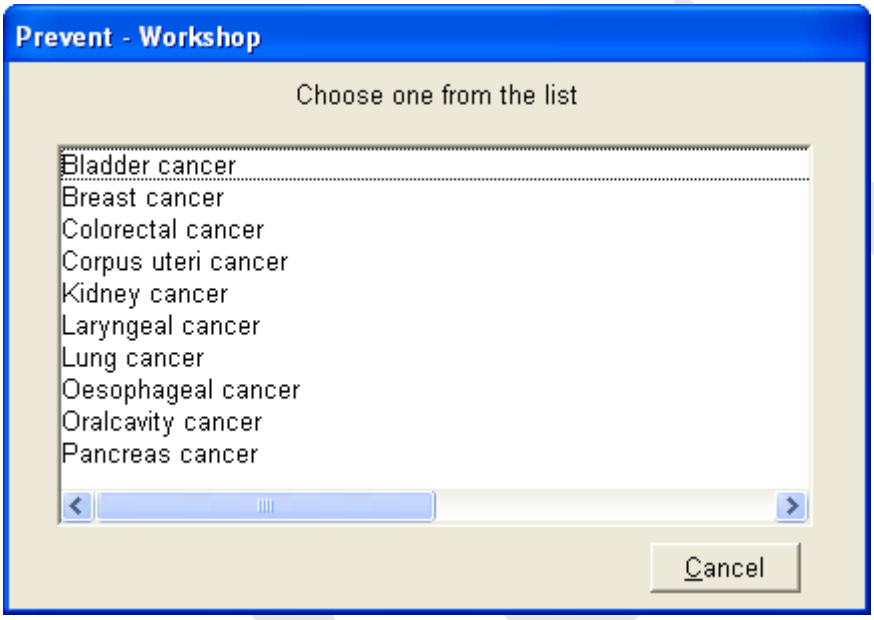


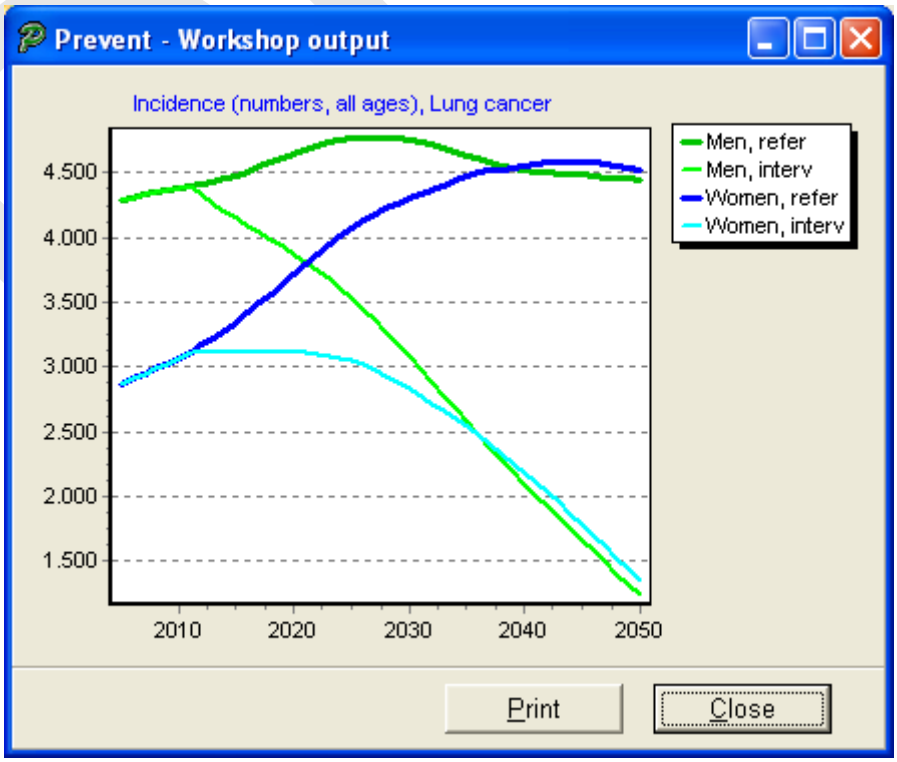
By clicking on a line the corresponding graph will be displayed.

Results – Diseases – Incidence (numbers, all ages)

This graph displays the Incidence (absolute numbers, all ages for the selected diseases), starting at the base year.

Depending on your definition of one or more diseases, you must select the disease to be displayed. 'Refer' displays the incidence of the disease without intervention, and 'interv' displays the incidence of the disease with intervention as specified in the 'CatrfInterventions' table.



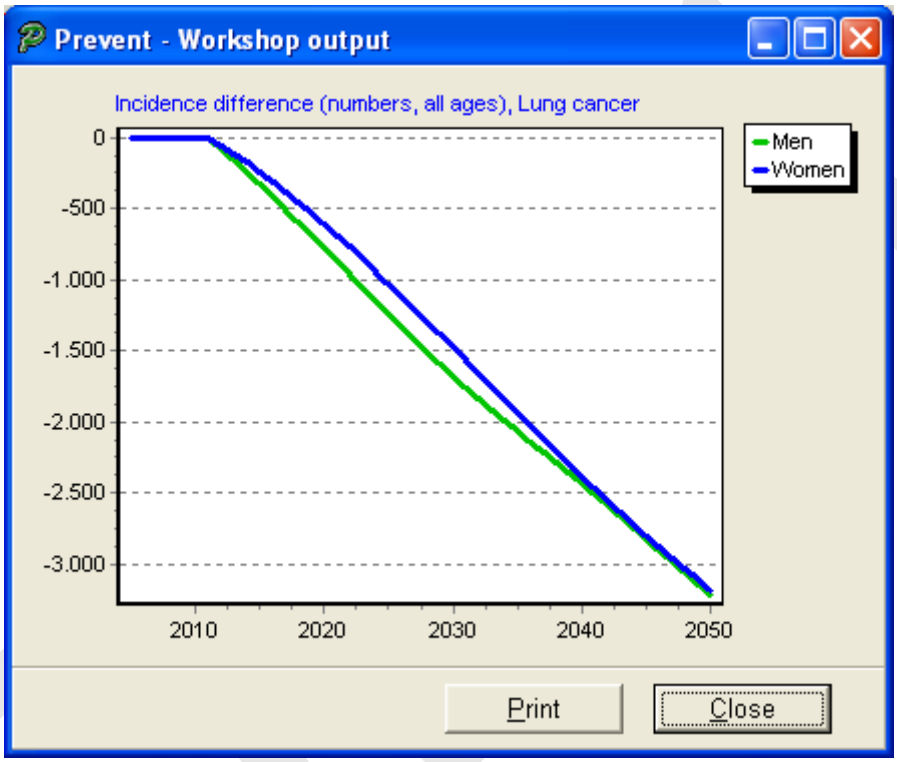


The graph can be printed by clicking the print button.

Results – Diseases – Incidence difference (numbers, all ages)

This graph displays the difference expected in the number of cases between intervention and scenario reference, starting at the base year.

Depending on your definition of one or more diseases, you will have to select the disease before the corresponding graph will be displayed.



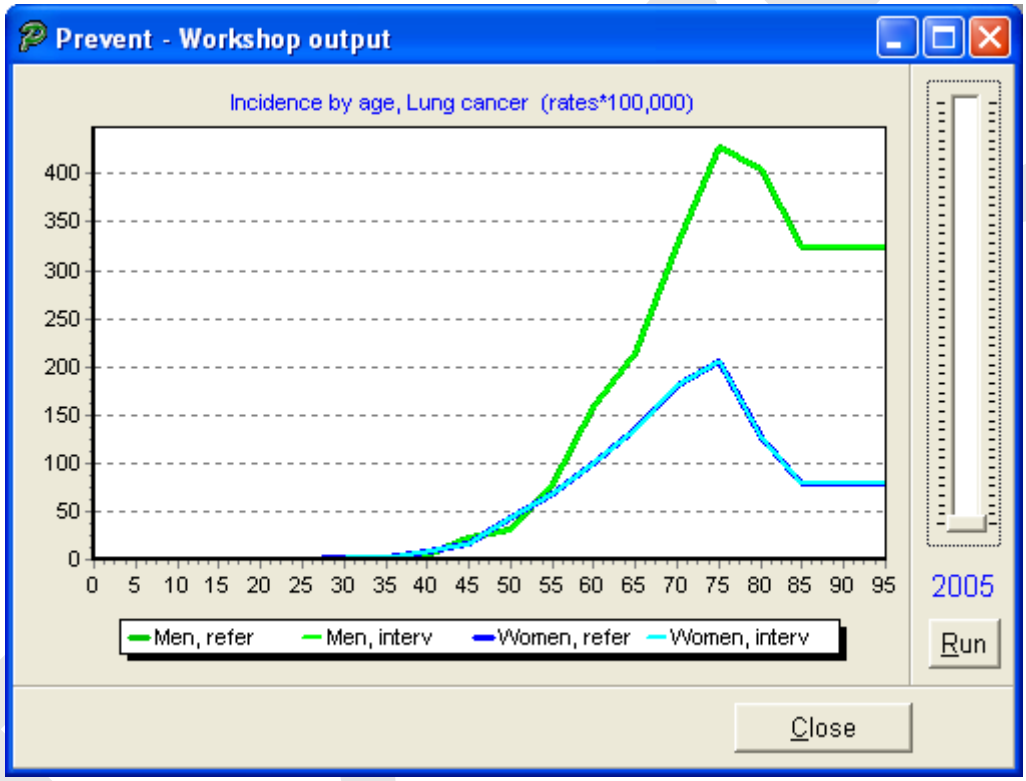
The graph can be printed by clicking the print button.

Results – Diseases – Incidence by age

This graph displays the Incidence by age, starting at the base year.

Depending on your definition of one or more diseases (in CatrfCats, CatrfExposure, CatrfRelRisk and CatrfInterventions), you will have to select the disease.

'Refer' displays the incidence of the disease without intervention, and 'interv' displays the incidence of the disease with intervention as specified in the 'CatrfInterventions' table.

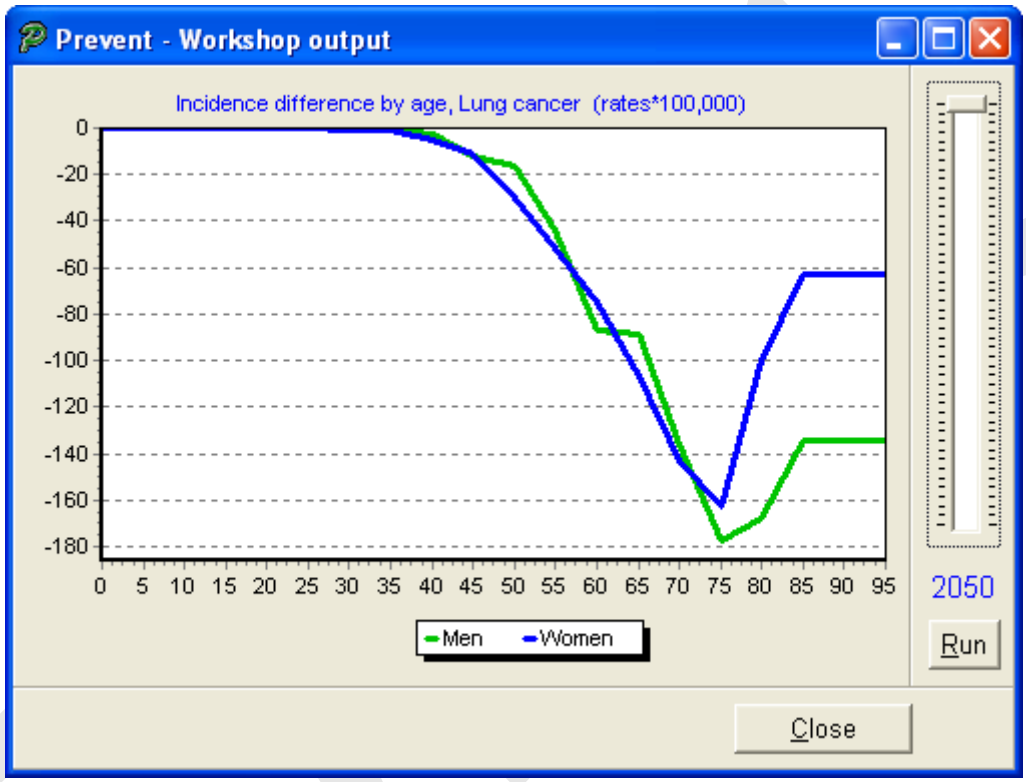


By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Results – Diseases – Incidence difference by age

This graph displays the difference in expected incidence rates between the intervention and the reference scenario by age, starting at the base year.

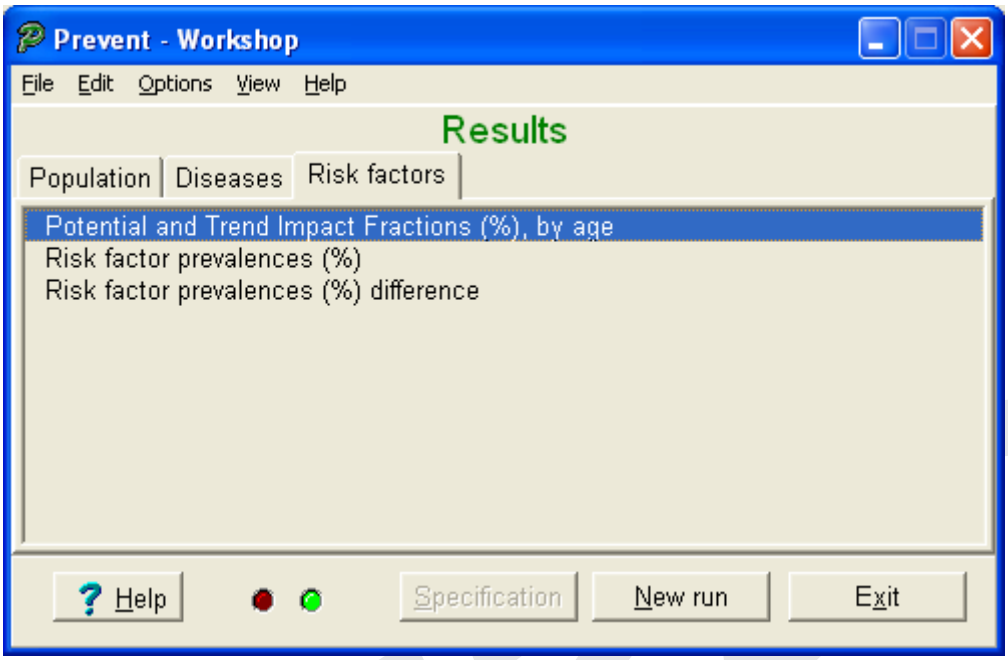
Depending on your definition of one or more diseases, you will have to select the risk factor before the corresponding graph will be displayed.



By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Results – Risk factors

In the **Risk factors** tab, each chart is presented as a singe line:

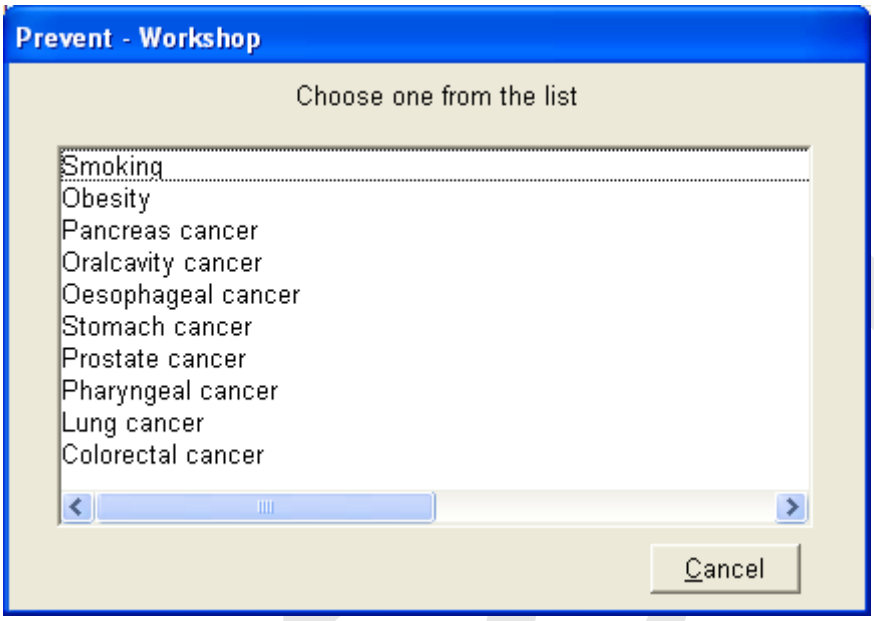


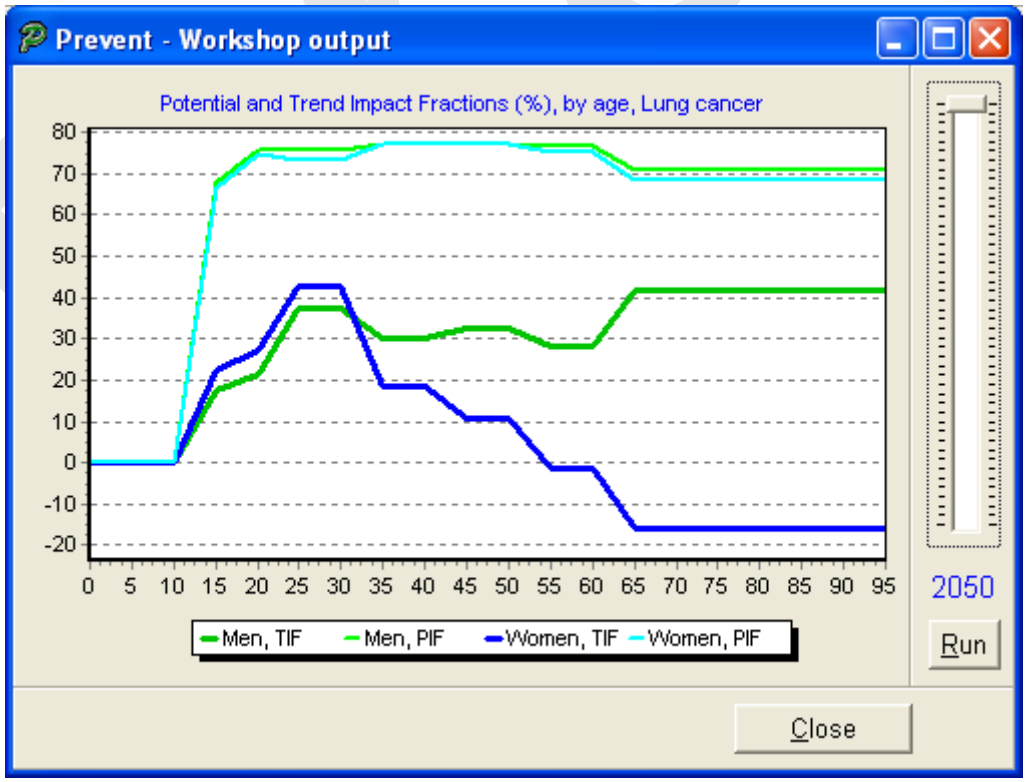
By clicking on a line the corresponding graph will be displayed.

Results – Risk factors – Potential and Trend Impact Fractions (%)

This graph displays the Potential and Trend Impact Factors (%), starting at the base year.

Depending on your definition of one or more risk factors or diseases, you will have to select the risk factor before the corresponding graph will be displayed.





By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the

slider on the right to a certain year.

PIF=Potential Impact Fraction, TIF=Trend Impact Fraction

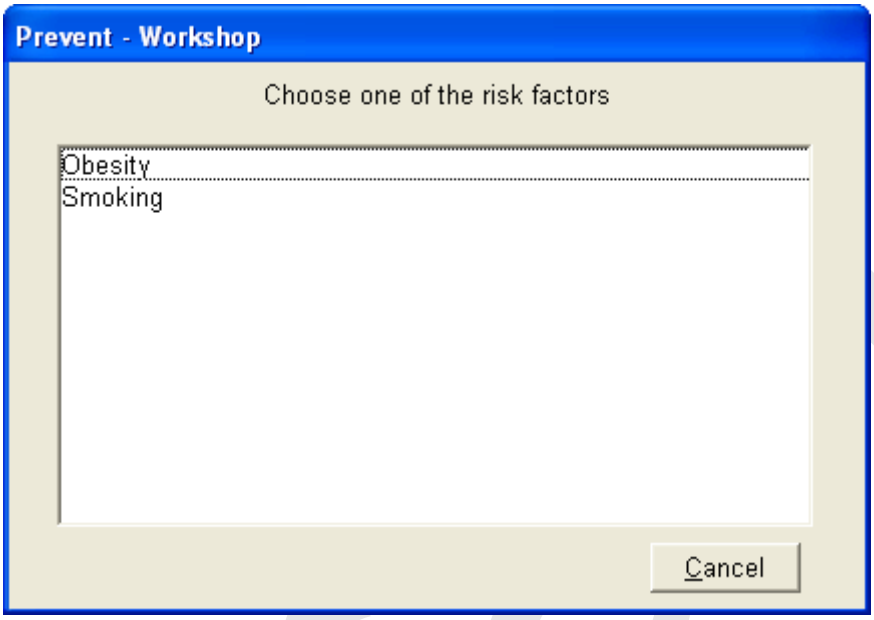
Potential and Trend Impact Fractions are variables that translate risk factor changes to disease changes.

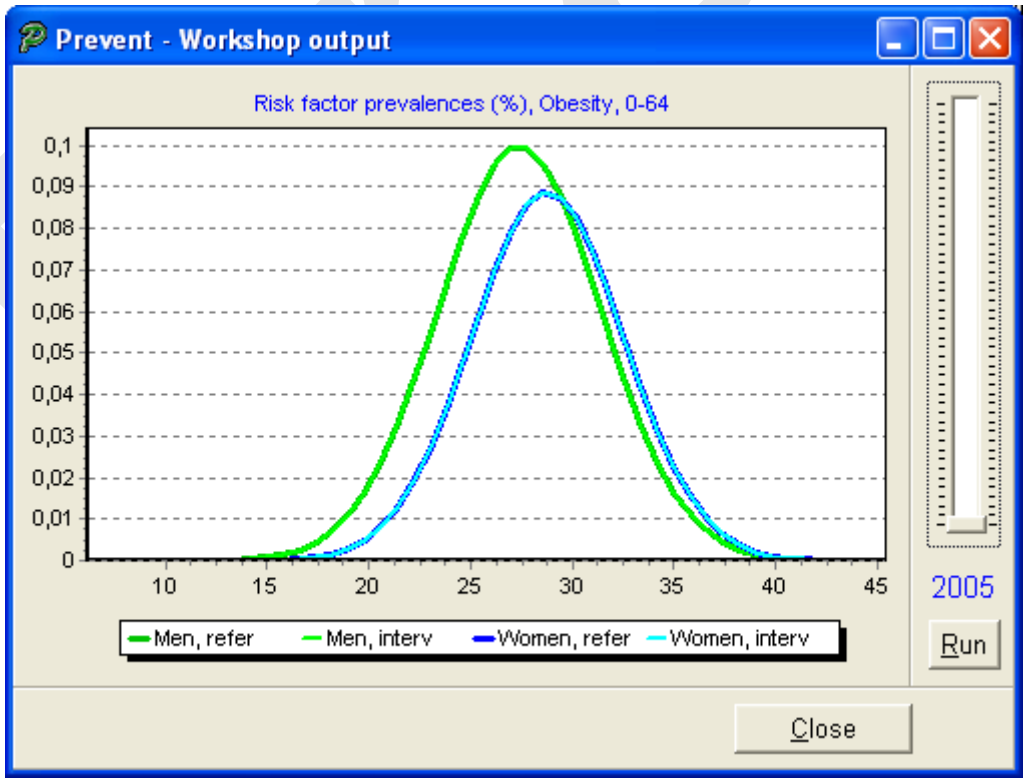
To the reference population, only the trend impact fraction is applied.

To the intervention population, both trend and potential (labelled interv) impact fractions are applied.

Results – Risk factors – Risk factor prevalences (%)

This graph displays the Risk factor prevalences (%), starting at the base year. Depending on your definition (in ConrfExposure & ConrfInterventions) of one or more risk factors, you will have to select the risk factor before the corresponding graph will be displayed.



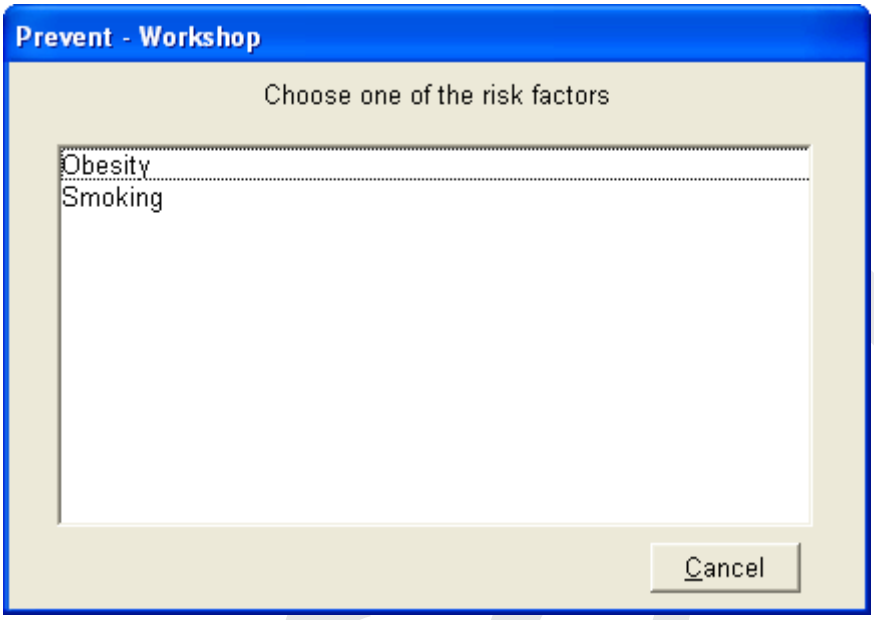


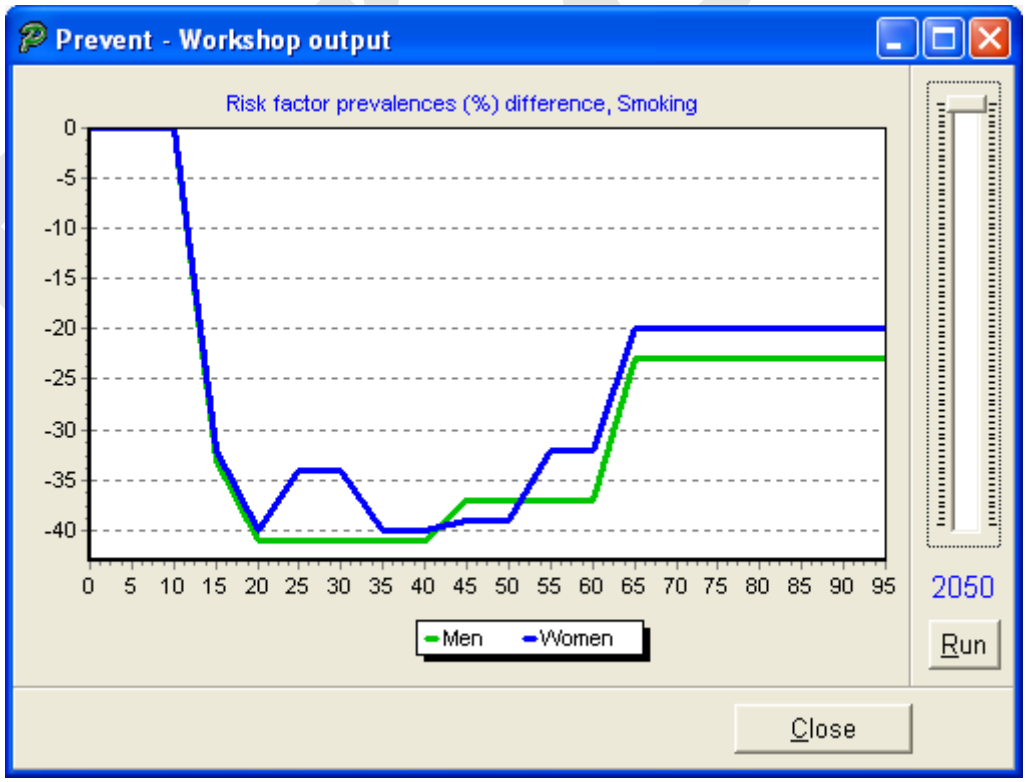
By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Graph that is produced for a categorical risk factor e.g. cigarette smoking would be the graph that add all Catno ≠ 0 in the CatrfExposure. Therefore if this group is categorised as ex-smokers and current smokers, please beware that graphical illustration includes ex-smokers.

Results – Risk factors – Risk factor prevalences (%) difference

This graph displays the changes in Risk factor prevalences (%) over time by age, starting at the base year. Depending on your definition (in ConrfExposure & ConrfInterventions) of one or more risk factors, you will have to select the risk factor before the corresponding graph will be displayed.





By clicking the **Run** button the changes over time for the number of years as specified in 'ProjectionMax' in the 'GeneralTab' table, can be observed. It is also possible to manually move the slider on the right to a certain year.

Known issues

As mentioned above, the current version of Prevent is not yet stable. A number of 'known issues' (otherwise known as 'bugs') exist, probably in addition to as yet unknown ones.

- In the output of the prevalence of categorical risk factors all categories of numbers 1 and higher are lumped together (in the calculations they are distinguished, though).
- Setting the 'lookback' field in Table DiseasesAndRiskfactors >0 for any disease or risk factor that has a risk factor itself causes the timing of the effect of changes in the disease or risk factor to go awry.

To do list

As mentioned above, Prevent is an ongoing work. A (not-exhaustive) list of things to do (with no guarantees and certainly no deadlines).

- Currently the potential impact fractions (PIFs) that drive the changes in disease variables are calculated from cross-sectional risk factor prevalences. For many risk factors lifetime exposure is a much better indicator for risk. This would require to calculate the PIFs from the total cohort exposure.
- The PIFs affect incidence only. In many cases also case fatality and, when applicable, remission will be affected by risk factor exposure. The RelRisk table already anticipates an extension to these variables.
- Combinations of two or more distributions are often a better way to describe exposure. For example alcohol, where the group of abstainers have 0 exposure, while drinkers exposure is probably best described by a skewed distribution like a lognormal.
- Implementation of a 'Scenario' feature, that would allow to bundle all settings of trends and interventions into a single scenario.

Literature

The first versions of Prevent only allowed risk factor change to affect disease specific mortality. The current version has the same functionality when in table DiseasesAndRiskfactors the field MortalityOnly= TRUE. When MortalityOnly= FALSE the risk factors influence disease incidence, but this is done in exactly the same way as mortality when MortalityOnly= TRUE. Therefore the publications based on earlier versions of Prevent are still relevant.

1. Gunning-Schepers L. The health benefits of prevention: a simulation approach. Health Policy 1989;12(1-2):1-255.
2. Gunning-Schepers LJ, Barendregt JJ, Van Der Maas PJ. Population interventions reassessed. Lancet 1989;1(8636):479-81.
3. Gunning-Schepers LJ, Barendregt JJ. Timeless epidemiology or history cannot be ignored. J Clin Epidemiol 1992;45(4):365-72.
4. Naidoo B, Thorogood M, McPherson K, Gunning-Schepers LJ. Modelling the effects of increased physical activity on coronary heart disease in England and Wales. J Epidemiol Community Health 1997;51(2):144-50.
5. Gunning-Schepers LJ. Models: instruments for evidence based policy. J Epidemiol Community Health 1999;53(5):263.
6. Mooy JM, Gunning-Schepers LJ. Computer-assisted health impact assessment for intersectoral health policy. Health Policy 2001;57(3):169-77.

The following three publications also refer to the mortality only version, but are by an independent researcher.

7. Brønnum-Hansen H. How good is the Prevent model for estimating the health benefits of prevention? J Epidemiol Community Health 1999;53(5):300-5.
8. Brønnum-Hansen H, Juel K. Estimating mortality due to cigarette smoking: two methods, same result. Epidemiology 2000;11(4):422-6.
9. Brønnum-Hansen H. Predicting the effect of prevention of ischaemic heart disease. Scand J Public Health 2002;30(1):5-11. The extension to morbidity with the Remission field in table DiseasesAndRiskfactors put to FALSE is based on a disease model described in the following publication:
10. Barendregt JJ, Oortmarssen GJ, van, Hout BA, van, Bosch JM, van den, Bonneux L. Coping with multiple morbidity in a life table. Mathematical Population Studies 1998;7(1):29-49.

Several publications with additional details or applications are:

11. Barendregt JJ, Bonneux L, van der Maas PJ. The health care costs of smoking. N Engl J Med 1997;337(15):1052-7.
12. Barendregt JJ, Bonneux L, Maas PJ, van der. When does nonsmoking save health care money? The many answers to a simple question. In: Jeanrenaud C, Soguel S, editors. Valuing the cost of smoking. Assessment methods, risk perception and policy options. Boston, Dordrecht, London: Kluwer Academic Publishers; 1999. p. 75-91.
13. Barendregt JJ. Prevent: the technical background. In: Public Health Models: Tools for health policy making at national and European level. Amsterdam: ISG/UvA; 1999.
14. Barendregt JJ, Bonneux L, Baan CA. The effects and costs of smoking cessation. In: Public Health Models: Tools for health policy making at national and European level. Amsterdam: ISG/UvA; 1999.

The disease model used when the Remission field in table DiseasesAndRiskfactors put to TRUE is the same as the model used by DisMod II, and described in:

15. Barendregt JJ, Oortmarssen GJ, van, Vos T, Murray CJL. A generic model for the assessment of disease epidemiology: the computational basis of DisMod II. Population Health Metrics 2003;1(1):4.

The concept of theoretical minimum risk is developed and applied in:

16. Ezatti M, Lopez AD, Rodgers A, Murray CJL, editors. Comparative Quantification of Health Risks. Global and Regional Burden of Disease Attributable to Selected Major Risk Factors. Geneva: World Health Organization; 2004.

A description of and Fortran code for the Romberg numerical integration routine can be found in:

17. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes in Fortran 77: The Art of Scientific Computing. Cambridge: Cambridge University press; 1992. The Fortran code was converted to Object Pascal for implementation in Prevent.

Internal consistency of epidemiological variables is discussed in:

18. Barendregt JJ, Ott A. Consistency of epidemiologic estimates. European Journal of Epidemiology 2005;20(10):827-832.

Index