

NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT
BILTHOVEN, THE NETHERLANDS

Report no. 266750001

The chronic diseases modelling approach

RT Hoogenveen, AEM de Hollander,
MLL van Genugten

May 1998

This investigation has been performed by order and for the account of the Board of Directors of RIVM

National Institute of Public Health and the Environment (RIVM), PO Box 1, 3720 BA Bilthoven. The Netherlands

Telephone: +31 30 2749111, Telefax: +31 30 2742971

MAILING LIST

- 1 Directeur-Generaal RIVM
- 2 Directeur-Generaal van de Volksgezondheid
- 3 Hoofdinspecteur voor de Gezondheidszorg
- 4 Dr J.J.M. Barendregt, Instituut Maatschappelijke Gezondheidszorg, EUR, Rotterdam
- 5 Dr. J. De Beer, Centraal Bureau voor de Statistiek, Voorburg
- 6 Prof dr L. J. Gunning-Schepers, Instituut Sociale Geneeskunde, UvA, Amsterdam
- 7 Dr E. Tabeau, Nederlands Interuniversitair Demografisch Instituut, Den Haag
- 8 Prof dr O. J. Vrieze, vakgroep Algemene Wetenschappen, RL, Maastricht
- 9 Prof dr F. J. Willekens, Demografisch Instituut, RUG, Groningen
- 10 Prof dr M. Wolfson, Statistics Canada, Montreal
- 11 Prof dr ir D. Kromhout, directeur Sector Volksgezondheidsonderzoek, RIVM, Bilthoven
- 12 Dr P. G. N. Kramers, hoofd Centrum voor Volksgezondheid Toekomst Verkenningen
- 13 Dr ir J. C. Seidell, hoofd Centrum voor Chronische Ziekten en Milieu-Epidemiologie
- 14-23 Leden Werkgroep Strategisch Overleg Modellen Sector 2, RIVM
- 24-26 Auteurs
- 27 Bureau Voorlichting en Public Relations RIVM
- 28 Bibliotheek RIVM
- 29 Bureau Rapportenregistratie RIVM
- 30-100 Bureau Rapportenbeheer RIVM

TABLE OF CONTENTS

Mailing list	3
Preface	7
Summary	9
Samenvatting	11
1. Introduction	13
2. The chronic diseases model structure	17
2.1. The general structure	17
2.2. The population and risk factor modules	19
2.3. The disease modules describing morbidity and mortality	19
2.4. Main model characteristics	20
2.5. Risk factors and diseases being distinguished	23
2.6. Main model parameters	23
3. Some model analyses and exercises	27
3.1. Model validation	27
3.2. Some general model validation results	28
3.3. First example application: comparing cause-specific mortality time series extrapolations to model predictions	30
3.4. Second model application: the effectiveness of different policies to reduce tobacco smoking in terms of future smoking prevalence rates and burden of disease	32
3.5. Third model application: the effectiveness of different policies on increasing physical activity	34
4. General comments and conclusions	37
4.1. Comments on the model characteristics	37
4.2. Future model applications and development	40
Appendix A. The main model equations	43
A.1. The basic model variables	43
A.2. The basic model equations	43
A.3. The interpretation of the state transition rates	44
A.4. Distinguishing causes of death	44
A.5. Computer model implementation of basic model equations	45
A.6. Some remarks	46
Appendix B. Distinguishing causes of death and competing death risks	49
B.1. The main model assumptions related to mortality	49
B.2. The main model equations	49
B.3. Some remarks	50
Appendix C. Relative risks in the model structure	53
C.1. The risk factor-specific incidence rates	53
C.2. Comparing relative risks for the outcome disease incidence and mortality	53
C.3. Risk ratio and relative risk	57

C.4. Comorbidity	58
C.5. Multiplicative or additive risk model	58
C.6. Some remarks	60
Appendix D. Change of disease prevalence over time and age	61
D.1. The main model equations	61
D.2. Relating time trends in disease incidence, prevalence and case fatality rate	62
Appendix E. Health outcome indicators	63
E.1. Life and remaining life expectancy	63
E.2. Reduction of life expectancy due to high risk	63
E.3. Other versions of life expectancy	64
Appendix F. Definitions	65
Appendix G. Abbreviations	67
Literature	69

PREFACE

Simulation models based on the multistate lifetable method are increasingly used to estimate the effects over time and age of public health intervention programs (for example Weinstein et al., 1987; Gunning-Schepers, 1988; Barendregt, Bonneux, 1998). These models describe several population risk factors and disease categories simultaneously taking into account some main integrative aspects. The models share the same mathematical methodology, but differ in the selection of risk factors, diseases, model specification and data used. Also at the National Institute of Public Health and the Environment (RIVM) we apply these types of models to illustrate the public health relevance of risk factors and the health gains associated with interventions on them. The example model applications are mainly presented in relatively small formats, such as scientific articles or paragraphs in books. So far a thorough account of the structure and basic assumptions of the RIVM applications of the chronic diseases model structure has lacked. This report is meant to fill this gap. This report addresses those who are already working with these types of models and those that have plan to do so. This report aims at: (1) giving a mathematical-methodological introduction to the chronic diseases modelling approach, (2) showing how so-called integral public health aspects can be described formally, (3) presenting a methodological basis upon which models for specific risk factors and diseases can and will be developed.

SUMMARY

A mathematical model structure is described that can be used to simulate the changes of the Dutch public health state over time. The model is based on the concept of demographic and epidemiologic processes (events) and is mathematically based on the life-table method. The population is divided over several states, over several risk factor classes and disease stadia. State transitions over time are possible due to birth, aging, migration, mortality, transitions between risk factor classes, disease incidence, progression and remission. The main model parameters are initial population numbers, initial risk factor class and disease stadium prevalence rates, one-year transition rates between the risk factor classes and disease stadia, and risk factor-cause-specific relative risks.

The model is used to describe the public health effects of possible intervention programs. These effects can be defined in terms of classic epidemiologic morbidity and mortality figures, but also in terms of life and health expectancy. The main model steering parameters are the one-year transition rates between the risk factor classes (concerning primary prevention) and the disease progression, remission and case fatality rates (concerning secondary prevention). Several examples of model applications are described: comparing trend extrapolations and model predictions on cause-specific mortality, and comparing the effects of different intervention programs on population smoking and physical activity levels.

In appendices several aspects of the model structure are worked out. The model equations are formally derived by expliciting the necessary assumptions. The central assumption is that of conditional independence. It is described how integrative health measures can be calculated from the model output. It is shown how competing death risks are taken account of. Extra attention is paid to the use of relative risks in the model structure and the relation with comorbidity. Finally the main model equation on the disease prevalence is described and how this equation can be used to estimate the case fatality rate parameters.

SAMENVATTING

Dit rapport geeft een beschrijving van een wiskundige modelstructuur, waarmee veranderingen van de gezondheidstoestand van de Nederlandse bevolking gesimuleerd kunnen worden. Het model is gebaseerd op het concept van demografische en epidemiologische processen (gebeurtenissen) en is afgeleid van de overlevingstafel.

De bevolking wordt verdeeld over verschillende mogelijke toestanden, namelijk voor onderscheiden risicofactoren in verschillende klassen en voor onderscheiden ziekten in één of meer stadia. Toestandsveranderingen zijn mogelijk ten gevolge van geboorte, veroudering, migratie, sterfte, verandering van risicofactor-klasse, incidentie, voortschrijding van de ziekte en remissie. De belangrijkste modelparameters zijn initiële bevolkingsaantallen, initiële risicofactor en ziekteprevalentiefractionen, éénjaars overgangskansen tussen de risicofactor-klassen en ziektestadia, en risicofactor-oorzaak-specifieke relatieve risico's.

Het model wordt gebruikt om de gezondheidseffecten door te rekenen van mogelijk beleidsmaatregelen, interventies etc. De effecten kunnen gedefinieerd worden in termen van klassieke epidemiologische ziekte- en sterfematen, maar ook in termen van levensverwachting en ziektelast. De betreffende model-stuurparameters zijn met name de éénjaars overgangskansen tussen de risicofactor-klassen (als het gaat om primaire preventie), en de ziekteprogressie, -remissie en case fatality rates (als het gaat om secundaire preventie). Enkele voorbeelden worden beschreven van verschillende modeltoepassingen: een vergelijking van trendextrapolaties en modelmatige vooruitberekeningen voor oorzaak-specifieke sterfte, en een vergelijking van de effectiviteit van verschillende mogelijke anti-roken en meer-bewegen campagnes.

In bijlagen worden verschillende aspecten verder uitgewerkt. Aangegeven wordt onder welke veronderstellingen de modelvergelijkingen afgeleid kunnen worden. Hierbij staat met name de veronderstelling van conditionele onafhankelijkheid centraal. Beschreven wordt hoe integrale gezondheidsmaten uit de modeluitvoer berekend worden. Toegelicht wordt hoe rekening gehouden wordt met het aspect van concurrerende doodsoorzaken. Extra aandacht wordt geschonken aan het gebruik van relatieve risico's in de modelstructuur en de relatie met comorbiditeit. Tenslotte wordt beschreven hoe de kern-modelvergelijking betreffende de ziekteprevalentie luidt en hoe deze gebruikt kan worden om case fatality rates te schatten.

1. INTRODUCTION

Several examples of simulation models on chronic diseases have been developed and used for public health policy evaluations, for example Weinstein et al., 1987; Gunning-Schepers, 1988; Barendregt, Bonneux, 1998. Also at the National Institute of Public Health and the Environment (RIVM) we are applying these types of models as a tool to describe chronic diseases and their relations to population risk factors. The tool can be used to describe the effects of intervention programs on public health status. Some example scientific and policy questions that can be dealt with are:

- Predicting trends in the Dutch population structure, risk factor and disease prevalence numbers taking into account the epidemiologic relationships. For example the trend in coronary heart disease mortality resulting from trends in the smoking, hypertension and hypercholesterolemia prevalence rates.
- Analyzing the effects of prevention and intervention programs in disease-specific terms and in terms of the avoidable burden of disease.
- Analyzing the effects of competing morbidity and mortality risks.
- Performing scenario studies on trends in demography, risk factor and disease prevalence resulting from autonomous trends or public health prevention and intervention programs.

The model structure has been developed to integrate several public health aspects, i.e. to describe several disease risk factors and their effects on several diseases and causes of death simultaneously. The main model parameters are initial risk factor and disease prevalence rates, disease incidence and case fatality rates, and relative risks. Public health intervention programs can be simulated by changing the model steering parameters. The main steering parameters are the rates of change of the risk factor prevalence rates and the disease progression and case fatality rates. The model output variables are time series of disease incidence, prevalence and mortality numbers and public health indicator variables. In this report the general model structure is described, together with some results of specific model exercises.

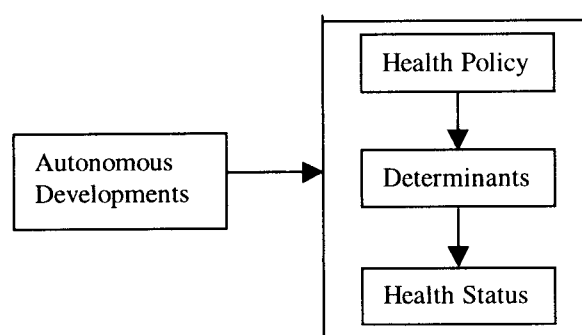


Figure 1: the VTV-conceptual model

The model structure is related to the conceptual model structure devised for the 'Public Health Status and Forecasts' (PHSF) (RIVM, 1997) that is used to frame the different aspects of public health and public health policy (see figure 1). The model shows that health status is defined by determinants that are affected by health policy. Autonomous developments may affect these three public health aspects and their relations. The health status can be thought of as forming several layers. These layers are disease-specific indicators like disease incidence and prevalence figures, the consequences of diseases and disorders for physical, mental and social functioning, mortality figures, and measures that combine information on mortality and quality of life. The determinants can be split in two groups: health care (somatic and mental) and prevention determinants. The prevention determinants can be divided in exogenous determinants (i.e. physical environment, lifestyle, and social environment) and endogenous determinants (both hereditary and acquired).

The model shows that health status is defined by determinants that are affected by health policy. Autonomous developments may affect these three public health aspects and their relations. The health status can be thought of as forming several layers. These layers are disease-specific indicators like disease incidence and prevalence figures, the consequences of diseases and disorders for physical, mental and social functioning, mortality figures, and measures that combine information on mortality and quality of life. The determinants can be split in two groups: health care (somatic and mental) and prevention determinants. The prevention determinants can be divided in exogenous determinants (i.e. physical environment, lifestyle, and social environment) and endogenous determinants (both hereditary and acquired).

The mathematical model structure chosen to describe these demographic and epidemiologic processes simultaneously is the system-dynamic multistate model based on the life-table method. The general model structure is presented in figure 2.

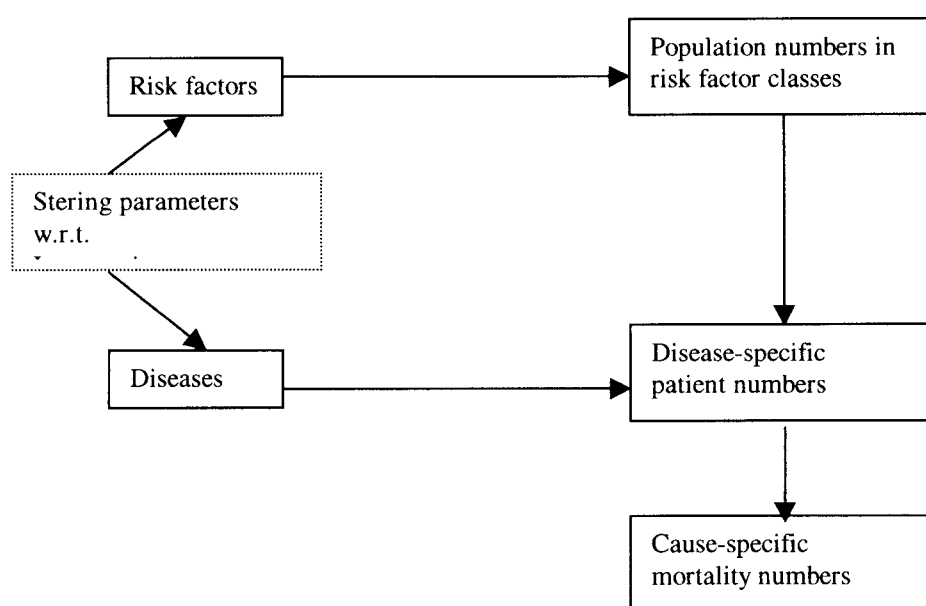


Figure 2: the general chronic diseases model structure

A demographic model describes the annual changes of the population numbers specified by gender, age, and risk factor classes due to birth, mortality, aging and changes of risk factor levels within individuals. For several chronic diseases modules can be developed that describe the annual changes of the prevalence numbers due to disease incidence, progression, remission, and mortality. The disease modules are linked to the demographic module.

Apart from demographic input data the main input data are: disease incidence and prevalence and cause-specific mortality figures, risk factor prevalence figures and risk factor and disease-specific relative risks. All figures are specified by gender and age, whenever possible.

The contents of the report are the following. In chapter 2 the chronic diseases model structure is presented in more detail. The model structure can be divided in relatively separate modules, i.e. a demographic module, and modules for each risk factor and disease category separately. The main model characteristics are worked out. The main model parameters and steering parameters are summarized. In chapter 3 some model analyses and exercises are presented. Some general methods are described that can be used to validate the model. As an example the results are presented of sensitivity analyses for some assumptions that are introduced to simplify the model structure. The chapter is concluded with three example applications. One on comparing cause-specific mortality time series extrapolations to model predictions. Two others on the public health effects of different policies to reduce tobacco smoking or promote physical activity respectively. The main text is ended with some general comments and conclusions.

In several appendices attention is paid to specific model aspects. In Appendix A the assumptions are shown that are needed to derive the main mathematical model equations. In Appendix B the way is described the aspect of competing death risks is taken account of. Appendix C presents the central role of relative risks in the model structure. For example, mortality selection and comorbidity as described by the model structure are directly related to these relative risks. In Appendix D the main model equation on disease prevalence is presented. It can be used to estimate cause-specific case fatality rates. The report is concluded with a list of definitions, abbreviations and literature used.

2. THE CHRONIC DISEASES MODEL STRUCTURE

2.1. The general structure

The model structure chosen to describe the relations between several risk factors and diseases simultaneously is a system-dynamic multistate model based on the life table method, also known as the mover-stayer / increment-decrement model. The main modelling characteristics are:

Process model

The model is based on the concept of demographic and epidemiologic processes. The demographic processes being described are aging, birth, migration and mortality. The epidemiologic processes being described are changing risk factors over time and age, disease incidence (partly depending on the risk factors), disease progression and mortality.

Intended use for prediction and evaluation of intervention measures

The intended use of the model is to describe the future trends in morbidity and mortality related to autonomous trends in the key risk factors, and to assess the public health effects of interventions on these risk factors. In describing these trends and effects the model takes account of demographic and epidemiologic aspects like aging, mortality selection, competing death risks and comorbidity.

Model maintenance

The model structure is simple and flexible to maintain and to work out for specific applications. The processes related to specific risk factors and diseases are described by specific subsets of model equations (modules) that can be handled separately. Introducing or omitting specific risk factors or diseases does not mean that all model parameters have to be estimated again.

Internal consistency

The model structure satisfies several internal consistency conditions. For example consistent dependency relation between the risk factors and diseases, no 'double counts' with respect to mortality, consistent consequences when changing parameter values.

Parameter identifiability

The model parameters have been defined in terms of demographic and epidemiologic figures for which data are principally available. The initial population numbers are available from national statistics, risk factor prevalence rates from national registrations and regional monitoring projects, and disease incidence and prevalence rates from morbidity registration projects. Disease remission and progression rates are available from literature. Case fatality rates are available from literature or can be estimated (see Appendix D). Relative risks are available from literature or follow-up studies.

The main structure of the chronic diseases model has already been presented in figure 2. It shows the main model states and state transitions. The state numbers are the population numbers in the model states that have been defined according to the risk factor classes and disease states being distinguished. I.e. risk factor class population numbers, and disease prevalence numbers. The state transition numbers are the population numbers that ‘flow’ between the model states during successive years. I.e. subgroups moving between risk factor classes, disease incidence and recovery numbers, and mortality numbers.

The more detailed chronic diseases model structure is presented in figure 3. Several aspects of the model structure are described in more detail in the following paragraphs.

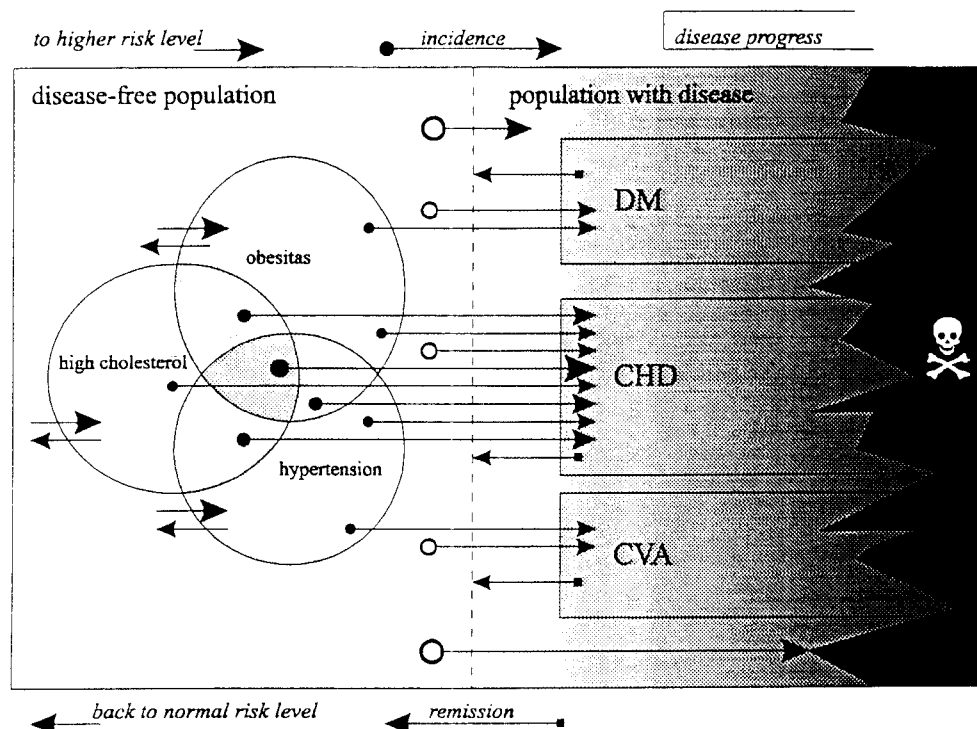


Figure 3: the detailed chronic diseases model structure

2.2. The population and risk factor modules

The core of the chronic diseases model structure is a demographic module that describes the annual changes of the population numbers specified by risk factor classes due to birth, mortality, migration, aging and changes of risk factor levels within individuals ('risk behaviour'). The population is initially structured using demographic data from Statistics Netherlands (CBS). The risk factor prevalence rates are estimated based on Dutch registration and monitoring projects, specified by gender and age whenever possible. For every one-year time step the following changes of the population numbers are calculated:

- The numbers of newborns following CBS forecasts.
- Increasing population numbers due to the net migration rate.
- Decreasing population numbers due to mortality, specified by gender, age and cause of death. The disease modules describe the cause-specific mortality, the mortality due to all other causes is generated externally.
- Transitions between the risk factor classes using one-year transition rates. The risk factor class transition rates have been estimated from longitudinal studies or time series of cross-sectional studies.
- New subpopulation numbers specified by age due to aging.

2.3. The disease modules describing morbidity and mortality

The demographic module can be linked to specific disease modules. Each disease module describes the morbidity and mortality numbers over time for one specific disease category:

- Disease incidence using risk factor prevalence numbers and disease-specific relative incidence risks.
- Cause-specific mortality using case fatality rates.
- Depending on the specific disease and the number of disease states being distinguished disease remission and progression are described. The transition numbers are calculated using one-year remission rates and disease progression rates.

Patients have a risk of dying due to the disease. They can also die of all other causes, whether causes described by any other disease module or truly remaining causes. The 'healthy' persons, i.e. those who are not in any disease module, can only die of the remaining causes. Note: the disease modules describe the excess mortality for those having the disease compared to those without the disease. Only for some highly fatal disease categories such as lung cancer the disease is the one primary cause of this excess mortality (see also Appendix D).

2.4. Main model characteristics

System-dynamic (multistate) model

The chronic diseases model structure is based on system-dynamic and multistate concepts. The model describes parallel population cohorts over time. State variables are introduced with respect to:

- risk factor classes, for example with respect to gender, smoking, bloodpressure, cholesterol level etc.,
- cause-specific disease states, for example with/without CHD, CVA etc.

We have assumed a discrete number of risk factor classes and disease states. This assumption is not restrictive because the number of classes is unrestricted. However, in the model implementation most risk factors and diseases are assumed dichotomous or trichotomous (normal/high risk, or normal/medium/high risk, or without/with disease, without disease/in acute/chronic disease state respectively). Model state transitions are described, due to:

- 'risk behaviour', i.e. transition between two risk factor classes,
- cause-specific disease incidence, i.e. from 'healthy' (=without specific disease) to 'with disease',
- disease progression, i.e. between disease states,
- remission, i.e. from 'with disease' back to 'healthy',
- cause-specific mortality, i.e. from alive (maybe also 'with disease') to the absorbing state 'dead'.

The state transitions are described using transition rates. Disease incidence and mortality rates are available from national registrations or sentinel studies, risk group transition rates are estimated from or longitudinal studies or time series of cross-sectional studies.

Deterministic/stochastic model

The underlying mathematical model is stochastic, i.e. the model equations describe the change over time of the probability distribution function over the model states for any individual, conditional on the initial distribution. However, the model is implemented deterministically, i.e. the model describes the change of the mean population numbers in the model states. Setting the migration to zero the ratio of the population numbers in the model states to the initial population numbers equals the probability distribution function for any initial person. This equivalence is valid because all model equations are linear with respect to the risk factor class and disease prevalence numbers (see Appendix A). In comparison, this equivalence is not valid for infectious disease models that are based on the Anderson-May model structure, because the number of newly infected persons is assumed to be proportional to both the numbers of susceptible and infected persons.

Markov-property

The model satisfies the Markov-property: conditional on the current model state (i.e. conditional on gender, age, risk factor classes and disease states) the probability distribution of the model states one time-step ahead is independent on the past model states. One consequence of this assumption is that the residence time in a model state has no effect on the (outflow) state transition probabilities. For example, conditional on gender, age etc, the survival probability of any stroke-patient is assumed independent on his past disease period.

Assumption of conditional independence

The model state variables can be distinguished into two disjoint subgroups, those to be interpreted as 'risk factors' and those to be interpreted as 'disease states'. According to the (in the context of public health modelling common) conditional independence assumption, the disease transition rates are assumed mutually independent, conditional on the risk factors. For example the survival probability of any CHD-patient does not depend on its disease state with respect to COPD, conditional on gender, age, and smoking. As a result of this conditional independence assumption the joint probability function for all disease prevalences is equal to the product of the marginal probability functions, conditional on the risk factors. The conditional independence assumption is empirically valid for most combinations of specific risk factors and diseases (for counter-example, see below).

Comorbidity

Comorbidity is defined as the prevalence rate for one disease being dependent on the prevalence rate for another disease. Due to the assumption of conditional independence the CDM model structure describes comorbidity resulting from common risk factors, i.e. age and the epidemiologic risk factors being distinguished. There may be dependency relations between diseases that cannot be explained by the common risk factors. For example, DM is a risk factor for CHD independent from gender, age, and Body Mass Index. This dependency relation can only be modeled by omitting the conditional independence assumption resulting in a more complex model structure.

Competing death risks

Competing death risks are described explicitly by distinguishing different causes of death. Individuals may die from any cause during their life. Due to the assumption of conditional independence the cause-specific mortality rates are independent, conditional on age and the risk factors. Because any individual may die from any cause during his life, being saved from any specific cause of death due to a specific intervention program means he will die from the same cause or any other cause on a higher age. Persons being saved are assumed to be identical to those already having survived, conditional on the risk factors (see also Wong, 1977).

Other assumptions with respect to morbidity and mortality

We assume that for any person his cause of death is unique. The cause-specific mortality is assumed to be proportional to the related disease incidence and prevalence. Because of lack of good data and to simplify the mathematical model equations it is also assumed that the disease progression and cause-specific case fatality are independent on the risk factors. The last assumption is not essential. Analogous model equations can be derived assuming that the disease progression and mortality are risk factor-dependent.

Costs and effects

Costs and effects are the concepts to characterize and compare different intervention programs. More information on costs and effects in public health and health care research, on relating them in terms of cost-effect measures, and on using these concepts in intervention program evaluations and scenario studies can be found in the literature (for example Drummond et al., 1987; Rutten et al., 1993). In this paragraph we only mention the way costs and effects can be related to some key model variables.

Several types of costs can be distinguished. We present two examples of possible cost calculations in relation to the chronic diseases model, one for a risk factor prevention program, the other for a health care intervention program. Of course the costs being distinguished and cost calculations being performed are the consequences of the perspective chosen for the evaluation. The direct costs of implementing both programs are generally fixed and variable. The variable costs of the prevention program are related to changes in the risk factor class population numbers. The direct costs of the intervention program are related to the changes in the disease progression, remission and/or case fatality numbers. However, due to competing death risks also indirect costs of the intervention program can be defined. These costs are related to the changes in the future disease incidence and prevalence numbers, both for the specific disease and other diseases.

The 'simple' effect measures are the model output morbidity and mortality numbers. Using the life table method the output mortality rates can be used to calculate life expectancy figures. When combining these mortality rates with output disease prevalence rates and DALY-coefficients it is also possible to calculate healthy life expectancy and disabled life expectancy figures, aggregated or specified by disease (see Appendix E).

2.5. Risk factors and diseases being distinguished

The actual chronic diseases model version (4) describes the following causes of morbidity and mortality together with the related risk factors and diseases.

Table 1: risk factors and diseases in the chronic diseases model

	smoking	bloodpressure	cholesterol	BMI
COPD	x			
CHD	x	x	x	x
stroke	x	x		
diabetes				x
lung cancer	x			

Apart from these classical epidemiologic risk factors the risk factor social-economic status (SES) based on educational level is introduced in the chronic diseases model structure. A hierarchical relationship is assumed between these risk factors: the three classical risk factors are assumed independent, conditional on SES. This assumption is worked out in the following way: the initial prevalence rates and the transition rates depend on the SES-status. The population distribution over the SES-states being distinguished is externally given. More information over the model structure related to SES can be found in (Hoogenveen et al., 1998b).

2.6. Main model parameters

The main model parameters are the initial population numbers in the model states, i.e. specified by gender, age, risk factor classes and disease states, and the transition rates between these states. The transition rates describe the relative changes over time of the population numbers in the model states due to aging, change of risk factor class within individuals, disease incidence, progression and remission, and mortality. A list of the main model parameters together with the main data sources is presented below. All parameters are specified by gender and age. All parameters are assumed time-constant (see also §4.1).

initial numbers and rates

initial population numbers

data from Statistics Netherlands (CBS)

initial risk factor prevalence rates

smoking (never/smoker/former): data from Dutch Smoking Foundation (1986-1995)
serum cholesterol level (normal/high): Morgen (Blokstra, 1997), ERGO (Mennen, 1995)
bloodpressure level (normal/high): Morgen, ERGO
Body Mass Index (normal/high): Morgen, ERGO

initial disease prevalence rates (note 1)

coronary heart diseases (CHD) (myocard infarct (MI), angina pectoris (AP))
chronic obstructive pulmonary diseases (COPD) (asthma (AST), chronic bronchitis (CB))
stroke (CVA)
lung cancer
diabetes mellitus

transition rates

with respect to risk factors:

transition rates between risk factor classes within individuals (note 2)

with respect to diseases:

disease incidence rates (note 1)
disease progression rates (if two or more disease states are distinguished)
disease remission rates (if existent)
case fatality rates (note 3)

relative risks

for all relevant combinations of risk factors and disease types

main steering parameters (note 4)

transition rates between risk factor classes within individuals
disease progression and case fatality rates

Notes:**1 Initial disease prevalence and incidence rates**

These rates are based on registrations in Dutch general practice (Gijssen et al., 1997; Hart et al., 1996; RIVM, 1994; RIVM, 1997).

2 Transition rates between risk factor classes

The transition rates between risk factor classes are ideally estimated from data from longitudinal studies. However, these data are rarely available. The transition rates can be estimated approximatively from series of cross-sectional studies (Hoogenveen et al., 1998a). The estimated transition rates have to be interpreted carefully, especially in case of absolutely large transition rates and/or relatively large changes over time.

3 Case fatality rates

It is hard to find empirical data on the cause-specific case fatality rates for two reasons. (1) The rates describe the excess hazard for persons with a specific diseases relative to those without the disease. For highly fatal diseases such as lung cancer the excess hazard is almost equal to the registered mortality with the disease as the one primary cause of death. For other less fatal diseases the registered primary cause of death mortality is smaller. (2) Patients may be more susceptible to mortality from other causes than persons without the disease (Zahl, 1997). However, under specific assumptions the case fatality rates can be estimated from data on disease incidence, prevalence, and, if existent, remission (see Appendix D).

4 Main steering parameters

The steering parameters are the model parameters that represent the ways the public health status can be influenced by policy measures. Two types of intervention measures can be distinguished: those aimed at reducing the public health risks (primary prevention), and those aimed at reducing the disease progression or, if possible, at increasing the disease remission (secondary prevention). The model parameters related to these intervention masures are the transition rates between the risk factor classes and the disease progression, case fatality and remission rates respectively.

3. SOME MODEL ANALYSES AND EXERCISES

3.1. Model validation

Model validation is defined as proving the reliability, accuracy and functionality of the model within the specified domain of model applications (Giessen van der, 1997). Following this definition several aspects of model validation can be distinguished:

- conceptual validation, i.e. with respect to the model assumptions and structure,
- functional validation, i.e. with respect to the functionality of the model and the model results,
- data validation, i.e. with respect to the data.

These three, especially the first two aspects of model validation may be overlapping. Many validation methods have been described that deal with these three aspects of model validation (Hoogenveen, Brouwer, 1989). We present some methods that are relevant and useful to chronic diseases modelling.

Face validity (expert judgement)

Expert judgement is given on the model assumptions, the selection of risk factors and disease types, the data sources being used, and the future trend assumptions on the model parameters.

Sensitivity analysis

This means analyzing the sensitivity of the model output variables to changes in the model parameters. Sensitivity analysis gives insight in the relevance of specific parameters and in the uncertainty of the model results.

Historical validation

Model results are compared to historical data. Historical validation is related to the statistical concept of cross-validation. The data available are divided in two parts: those to identify the model parameters, and those to validate the model. Historical validation is often difficult in the field of public health modelling, because detailed data may be hard to find and are used for parameter identification only. However, some simple forms of historical validation may be applicable, for example comparing the calculated future trends in the risk factor and disease prevalence rates to past trends.

Comparison with other models

The model results are compared to those of other models on the same public health aspect. Different model results may be caused by different model assumptions/structures and/or different data being used.

3.2. Some general model validation results

The chronic diseases model structure is complex: several risk factors and diseases are combined within one model structure. Following the methodology of multistate modelling, model states have to be defined for all risk factors and diseases simultaneously. Due to the assumption of conditional independence the model structure can be simplified substantially: only model states have to be defined for each disease separately together with the risk factors for that specific disease. Still the model structure is complex and more model simplifications are needed. Several model exercises have been performed to analyse the sensitivity of the model results to extra assumptions simplifying the model structure.

Model results with respect to diseases

- Assuming a multiplicative or additive risk model, or omitting any number of risk factors or other diseases has no effect on the disease prevalence rates on the population level. The reason is that the risk factor class-specific disease incidence rates are defined conditional on the mean rates.
- Not taking into account the higher mortality risks of persons having one disease for causes of death with common risk factors (mortality effect of comorbidity) has almost no effect on the disease prevalence rates on the population level. This allows us to simplify the complex mortality calculation step within the disease modules and to apply the population cause-specific mortality rates instead of the risk factor class specific ones (see also Appendix A).
- Changing the multiplicative to the additive risk model results in significantly smaller comorbidity coefficients.

Notes:

- Univariate relative risks are mostly estimated in the context of (i.e. corrected for) other confounding risk factors. The method of correction defines which risk model can be applied to combine the univariate relative risks to (approximately) estimate the risk of combined exposure. Mostly a multiplicative risk model is assumed. Estimating univariate relative risks assuming an additive risk model will generally result in higher univariate risk levels. Therefore the results on comorbidity described above overestimate the differences between the model results for both risk models.

Model results with respect to risk factors

- When changing the multiplicative to the additive risk model the risk factor prevalence rates increase for higher ages. The explanation is simple: the mortality rates for persons with combined risks decrease, and so the marginal prevalence rates increase.
- Assuming a multiplicative risk model omitting one or more diseases may result in increasing prevalence rates for the high risk classes for higher ages. The explanation is again simple: the mortality rates for persons with high risk levels decrease because the mortality due to all remaining causes is assumed to be risk factor-independent.

3.3. First example application: comparing cause-specific mortality time series extrapolations to model predictions

The first example application of the chronic diseases model concerns comparing the model output cause-specific mortality rates to estimations based on time series extrapolations (see also VTV, 1997, theme 7, pp 128-135). The method of dynamic parametrization (Tabeau, 1995) has been used to estimate the functional relationship of cause-specific mortality rates to time and age. Substituting future time years results in future mortality rate estimations. The chronic diseases model has been used to estimate future cause-specific mortality rates based on several assumptions on the future risk factor values (scenarios). The causes of death being distinguished are coronary heart disease (CHD) and lung cancer. The risk factors being distinguished are smoking, hypertension and hypercholesterolemia. In the 'reference smoking scenario' the smoking transition rates are assumed to be constant over time. In a 'feasible anti-smoking scenario' it is assumed that the start-smoking rates decrease 20% in a period of three years and then remain constant. Combined with increasing stop-smoking rates, this scenario results in a 14% decrease of smoking prevalence rates in the first year, followed by a 1% annual decrease. In an 'extreme anti-smoking scenario' it is assumed that the start-smoking rates decrease and the stop-smoking rates increase, both with a factor 5. In alternative scenarios for the other risk factors it is assumed that the hypertension prevalence rate decreases from 5% to 2% in 2015 for males and from 3.5 to 1.4% for females, and for hypercholesterolemia from 10% to 4% for males and from 8% to 4% for females. In a 'CHD scenario' it is assumed that better medical treatment results in further decreasing case fatality rates of both CHD incidence and prevalence.

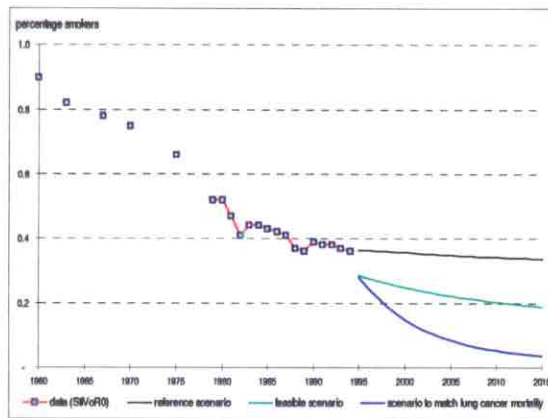


Figure 4a: smoking prevalence data and model predictions

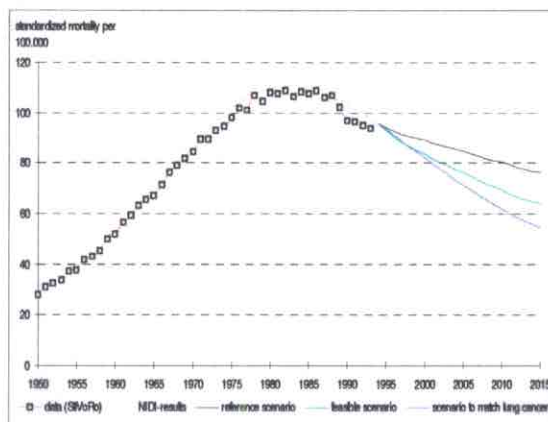


Figure 4b: lung cancer mortality data and predictions

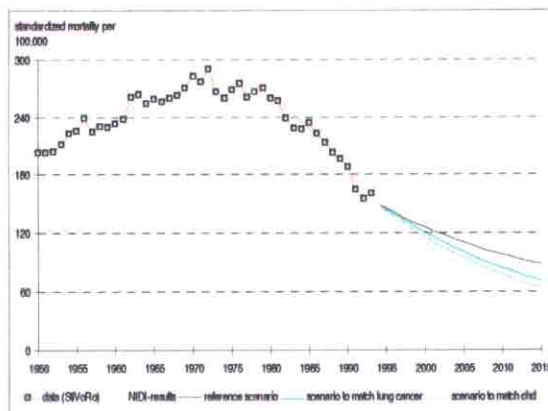


Figure 4c: CHD mortality data and predictions

Figure 4: comparing cause-specific mortality time series extrapolation to model predictions (males)

For lung cancer the reference scenario shows a much smaller decrease of mortality for males than the time series extrapolation (see figure 4b). Only the extreme 'anti-smoking scenario' results in similar results. For CHD the reference scenario shows a much smaller decrease of mortality than the time series extrapolation (see figure 4c). Similar results can only be found when combining the 'extreme anti-smoking scenario', the alternative scenarios for hypertension and hypercholesterolemia, and the CHD scenario on increased survival.

Two important comments can be given on the differences between the time trend extrapolations and the model results for the scenarios being distinguished. (1) The trend extrapolations are mostly influenced by the decrease of smoking prevalence rates over the past decades, whereas the model results are mostly influenced by the change over the past five years (see figure 4a). (2) For CHD not only trends in risk factors are important to describe changing mortality rates, but also trends in prognosis. Both trends can be distinguished better by the simulation model, than by time series extrapolation.

3.4. Second model application: the effectiveness of different policies to reduce tobacco smoking in terms of future smoking prevalence rates and burden of disease

As in most developed countries the smoking attributable burden of disease is substantial in the Netherlands. For the Dutch government it is top priority to discourage tobacco smoking to prevent smoking related morbidity. An assessment is made of the public health gain of different policy measures to reduce smoking prevalence: campaigns aimed at keeping young people from starting; campaigns aimed at smoking cessation; and price/tax measures.

The behavioral change related to smoking is modelled using starting and quitting transition rates. The transition rates are fitted on the 1987-1995 time series of smoking prevalence data (see figures 5c&5d). The starting rates are highest for 15-19 years old, then they are declining to zero from age 35 years on. For males the decline is steeper than for females, which means that girls generally start smoking at higher age. From Dutch registry data on morbidity (Gijzen et al., 1997) the number of patients with lung cancer, CHD, stroke and COPD are known. Disease incidence rates for smokers are higher than for former-smokers and never-smokers. Smoking related rate ratios are drawn from international literature (Shopland et al., 1991).

Different scenarios on smoking intervention measures are described:

- A Health promotion aimed at keeping young people from starting. A 20% decrease of the starting rates can be reached in three years' time and can also be maintained.
- B Health promotion aimed at smoking cessation. Stopping rates are increased such that the smoking prevalence rates decrease with 14% in the first year and continues to decrease with 1% annually.
- C CA price increment of 50% results in decreasing starting rates and increasing stopping rates using price elasticity figures (-1.2% for starters, .08% for male stoppers and .23% for female stoppers). The effects diminish annually with 3%.

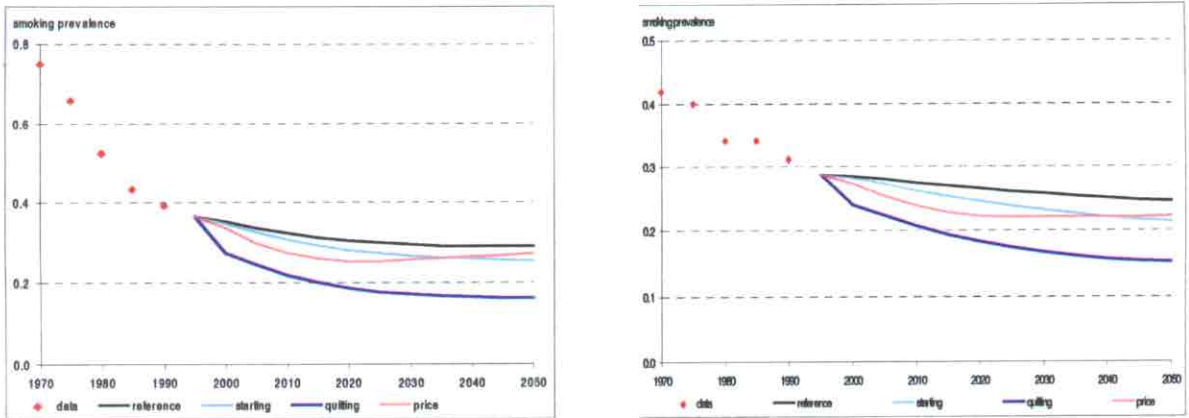


Figure 5a: smoking prevalence data and predicted rates for males and females

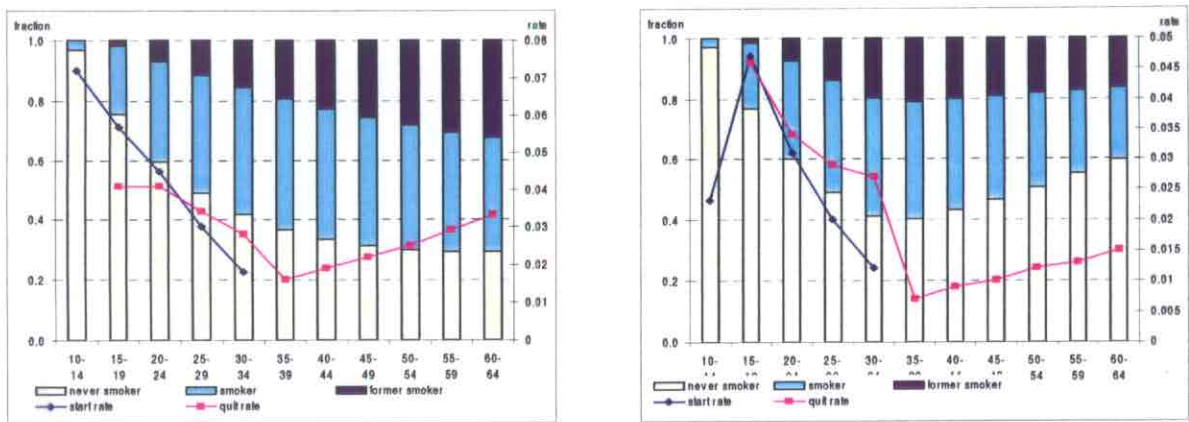


Figure 5b: smoking class prevalence rates and transition rates for males and females

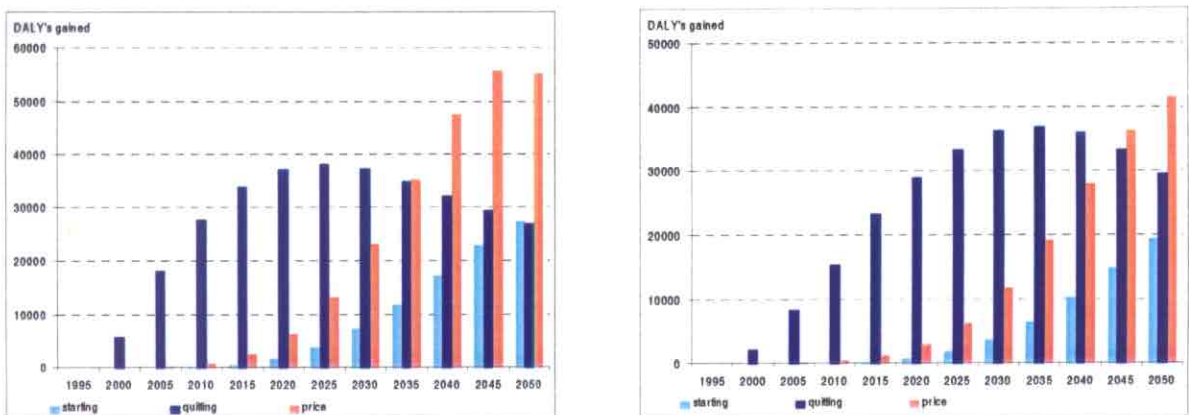


Figure 5c: DALYs gained according to the scenarios for males and females

Figure 5: the effects of different scenarios on smoking behaviour (males and females)

In the reference situation the smoking prevalence rate for males decreases from 36% in 1995 to 30% in 2020 and 29% in 2050, for females it decreases from 29% in 1995 to 27% in 2020 and 25% in 2050 (see figures 5a&5b). The policy scenarios A and B result in slightly and significantly respectively smaller smoking prevalence rates. The price scenario C result in reductions that are initially larger than those in scenario B but in the end are smaller. The number of Disability Adjusted Life Years (DALY's) gained are initially largest for policy scenario B, but approximately after 50 years they are largest for scenario C (see figures 5e&5f). For scenario A the DALY-effects are smallest.

3.5. Third model application: the effectiveness of different policies on increasing physical activity

Physical activity has positive effects on the prevention and treatment of several important chronic diseases such as CHD, diabetes and osteoporosis. The normative level that distinguishes active from inactive persons has been defined by doing 'moderate-intensity activities during at least 30 minutes every day'. In 1994 the (CHD, diabetes and colon cancer) mortality fraction attributable to physical inactivity in the Netherlands is approx. 8000 persons. The public health effects have been estimated of several scenarios on physical activity programs aimed at specific population subgroups and/or activity levels. The physical activity classes being distinguished are: sedentary, light, moderate and vigorous activity (see figure 6a). These scenarios are:

- A Promoting physical activity at younger ages
- B Promoting physical activity in the total population such that every person moves to next-favourable class
- C Promoting physical activity in the total population such that sedentary and light active persons move to class 'moderately active'
- D Promoting physical activity at higher ages

As a first example the effects of the four scenarios are calculated in terms of DALY's gained compared to the reference scenario. It turns out the effects of the scenarios B and C do not differ much.

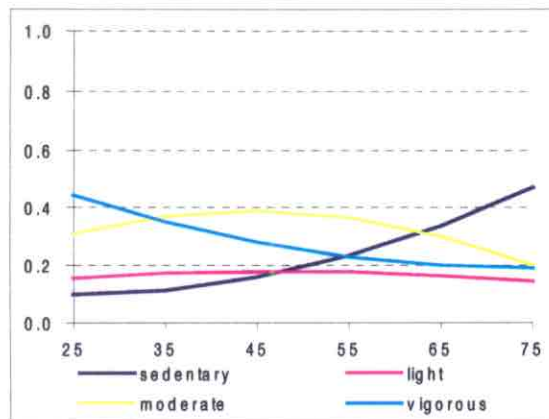


Figure 6a: physical activity class prevalence rates over age (males)

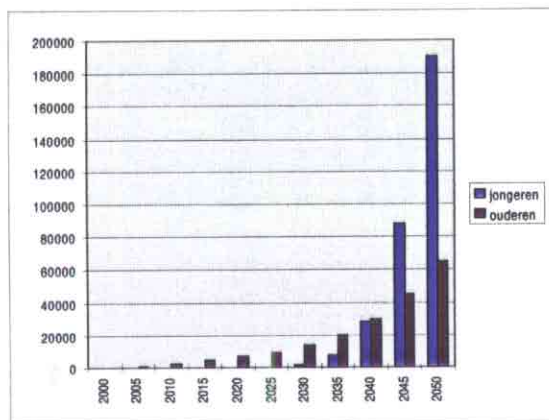


Figure 6b: quality-adjusted life years gained for CHD and diabetes mellitus (males)

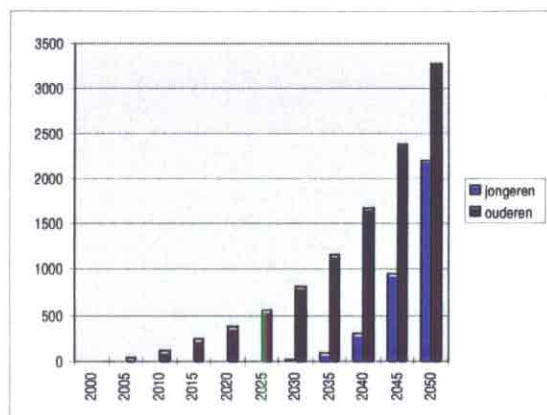


Figure 6c: mortality numbers gained for CHD and diabetes mellitus (males)

Figure 6: the effects of different scenarios on physical activity (males)

4. GENERAL COMMENTS AND CONCLUSIONS

4.1. Comments on the model characteristics

System-dynamic model structure

The system-dynamic multistate structure underlying the chronic diseases model is highly applicable to the intended use of the model, i.e. future trend extrapolations and effect estimations of risk factor intervention programs, taking explicitly into account some main demographic and epidemiologic processes (high model 'face validity'). It is also flexible with respect to changing submodules related to specific risk factors or diseases, or introducing other risk new factors or diseases.

Internal (in)consistencies

The model structure is highly internally consistent. For example changing or omitting specific risk factors or disease categories has almost no impact on disease related model outcome variables on the population level. However, omitting disease categories does have effect on the risk factor prevalence rates for higher ages. The reason is that it is assumed that the mortality due to all remaining causes is independent on the risk factors. So omitting a disease results in decreasing risk factor-specific mortality rates and consequently increasing prevalence rates, especially for the higher ages.

The choice of the model of combining relative risk values to describe the risk of 'combined exposure', i.e. additive or multiplicative risk model, has no effect on the disease prevalence rates. However, it does have effect on the high risk prevalence rates for higher ages. The reason is that assuming an additive instead of multiplicative risk model will result in decreasing risk levels for 'combined exposure'. Therefore also the marginal risk factor prevalence rates increase. This effect is overestimated because assuming an additive instead of multiplicative risk model also results in higher univariate relative risk values.

Assuming the cause-specific case fatality and mortality due to all remaining causes independent on the risk factors

These assumptions have been made for several reasons. (1) They are common in the context of public health modelling (Barendregt, Bonneux, 1998). (2) There are hardly data to reject these assumptions, especially the first one. (3) They simplify the model equations. However, they are not essential to the model structure meaning that releasing them does not violate the assumption of conditional independence. These assumptions can be dealt with following the way comorbidity is described within the model structure (see Appendix C).

The assumption that the mortality to all remaining causes is independent on the risk factors results in an internal model inconsistency. That is, omitting an important cause of death that is strongly dependent on a risk factor such as CHD with respect to smoking, results in smaller mortality risks for the high risk class and consequently in increasing prevalence rates (see also §3.2). Therefore, to maintain the internal consistency of the model structure, the mortality to all other causes have to be made dependent on the risk factors taking into account the causes of death being distinguished.

Comorbidity

The chronic diseases model structure describes comorbidity due to common risk factors, i.e. age and the epidemiologic risk factors being distinguished. It turns out that the effects of comorbidity on the disease prevalence rates on the population level are marginal. This allows us to simplify the complex mortality calculation step within the disease modules and to apply the population cause-specific mortality rates instead of the risk factor class specific ones.

Data availability

The main model parameters can be split in four groups. (1) Initial population numbers, and risk factor and disease prevalence rates. (2) Disease-specific incidence, recovery and case fatality rates. (3) Transition rates between risk factor classes. (4) Risk factor- and cause-specific relative risks.

Data on the group (1) parameters are principally available, although the data quality on disease prevalence rates and risk factor prevalence rates for higher ages may be poor. (2) Disease incidence rates are generally available, recovery rates may be hard to find, and cause-specific case fatality rates are hard to estimate. However, the cause-specific case fatality rates can be estimated using disease incidence and prevalence data (see Appendix D). (3) These data are hard to find because there have been no representative longitudinal studies on risk factors for sufficient long time periods in the Netherlands so far. However, these transition rates may be estimated from time-series of cross-sectional studies (Hoogenveen et al., 1998a) or by assuming time-constant risk factor prevalence rates. (4) Relative risks are generally available for most combinations of specific important risk factors and disease categories. However, relative risks for all remaining causes of death and relative risks for higher ages still need to be assessed.

Time trends

Almost all model parameters are assumed to be time-constant. One exception is the CHD case fatality rate, that is assumed to decrease over time. This does not mean that the output prevalence rates are time-constant due to the system-dynamic model structure. The risk factor prevalence rates change over time because the risk factor transition rates may have changed in the past. When the disease prevalence rates change also the disease prevalence and mortality rates will change. Recognizing trends apart from the system-dynamic effects needs more attention.

Time horizon

The time horizon of the chronic diseases model structure is a complex issue. It depends on the use of the model, i.e. whether for estimating public health trends or for estimating the effects of public health intervention programs. In the case of trend estimations the main model results are absolute disease prevalence and mortality figures. The time horizon may be 15 years at most and according to general statistical extrapolation rules it has to be even smaller. In the case of effect estimations the results are changes in these figures due to the intervention program and compared to the reference situation. Due to time delays the effects of risk factor intervention programs may only become apparent after long time periods such as 40 years. So in order to describe the full effects of intervention programs the time horizon has to be chosen sufficiently large, often more than 15 years.

Aspects of time delay

Several aspects of time delay can be distinguished.

- 1 Between entering a high risk group and increasing risk level, for example between starting to smoke and increased risk of lung cancer.
- 2 Between leaving a high risk group and decreasing risk level, for example between cessation of smoking and decreased risk of lung cancer.
- 3 Between disease incidence and mortality.

The system-dynamic structure of the chronic diseases model results automatically in time delays with respect to risk factor and disease prevalence. The risk factor and disease prevalence numbers can be interpreted as volume variables that only respond with delay times to changes in the inflow and outflow rates.

The standard methodology of implementing time delays in Markov-type models is introducing extra model states for all duration times being distinguished. For example, in Prevent (Gunning-Schepers, 1988) the former smokers have been divided according the time since stopping. This way of implementing time delays has several disadvantages: many extra model states and model parameters are introduced. However, some alternative implementation forms are available, for example model state parameters that depend on the mean time duration. The mean duration increases due to inflow of new persons and decreases over time.

Time delays are important on the short run. For long time periods when all persons have reached the same maximum time duration, the time delay is not important any more.

Relative risks with respect to mortality or incidence

The relative risk parameters of the chronic diseases model structure refer to the outcome 'disease incidence', whereas most relative risk data refer to the outcome mortality. There is not much known about the relation between these relative risks, but some results have been found (see Appendix C.2). These results are based on analyses on data from the Zutphen-study (Keys et al., 1967), simulation results of the chronic diseases model, and on analyses on data from the Framingham-study (Kannel et al., 1987) presented in literature. We conclude that there is some evidence that relative incidence risks are smaller than relative mortality risks for several combinations of risk factors and disease types, but that age-effects cross this relation.

4.2. Future model applications and development

Several aspects of future applications can be mentioned (see also RIVM, 1997).

Availability of data from longitudinal studies

The chronic diseases model structure is process-oriented: it describes the change of the public health state over time as a function of demographic and epidemiologic processes. Data from longitudinal studies are ideally used to estimate the model parameters and to validate the model. However, in practice only data from time series of cross-sectional studies are available. An increase of the number of Dutch longitudinal and follow-up studies encourages the further development and refinement of system-dynamic simulation models on population risk factors and chronic diseases.

Statistical aspects of chronic disease modelling

Generally little attention is paid to the interpretation of chronic disease models and model results in mathematical-statistical terms. A longer tradition on this model interpretation exists in the field of trend extrapolations based on data time series, for example in demographics (Willekens, 1997). However, the lack of attention in chronic disease modelling is not justified: they deal with stochastic events, the model parameters have to be estimated, and the models produce predictions. Interpreting the model in mathematical-statistical terms may be useful to select and estimate model parameters, and to make quantitative judgements on the reliability of the model results.

Costs and effects of intervention programs

The chronic diseases model structure has been developed to describe the effects of intervention measures on disease prevalence and mortality taking into account several risk factors and disease types simultaneously. The intervention measures are described as changing model parameters, i.e. changing transition rates between risk factor classes and disease progression, case fatality or, if existent, remission rates. However, this effect estimation is only part of the evaluation of intervention programs. Other important issues that have to be dealt with are: relating specific 'real world intervention measures' to the changes in the model parameters, quantifying the costs of the intervention programs and unifying the effects to make different intervention programs comparable. One well-known unifying concept is the costs per (quality adjusted) life year gained. This concept has already extensively been used in the field of health-economics (Drummond et al., 1987).

Model applications

The chronic diseases model structure based on the concepts of states and state transitions (flows) is well developed to calculate the public health effects over time and age of intervention programs taking into account aspects like postponement of and competition between events (morbidity and mortality risks) and clustering and interaction of risk factors and diseases. Examples of applications are estimating the future rise of heart failure, stroke trends in an aging population, and the health care costs of smoking (Barendregt, Bonneux, 1998), effects on osteoporosis and CHD prevalence of oestrogen suppletion to postmenopausal women, and the public health effects of anti-smoking campaigns (RIVM IV, 1997). The consequence of using this type of state-transition models is the need of data on the transitions. These data have to come basically from longitudinal studies.

A general issue concerning model development and applications is model complexity. One general rule of parsimony is known that has been stated in several ways, for example: a model should be as simple as possible. What is simple and what is possible depends on the research question and the data available (see also Barendregt, Bonneux, 1998). Sensitivity analyses can be helpful in developing models.

APPENDIX A. THE MAIN MODEL EQUATIONS

A.1. The basic model variables

The basic model variables describe the state of any individual, i.e. with respect to its survival, the risk factors and the disease states. They are stochastic indicator variables. Note: throughout the appendix we present stochastic variables in bold capitals, realisation values are presented in regular capitals.

I(t) stochastic survival indicator variable

D(t) stochastic disease state vector variable, for example with(1)/without(0) CHD, stroke etc.

Z(t) stochastic risk factor state vector variable, for example never-smoker/smoker/former-smoker, with(1)/without(0) hypertension etc.

Note: **D(t)** and **Z(t)** are only defined for persons being alive (**I(t)=1**).

A.2. The basic model equations

The state probability distribution function P is defined implicitly, i.e. conditionally dependent on its latest value. For the initial time point the probability distribution function is given explicitly.

model state probability on time $t+\Delta t$ $P(\mathbf{I}(t+\Delta t)=1, \mathbf{D}(t+\Delta t)=\mathbf{D}, \mathbf{Z}(t+\Delta t)=\mathbf{Z}; t+\Delta t) \equiv$

$$P(\mathbf{D}, \mathbf{Z}; t+\Delta t) \approx$$

model state probability on time t $P(\mathbf{D}, \mathbf{Z}; t)$

inflow from other disease states $+ \sum_{n(\mathbf{D})} f(\mathbf{D} \leftarrow n(\mathbf{D}); \mathbf{Z}; t) P(n(\mathbf{D}), \mathbf{Z}; t) \Delta t$

inflow from other risk factor states $+ \sum_{n(\mathbf{Z})} f(\mathbf{Z} \leftarrow n(\mathbf{Z}); \mathbf{D}; t) P(\mathbf{D}, n(\mathbf{Z}); t) \Delta t$

outflow to other disease states $- \sum_{n(\mathbf{D})} f(n(\mathbf{D}) \leftarrow \mathbf{D}; \mathbf{Z}; t) P(\mathbf{D}, \mathbf{Z}; t) \Delta t$

outflow to other risk factor states $- \sum_{n(\mathbf{Z})} f(n(\mathbf{Z}) \leftarrow \mathbf{Z}; \mathbf{D}; t) P(\mathbf{D}, \mathbf{Z}; t) \Delta t$

mortality $- f(\mathbf{D}, \mathbf{Z}; t) P(\mathbf{D}, \mathbf{Z}; t) \Delta t$

with: \mathbf{D} vector of disease state variables D_i , i.e. with($D_i=1$)/without($D_i=0$) CHD, stroke, etc.

\mathbf{Z} vector of risk factor class variables Z_j , i.e. with($Z_j=1$)/without($Z_j=0$) hypertension, etc.

$P(\mathbf{D}, \mathbf{Z}; t)$ the probability of being in disease state \mathbf{D} and in risk factor state \mathbf{Z} on time t

$D \leftarrow n(D)$ transition to disease state D from one of the neighbour disease states $n(D)$
 f disease and risk factor transition rates and mortality rates respectively

Notes:

- The model is a cohort model. So time may be interpreted as age.
- The model equations describe the change in the probability function for any individual. The probability of being dead is the complement value.
- When the time interval Δt is sufficiently small, not more than one transition may take place. So for any model state the neighbour states are those states that differ for only one state variable.
- Initially $I(t_0)=1$. For the other model state variables the initial conditions are age-dependent.

A.3. The interpretation of the state transition rates

Depending on the specific state transitions the rates f have the following interpretation:

$D_i \rightarrow n(D_i)$ interpretation

0 1 disease i incidence; notation: inc_i

1 0 disease i remission; notation: rem_i

$Z_j \rightarrow n(Z_j)$

0 1 risk factor j increase (from normal to high risk factor class)

1 0 risk factor j decrease (from high to normal risk factor class)

alive \rightarrow dead

$D_i=0$ 'acute' mortality, assumed to be proportional to disease incidence

1 'chronic' mortality, assumed to be proportional to disease prevalence

A.4. Distinguishing causes of death

With respect to mortality we distinguish several causes of death. These causes are the diseases being described explicitly and the remaining causes. So the mortality rate is calculated as (cf. the model characteristics):

$$\text{mort}(D, Z; t) = \sum_i \text{let}_i(Z; t) \text{ind}(D_i(t)=1; D, z; t) + \text{mort}_{\text{ov}}(Z; t)$$

with: $\text{ind}(\cdot)$ indicator function; $\text{ind}(\text{TRUE})=1$, $\text{ind}(\text{FALSE})=0$

let_i the cause-specific case-fatality rate for disease i

mort_{ov} the mortality rate for all remaining causes

Notes:

- In the computer model implementation let_i and $mort_{ov}$ are assumed independent on the risk factors being distinguished in the model.
- It is assumed here that the cause-specific mortality is proportional to the disease prevalence. For some diseases there is also cause-specific mortality related to the disease incidence. Analogous model equations can be derived for this situation.
- Both assumptions are not essential to the model structure meaning that releasing the assumptions does not violate the conditional independence assumption.

A.5. Computer model implementation of basic model equations**With respect to the disease prevalence rate**

- Making use of the specific notations for the state transition rates,
- specifying the cause of death,
- applying the conditional independence assumption,
- assuming the disease progression for any patient, the cause-specific case fatality and the mortality due to all remaining causes being independent on the risk factors,
- assuming the cause-specific mortality to be proportional to the disease prevalence,
- assuming that the transitions between the risk factor classes are independent on the risk factors,

the differential equations for the disease prevalence probability function are described by:

$$\begin{aligned} \Delta PREV_i(t) / \Delta t \approx & \\ \text{incidence} & + \sum_z inc_i(Z;t) \{ POP(Z;t) - PREV_i(Z;t) \} \\ \text{remission} & - rem_i(t) PREV_i(t) \\ \text{case-fatality} & - let_i(t) PREV_i(t) \\ \text{other cause-specific mortality} & - \sum_z \sum_{j \neq i} mort_j(Z;t) PREV_i(Z;t) \\ \text{rest mortality} & - mort_{ov}(t) PREV_i(t) \end{aligned}$$

with: $PREV_i$ the probability of having at least disease i
 inc_i the disease i incidence rate, related to the disease-free population
 POP the survival probability
 $mort_j$ the cause (disease) j -specific mortality rate

Notes:

- $mort_i = let_i prev_i$, with: $prev_i$: the disease i prevalence rate = the probability of having the disease conditional on survival.

With respect to the risk group prevalence rate

- Again using the specific notations for the transition rates,
- specifying the cause of death
- applying the conditional independence assumption,
- assuming the mortality due to all remaining causes being independent on the risk factors,
- assuming that the transitions between the risk factor classes are independent on the risk factors,

the differential equations for the risk group probability function are described by:

$$\Delta POP(Z;t) / \Delta t \approx$$

cause-specific mortality

$$\sum_i mort_i(Z;t) POP(Z;t)$$

rest mortality

$$- mort_{ov}(t) POP(Z;t)$$

risk factor change

$$+ \sum_{n(Z)} f(Z \leftarrow n(Z);t) POP(n(Z);t)$$

A.6. Some remarks**Fixed vs variable cohorts (dynamic populations)**

The model equations describe the stochastic life course of one individual. The chronic disease model structure describes total population numbers. Therefore the model equations described above are extended in the deterministic population model version with:

- numbers of newborns during the simulation period (with age 0) or elder persons alive at the start of the simulation period (with age greater than 0)
- 'parallel' birth-cohorts to describe the total population,
- inflow and outflow of persons due to migration.

Cohorts vs cross-sections

Three 'time axes' can be distinguished: calendar time, age and year of birth (see Lexis-diagram). Only two of them are independent because of the relationship:

age = calender time - year of birth

The chronic diseases model structure is meant to describe the population within the rectangle defined by age and simulation (calender) time period. At each time point the change of the cohort distribution is computed conditional on the old one. Therefore, age increases parallel with time.

Marginal numbers compared to joint numbers

The original mathematical model structure has been defined in terms of the change of a joint probability distribution function over time. 'Joint' means joint with respect to all risk factors and diseases (causes of death) being distinguished. According to the conditional independence assumption the joint probability distribution function for all diseases is uniquely defined by the marginal disease-specific functions (prevalence rates), conditional on all risk factors. The prevalence rates describe the probability of having at least the specific disease.

The dependency of the prevalence rates on the population risk factors is used in the model equations to describe the disease incidence and mortality to all other causes of death being distinguished. This dependency relation can be approximated by assuming that the risk factor probability function for the persons having the disease (disease prevalence) or dying due to the disease (cause-specific mortality) is equal to that for the persons getting the disease (disease incidence). Following this approximation it is only necessary to describe the unconditional prevalence rates instead of the conditional ones for all risk risk factor classes being distinguished.

Assuming independently distributed risk factors the joint probability function for all risk factor classes is uniquely defined by the marginal risk factor prevalence rates. Introducing dependency relations between risk factors results in more complex model structures.

APPENDIX B. DISTINGUISHING CAUSES OF DEATH AND COMPETING DEATH RISKS

B.1. The main model assumptions related to mortality

With respect to mortality we distinguish several causes of death. These causes are the disease categories being distinguished explicitly and the remaining causes. The main assumptions underlying the chronic diseases modelling approach related to mortality are (see also §2.4):

One essential assumption:

- the assumption of conditional independence: the mortality rates for two different diseases are independent conditional on the risk factor variables, for example the CHD and stroke mortality rates are independent conditional on gender, age, smoking and bloodpressure status.

Other non-essential assumptions meaning that releasing them does not violate the conditional independence assumption:

- The cause-specific case fatality rate is independent on the risk factors, for example the CHD case fatality rate is independent on smoking and bloodpressure status.
- The mortality rate for all remaining causes is independent on the risk factors being distinguished.
- The cause-specific mortality is proportional to the disease prevalence.

B.2. The main model equations

For every one-year time step during the simulation period the mortality is calculated in the following way:

- A The cause-specific mortality for those having the related disease
- B The total mortality for all risk factor classes being distinguished
- C The mortality for all other causes for all disease states being distinguished

Notes:

- The order of calculating the mortality numbers is imperative.
- All model variables and parameters, whenever data are available, are specified by gender and age. To simplify the notation of the model equations we omit this specification as well as the time parameter.

A The cause-specific mortality for those having the related disease

$mort_i(Z) = let_i \cdot prev_i(Z)$ mortality due to cause of death i

with: i index over all causes of death (diseases) being distinguished

Z, z	any risk factor class
$mort_i(Z)$	cause-specific mortality rate for any person with risk factor Z
let_i	cause-specific case fatality rate for any person having the disease (patient)
$prev_i(Z)$	risk factor class-specific disease prevalence rate

B The total mortality for all risk factor classes being distinguished

$$mort(Z) = \sum_i mort_i(Z) + mort_{ov}$$

with: $mort(Z)$ risk factor class-specific mortality rate
 $mort_{ov}$ mortality rate for all remaining causes

C The total mortality for all disease states being distinguished

$$let_i prev_i + \sum_{j \neq i} \sum_z mort_j(z) prev_i(z) f(z) + mort_{ov} prev_i$$

with: j index over all other causes of death being distinguished,
 $f(z)$ risk factor prevalence rate

In the last equation the first term is the cause-specific mortality already described in A. The second term is the mortality due to all other causes of death being distinguished. The third term is the mortality due to all truly remaining causes. For calculation of the variables $mort_j(z)$ and $prev_i(z)$, see §C.3.

B.3. Some remarks

Three groups of causes of death being distinguished

In the formula under C three groups of causes of death are distinguished. (1) The specific disease as cause of death. The related mortality rate (conditional on having the disease) is described by the case-fatality rate let_i . This mortality rate is assumed independent on the risk factors. (2) The other causes of death (diseases) being distinguished. The mortality is described by mortality rates that are risk factor-dependent caused by the comorbidity due to the common risk factors. (3) The mortality due to all truly remaining causes of death. This mortality is assumed independent on the risk factors.

Order of calculation steps

The model calculation steps related to disease incidence and mortality are summarized in figure 4. Note that in this figure only transitions related to incidence and mortality are presented.

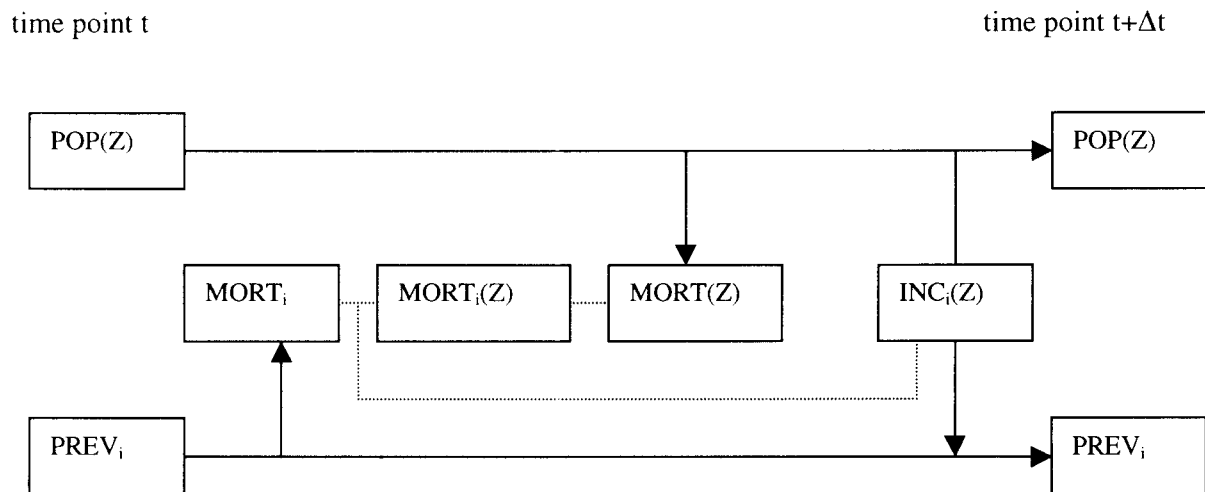


Figure 7: the order of calculating disease incidence and mortality numbers

Integrative aspects being taken account of

This way of mathematically describing risk factor-specific, cause-specific and total mortality takes account of the following aspects of integrating causes of death (see also §2.4):

- mortality selection: persons with high risk level have higher mortality risks,
- independent competing risks: all persons are 'exposed' during their lives to all causes of death,
- dependent competing risks: persons having a disease have higher mortality risks for any other cause compared to those not having the disease whenever both diseases (causes of death) have common risk factors.

APPENDIX C. RELATIVE RISKS IN THE MODEL STRUCTURE

C.1. The risk factor-specific incidence rates

Every person not having a disease has a probability of getting the disease. The disease incidence rate is dependent on the risk factors being distinguished using relative risks. The two main formulas being used to estimate the incidence rates are:

for starting time point t_0 $inc_{0i} = inc_i(t_0) / \sum_z RR_i(z) f(z;t_0)$

for any time point $t \geq t_0$ $inc_i(t) = \sum_z RR_i(z) inc_{0i} f(z;t)$

with: inc_i, inc_{0i} disease incidence rate and baseline incidence rate respectively,
 i index over diseases being distinguished,
 $RR_i(z)$ relative disease i incidence risk for risk factor class z
 $f(z;t)$ risk factor z prevalence rate

Notes:

- All model variables and parameters are specified by gender and age.
- The main model parameters used are the initial disease incidence rate $inc_i(t_0)$, the relative risks $RR_i(z)$ and risk factor prevalence rates $f(z;t)$. Note that the risk factor prevalence rates for $t > t_0$ are model output and not input.
- The relative risks and baseline incidence rates are assumed time-constant. Therefore changes in the total incidence numbers can only result from changes in total population numbers and risk factor prevalence rates.
- The relative risks used are related to the event 'disease incidence'. Most data on relative risks refer to the event 'cause-specific mortality'. Due to data restrictions both are assumed equal.

C.2. Comparing relative risks for the outcome disease incidence and mortality

Some analyses have been done on relative risks for the outcomes disease incidence and mortality. The summary results are presented in this paragraph. The results refer to Cox proportional hazards analyses on data from the Zutphen-study, computer model simulation results and some relative risks from literature based on analyses on data from the Framingham-study.

C.2.1. Zutphen-study

The Cox proportional hazards model has been fit on data from the Zutphen-study. The estimated regression coefficients are presented in table 2. Coefficients have been estimated for the risk factors systolic bloodpressure, serum cholesterol level, number of cigarette years, Body Mass Index, and age, and for the causes of morbidity and mortality CHD, stroke, other heart diseases, lung cancer and other cancer.

Table 2: regression parameters for outcome variables disease incidence and mortality

	CHD	CVA	othHD	lungc	other c
syst	18;12	14ns;18	15;11		
chol	22;21	21ns;13ns	;19		
cigyr	;03ns	;06ns	;10	16;15	07ns;07
BMI	10;06	13;08	09;		
age	10;05	13;09	09ns;05	05;06	10;08

Notation: x;y: regression coefficient (*.01) for outcome mortality and incidence respectively; blank: no value calculated; ns: not significant.

The differences in the estimated regression coefficients depend on the risk factor and cause of death selected. For bloodpressure, BMI, CHD and stroke the coefficients for the outcome incidence are significantly smaller than those for outcome mortality, for cholesterol non-significantly smaller. For cancer the differences are very small. Several reasons may explain these differences. (1) Different case fatalities: the larger the case fatality, the smaller the differences for the two outcome types. (2) Different patient structures. The more homogeneous the patient population, again the smaller the differences. (3) Age at mortality is always higher than age at incidence.

C.2.2. Computer model simulations

Some simulation experiments have been performed on a simplified version of the chronic diseases model that describes only one birth-cohort. The model results have been presented in table 3.

Table 3: input disease incidence rates and model output prevalence rates

age	males				females			
	CHD		lungcancer		CHD		lungcancer	
0	2.8	*	22.4	*	3	*	11.9	*
10	2.8	*	22.4	*	3	*	11.9	*
20	2.8	*	22.4	*	3	*	11.9	*
30	2.8	2.43	22.4	11.4	3	2.57	11.9	7.8
40	2.8	2.52	22.4	16.3	3	2.63	11.9	9.6
50	2.8	2.67	22.4	19.5	3	2.89	11.9	10.9
60	2.8	2.53	22.4	16.7	3	2.75	11.9	10.3
70	1.6	1.92	22.4	22.1	1.6	1.95	11.9	11.8
80	1.6	1.65	22.4	22.3	1.6	1.60	11.9	11.9
90	1.6	1.60	22.4	22.4	1.6	1.57	11.9	11.9
100	1.6	1.59	22.4	22.4	1.6	1.58	11.9	11.9

Notation: *: no relative risk calculated because mortality rate is too small.

The relative mortality risks converge over age to the relative incidence risks values, although changes over age cross these trends. Therefore age-dependent relative mortality risks may result from age-dependency of disease incidence, case fatality and/or mortality due to other causes.

C.2.3. Framingham Study

Based on data from the Framingham Study many relative risks, regression coefficients etc. have been presented in reports, books and scientific articles. Without having the intention to present a complete overview some results are presented related to the issue of relative incidence and mortality risks (see table 4). Only those risk values are presented that have been published within the same article.

Table 4: examples from disease incidence and mortality risks simultaneously presented in literature

Doyle et al. (1962):	Incidence rates (/1000 py)	
	smokers	non-smokers
AP adj	1.6	2.1
MI adj.	7.3	2.1
death from CHD	2.4	0.4

Framingham report section 30: coefficients of logistic regression model for outcome variable CHD

	incidence		mortality	
	univariate	multivariate	univariate	multivariate
SBP	.017	.012	.023	.015
DBP	.024	.021	.032	.024
Chol	.0058	.0058	.0060	.0061
Rel wght	.011	.010	.011	.008
Smoking	.12	.24	.18	.32

Notes: SBP: systolic bloodpressure (mmHg), DBP: diastolic bloodpressure (mmHg), chol: serum cholesterol level (mg/dl), rel wght: Metropolitan relative weight, smoking (number per day)

C.2.4 Some concluding remarks

The empirical results (Zutphen-study and Framingham-study) show that relative incidence risks are generally smaller than relative mortality risks, especially for the outcome coronary heart diseases. For cancer no significant differences have been found. These differences can be explained in terms of different case fatalities, population heterogeneity and/or mortality to other causes, in relation to the time difference between age at disease incidence and age at death.

The model gives conflicting results: larger relative incidence risks than mortality risks, assuming age-constant risk values. This assumption may be too strict. In case of relative incidence risks, that decrease over age, the relative mortality risks follow the incidence risks with a time-lag and so may still be larger for given age values.

C.3. Risk ratio and relative risk

The key model parameters that describe the dependency of the disease variables to the risk factor variables are the relative risk and related risk multiplier. The risk multiplier is defined as follows. For reasons of simplicity we omit the specification by gender and age:

$$RM_i(Z;t) \equiv RR_i(Z) / \sum_z RR_i(z) f(z;t)$$

with: RM_i risk multiplier
 i index over diseases (causes of death) being distinguished
 Z, z specific and any risk factor class respectively
 RR_i relative risk
 $f(.)$ risk factor prevalence rate

The relative risk describes the risk relative to the baseline rate, the risk multiplier describes the risk relative to the (mean) population rate. The time-dependent mean relative risk value is defined as:

$$RR_i(t) \equiv \text{mean}(RR_i;t) = \sum_z RR_i(z) f(z;t)$$

Using these parameters the relation between the total and risk factor-specific incidence rates becomes:

$$\text{inc}_i(Z;t) / \text{inc}_i(t) \approx RR_i(Z) \text{inc}_{i0} / \sum_z RR_i(z) \text{inc}_{i0} f(z;t) = RR_i(Z) / \sum_z RR_i(z) f(z;t) =$$

$$RM_i(Z;t)$$

Assuming that the risk factor distribution of the disease prevalence equals that of the disease incidence the same ratio value applies to the disease prevalence and cause-specific mortality:

$$\text{prev}_i(Z;t) / \text{prev}_i(t) \approx \text{mort}_i(Z;t) / \text{mort}_i(t) = RM_i(Z;t)$$

$$PREV_i(Z;t) / PREV_i(t) = \text{prev}_i(Z;t) \text{POP}(Z;t) / \sum_z \text{prev}_i(z;t) \text{POP}(z;t) =$$

$$RR_i(Z) f(Z;t) / \sum_z RR_i(z) f(z;t) = RM_i(Z;t) f(Z;t)$$

Using these relationships the formula for the disease incidence number becomes:

$$INC_i(t) = \sum_z INC_i(z;t) = \sum_z \text{inc}_i(z) \{ \text{POP}(z;t) - PREV_i(z;t) \} =$$

$$\sum_z RR_i(z) \text{inc}_{i0} f(z;t) \{ \text{POP} - RM_i(Z;t) PREV_i(t) \}$$

C.4. Comorbidity

The mathematical model structure based on the conditional independence assumption makes it possible to quantify the comorbidity resulting from common risk factors. We define the comorbidity coefficient (*com_coeff*) as the ratio of the prevalence rate for two diseases simultaneously to the product of the marginal disease prevalence rates, i.e. the ‘empirical’ combined prevalence rate to the one assuming independent separate disease rates. For reasons of simplicity the specification by gender, age and time has been omitted.

$$\text{com_coeff}_{ij}(t) = \text{prev}_{i\&j}(t) / \text{prev}_i(t) * \text{prev}_j(t) =$$

$$\sum_z \text{prev}_{i\&j}(z;t) f(z;t) / \{ \sum_z \text{prev}_i(z;t) * \sum_z \text{prev}_j(z;t) \} =$$

$$\sum_z \text{prev}_i(z;t) \text{prev}_j(z;t) f(z;t) / \{ \sum_z \text{prev}_i(z;t) * \sum_z \text{prev}_j(z;t) \} =$$

$$\sum_z \text{RR}_i(z) \text{RR}_j(z) f(z;t) / \{ \sum_z \text{RR}_i(z) f(z;t) \sum_z \text{RR}_j(z) f(z;t) \} =$$

$$\text{mean}(\text{RR}_i * \text{RR}_j; t) / \{ \text{mean}(\text{RR}_i; t) \text{mean}(\text{RR}_j; t) \}$$

with: i, j indices over diseases being distinguished

C.5. Multiplicative or additive risk model

So far we have only used relative risk values for risk factor classes that are defined on simultaneous combinations of specific risk factor values (‘combined exposure’), for example persons who do not smoke but have hypertension. In this paragraph we work out the functional relationship between the relative risk values for combined exposures to those for single exposures assuming a multiplicative or additive risk model.

$$\text{multiplicative risk model} \quad \text{RR}_i(Z_{1..n}) = \prod_j \text{RR}_i(Z_j)$$

$$\text{additive risk model} \quad \text{RR}_i(Z_{1..n}) = 1 + \sum_j \{ \text{RR}_i(Z_j) - 1 \}$$

We apply these risk models together with the assumption of independently distributed risk factors:

$$f(Z) = \prod_j f_j(Z_j)$$

with: $f(Z)$ joint risk factor prevalence rate
 j index over risk factors
 $f_j(Z_j)$ marginal risk factor j prevalence rate

Combining both risk models with the assumption of independently distributed risk factors results in:

multiplicative risk model

$$\text{mean}(\text{RR}_i) = \sum_j \sum_{k(j)} \Pi_l \text{RR}_i(z_{l,k(j)}) \Pi_l f(z_{l,k(j)}) = \Pi_j \sum_k \{ \text{RR}_i(z_{j,k}) f_j(z_{j,k}) \}$$

$$\text{RM}_i(Z) = \text{RR}_i(Z) / \text{RR}_i = \Pi_j \text{RR}_i(Z_j) / \Pi_j \sum_k \{ \text{RR}_i(z_{j,k}) f_j(z_{j,k}) \} =$$

$$\Pi_j \{ \text{RR}_i(Z_j) / \sum_k [\text{RR}_i(z_{j,k}) f_j(z_{j,k})] \} = \Pi_j \text{RM}_i(Z_j)$$

additive risk model

$$\text{mean}(\text{RR}_i) = \sum_j \sum_{k(j)} \{ 1 + \sum_l [\text{RR}_i(z_{l,k(j)}) - 1] \} \Pi_l f_i(z_{l,k(j)}) =$$

$$1 + \sum_j \sum_{k(j)} \{ \text{RR}_i(z_{j,k}) - 1 \} f_j(z_{j,k})$$

$$\text{RM}_i(Z) = \text{RR}_i(Z) / \text{RR}_i = \{ 1 + \sum_j [\text{RR}_i(Z_j) - 1] \} / \{ 1 + \sum_j \sum_{k(j)} [\text{RR}_i(z_{j,k}) - 1] f_j(z_{j,k}) \}$$

- with: i index over diseases (causes of death)
 k index over classes for given specific risk factor (domain of k depends on number of classes being distinguished for risk factor j)
 l index over risk factor classes for given indices k(j)
 $\text{RR}_i(z_{j,k})$ relative risk for class k of risk factor j
 $f_j(z_{j,k})$ prevalence rate for class k of risk factor j

The disease incidence and mortality rates for both risk models are:

baseline incidence rate

$$\text{inc}_{oi} = \text{inc}_i(t_o) / \text{mean}(\text{RR}_i; t_o)$$

disease incidence number

$$\text{INC}_i = \text{inc}_{oi} \text{POP} \{ \text{mean}(\text{RR}_i; t) - \text{mean}(\text{RR}_i * \text{RR}_i; t) / \text{mean}(\text{RR}_i; t) \}$$

with: $\text{mean}(\text{RR}_i * \text{RR}_i; t) =$

$$\text{multiplicative risk: } \Pi_j \sum_k \{ \text{RR}_i(z_{j,k}) \text{RR}_i(z_{j,k}) f_j(z_{j,k}) \}$$

$$\text{additive risk: } \sum_j \sum_k \{ 1 + \sum_l [\text{RR}_i(z_{l,k(j)}) - 1] \}^2 \Pi_l f_i(z_{l,k(j)})$$

mortality for those having disease i due to diseases (cause of death) j≠i

$$\sum_z \text{mort}_j(z) \text{prev}_i(z) f(z) = \text{mort}_j * \text{prev}_i *$$

$$\text{multiplicative risk: } \prod_j \sum_k \text{RR}_i(z_{j,k}) \text{RR}_j(z_{j,k}) f(z_{j,k}) / \{ \text{mean}(\text{RR}_i) \text{mean}(\text{RR}_j) \}$$

$$\text{additive risk: } \sum_j \sum_k \{ 1 + \sum_i [\text{RR}_i(z_{j,k}) - 1] \} \{ 1 + \sum_j [\text{RR}_j(z_{j,k}) - 1] \} \prod_i f_i(z_{j,k}) / \\ \{ \text{mean}(\text{RR}_i) \text{mean}(\text{RR}_j) \}$$

The last two terms are equal to the comorbidity coefficient that has been defined in §C.4.

C.6. Some remarks

Sensitivity of model results to relative risk values

The relative risks are used in the starting time point to disaggregate the disease incidence rate over the risk factor classes. This disaggregation is internally consistent meaning that aggregating the risk factor-specific incidence rates results in the original rate value. As a result of the disaggregation step the model results on the population level (disease prevalence rates, mortality rates) are almost insensitive to the relative risk values. However, for population subgroups the model results may be sensitive, for example the life expectancies for different risk factor classes.

D. CHANGE OF DISEASE PREVALENCE OVER TIME AND AGE

D.1. The main model equation

So far only the change over time of the probability of having at least one specific disease has been described (see §A.5). To describe the change of the disease prevalence rate one also has to take account of the change of the survival probability. A simple differential equation for the disease prevalence rate can be derived making two extra assumptions next to the assumption of conditional independence: there are no risk factors, and there is no recovery. Then we find:

$$\Delta \text{PREV}_i / \Delta t \approx \text{inc}_i (\text{POP} - \text{PREV}_i)$$

$$\Delta \text{POP} / \Delta t \approx - \text{let}_i \text{PREV}_i - \text{mort}_{\text{ov}} \text{POP}$$

$$\Delta \text{prev}_i / \Delta t \approx \Delta (\text{PREV} / \text{POP}) / \Delta t = (\text{inc}_i - \text{let}_i \text{prev}_i) (1 - \text{prev}_i)$$

with:	PREV_i	probability of having at least disease i
	prev_i	disease prevalence rate
	POP	survival probability
	inc_i	disease i incidence rate
	let_i	cause-specific mortality rate for those having disease i
	mort_{ov}	mortality rate for all remaining causes of death

Notes:

- The change of the disease prevalence rate over age is independent on the mortality due to all other causes of death, under the assumptions being made.
- The model equation has been derived for a birth cohort and prev describes the age-specific prevalence rate. The model equation can also be interpreted in terms of following a general cohort over time. Then the prevalence is stable (time-constant, $\Delta \text{prev} = 0$) if $\text{prev}_i = \text{inc}_i / \text{let}_i$, or, equivalently, if prevalence rate = incidence rate * mean disease duration.
- The model equation has been used in the DisMod-exercise by Murray c.s. (1996) to get internally consistent disease incidence, prevalence, and mortality data. The same exercise will be done for the Dutch situation (Hoogenveen et al., 1998c)
- The cause-specific case fatality rate describes the excess hazard rate for patients compared to persons without the disease. See also §2.6.

D.2. Relating time trends in disease incidence, prevalence and case fatality rate

The differential equation can be used to relate changes (trends) in disease incidence or case fatality rates to changes in disease prevalence and cause-specific mortality rates:

rates:	disease incidence	case fatality
reference situation	inc	mort
alternative situation	(1+ α) inc	(1+ β) mort

with: α , β : relative change parameters. Then the relative change in the disease prevalence and cause-specific mortality rate over time are:

$$\Delta \text{prev}_i \equiv (\text{prev}_{i,\text{alt}} - \text{prev}_{i,\text{ref}}) / \text{prev}_{i,\text{ref}} \approx (1 - \text{prev}_{i,0}) / \text{prev}_{i,0} (\text{inc } \alpha - \text{mort } \beta \text{ prev}_{i,0}) \Delta t$$

$$\Delta \text{mort}_i \equiv (\text{mort}_{i,\text{alt}} - \text{mort}_{i,\text{ref}}) / \text{prev}_{i,\text{ref}} \approx \beta \Delta t + \Delta \text{prev}_i$$

with: $\text{prev}_{i,\text{ref}}$, $\text{prev}_{i,\text{alt}}$ age-specific prevalence rate in reference and alternative situation respectively

Δprev_i relative change of age-specific prevalence rate when comparing alternative situation to reference situation

Δt small time period

Notes:

- The relation between changes in the input parameters to those in the output variables can be interpreted easily. For example, an increasing case fatality rate results in changing cause-specific mortality rates along two lines, one directly resulting in increasing mortality rates and one indirectly resulting in decreasing prevalence rates and therefore also decreasing mortality rates.
- The term $(\text{inc } \alpha - \text{mort } \beta \text{ prev}_{i,0})$ can be interpreted as the net growth. For small time periods the net growth is proportional to the time length.

APPENDIX E. HEALTH OUTCOME INDICATORS

The model outputs are risk factor prevalence rates, and cause-specific disease incidence, prevalence and mortality rates. These model outputs can be transformed (integrated) to health outcome indicators.

E.1. Life and remaining life expectancy

Standard life table method using total mortality rates.

E.2. Reduction of life expectancy ‘due to’ high risks

We can define the (remaining) life expectancy for specific population risk groups. However, there are several problems here: (1) Individuals enter the high risk groups at different ages. (2) Individuals may leave the high risk group at higher ages. (3) The number of life years is defined with respect to any (one) individual, the chronic diseases model structure describes only dynamic population groups. These problems can be dealt with by introducing a matching context in the life expectancy definition. For any individual entering the high risk group (a ‘case’) his remaining life expectancy is computed assuming he will stay in the high risk group until death. Also the remaining life expectancy is calculated of a ‘control’ that stays in the normal risk group. The difference between both the mean of the ‘cases’ and the ‘controls’ is defined as the reduction of life expectancy ‘due to’ the high risk.

The (time-continuous) formula for any risk factor Z is:

$$\int_{u \geq 0} \{ \text{POP}_0(u) - \text{POP}_0(u)^{\text{RR}(u)} \} \left\{ \int_{t \leq u} \text{inc}(t) P(Z=0;t) / \text{POP}_0(t) dt \right\} du /$$

$$\int_{t \geq 0} \text{inc}(t) P(Z=0;t) dt$$

with: POP_0 the survival fraction for the baseline risk group
 RR the relative risk for the high risk group
 inc the disease incidence rate
 $P(Z=0)$ the probability of normal risk given survival until age t

This implementation of differences in life expectancy between risk factor classes is based on a matching context over all ages. Of course it is also possible to define these differences only for one age point. In this case it is also assumed that the ‘cases’ and ‘controls’ maintain their risk factor status over life.

E.3. Other versions of life expectancy

Potential years of life lost due to cause i specific mortality:

$$\int_{t \geq 0} \text{POP}(t) \text{mort}_i(t) e(t) dt$$

with: mort_i mortality rate on age t
 $e(t)$ rest life expectancy on age t

Quality (disability) adjusted life years

Survival and disease prevalence fractions may be weighted by their quality of life (QALY) values. We work out the QALY (DALY)-calculation in more detail in relation to the comorbidity that is described in the model structure based on the assumption of conditionally independent disease prevalence rates. It is assumed that:

- the QALY (DALY) weight-value (the weighting of life years in terms of full health equivalents) is independent on the population risk factors conditional on the disease states,
- 'being dead' has QALY weight-value 0, DALY weight-value 1 respectively.

Because QALYs and DALYs can be treated mathematically in the same way, we will present only results in terms of QALYs. The QALY-value of the population can be computed under an additivity or multiplicativity assumption of the disease-specific QALY-values:

Additivity assumption:

The disease-specific reductions of the QALY-value are additive.

$$\sum_{D,Z} \text{QALY}(D) P(D,Z;t) = (1 - \sum_i \text{QALY}_i \text{prev}_i(t)) \text{POP}(t)$$

Multiplicativity assumption:

The disease-specific reductions of the QALY-values are multiplicative.

$$\sum_{D,Z} \text{QALY}(D) P(D,Z;t) = \prod_i (1 - \text{QALY}_i \text{prev}_i(t)) \text{POP}(t)$$

with: $\text{QALY}(D)$ the multivariate QALY-value,
 QALY_i the disease-specific QALY-value, $\text{QALY}(D) = \sum_i \text{QALY}_i \text{ind}(D_i=1)$.

APPENDIX F. DEFINITIONS

analysis

sensitivity - analyzing the effects on the model output of changes in the model input

case fatality rate excess hazard rate of persons with a disease compared to those without the disease

clustering joint high risk prevalence rate is larger than is expected assuming independent univariate high risk prevalence rates

comorbidity joint disease prevalence rate is larger than is expected assuming independent univariate disease prevalence rates

conditional independence assumption that disease rates are independent conditional on the risk factors
disease

- progression transition to disease state with more serious health effects

- remission transition of state 'with the disease' to state 'without the disease'

incidence

- number number of persons getting the disease during time interval

- rate fraction of population getting the disease during time interval

life

- expectancy expected rest number of years to live for any individual; mostly estimated for newborn

- table mathematical method using class-specific mortality rates to calculate life expectancy figures

Markov-property given the actual state the (probability distribution of the) future model states are (is) independent on the past states

mortality

rest - mortality due to other causes than those being described separately

- selection population characteristics change over time due to different individual mortality rates

prevalence

- number number of persons with disease on time point (starting point of time interval)

- rate fraction of population number with disease on time point

prevention

primary - policy measures on reducing risk levels to reduce morbidity and mortality

secondary - policy measures on reducing disease progression or promoting disease remission

risk

- behaviour change of risk factor values over time within individuals

competing death - distinguishing several causes of death simultaneously

- model model that relates relative risk for combined exposure to those for single exposures

- multiplier event probability of exposed persons relative to population probability

-
- relative - event probability of exposed persons relative to that of non-exposed persons
- scenario - possible future developments implemented as possible future model input values together with the related model output values
- model
- deterministic - given the mathematical model equations and model input values the model output is fixed
 - multistate - model that describes persons that can be in different states
 - parameter variable with fixed value
 - simulation - model that mimics events over time
 - stochastic - given the mathematical model equations and model input values the model output is stochastic
 - system-dynamic - model based on the concepts of states and state transitions
 - validation proving the reliability, accuracy and functionality of the model within the specified domain of model applications

APPENDIX G. ABBREVIATIONS

α, β	model parameters
BMI	Body Mass Index
CBS	Statistics Netherlands
CHD	coronary heart diseases
COPD	chronic obstructive pulmonary diseases
CVA	cerebrovascular accident = stroke
CVD	cardiovascular diseases
D	vector of states for all diseases being distinguished
DALY	Disability Adjusted Life Year
DM	diabetes mellitus
f	transition rate
f(.)	risk factor prevalence rate
i	index over diseases being distinguished; ov = remaining causes of death
inc	disease incidence rate
j	index over risk factors being distinguished
k	index
let	cause-specific case fatality rate
μ, mort	mortality rate for any individual
MORT	mortality number
n(.)	set of neighbouring states
P	probability
POP	probability of survival / population number
prev	disease prevalence rate
PREV	probability of disease prevalence / prevalence number
QALY	Quality Adjusted Life Year
rem	disease remission rate
RM	risk multiplier
RR	relative risk
S(.)	age-dependent survival probability
SES	social-economic status
t, Δt	time point, small time period
z, Z	vector of states for all risk factors being distinguished
Σ	summation
Π	product
\int	integral

LITERATURE

Ackerman E, Zhuo Z, Altman M, Kilis D, Yang J-J, Seaholm S, Gatewood L. Simulation of stochastic micropopulation models. I: the SUMMERs simulation shell. *Comput Biol Med*, vol 23:3 (1993), pp 177-198

Ackerman E. Simulation of micropopulations in epidemiology. Tutorial 1, Simulation: an introduction. Tutorial 2. ... Tutorial 3: simulation model evaluation methods. Tutorial 4: evaluations on simulation models. *International Journal of Bio-Medical Computing*. Vol 36/36/37/39 (1994/1994/1994/1995), pp 229-238/..195-204/219-229

Andersen PK. Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, vol 7 (1988), pp 661-670

Barendregt JJ, Bonneux L. De invloed van veranderingen in incidentie en overleving van cardiovasculaire ziekten op ziekteprevalentie en gezonde levensverwachting. [The effect of changes in incidence and survival of cardiovascular diseases on prevalence and healthy life expectancy.] *Tijdschrift Sociale Gezondheidszorg*, vol 72:8 (1994), pp 429-433 (in Dutch)

Barendregt JJM, Bonneux LGA. Degenerative disease in an aging population. Models and conjectures. Thesis. iMGZ, Erasmus University Rotterdam, 1998

Blokstra A, Smit HA, Verschuren WMM, Bueno de Mesquita HB, Seidell JC. Het Project Monitoring Risicofactoren en Gezondheid Nederland (MORGEN-project) Jaarverslag 1995. [The Risk Factors and Health State Monitoring Project (Morgen-project). Annual report 1995.] RIVM-report 263200005. RIVM, Bilthoven, 1997 (in Dutch)

Boulanger B. Design of a simulation shell for the WHO-MONICA project. *MedInfo* (1992), pp 1-5

CBS. Bevolking van Nederland naar burgerlijke staat en geslacht per leeftijd, 1 januari. Sterfte naar doodsoorzaak, Migratiesaldo, Aantal levendgeborenen, Administratieve correcties. [Statistics Netherlands: Dutch population numbers on jan. 1, Mortality numbers by cause of death, Migration numbers, Number of newborns, Administrative corrections.] *Maandstatistiek Bevolking* (in Dutch)

Checkoway H, Hickey JLS. Estimating the potential disease rate reduction under varying conditions of combined effects. *International Journal Epidemiology*, vol 17:3 (1988), pp 666-672

Chiang CL. Introduction to stochastic processes in biostatistics. Wiley, New York, 1968

Chiang CL. Competing risks in mortality analysis. *Annu. Rev. Public health*, vol 12 (1991), pp 281-307

Doyle JT et al. Cigarette smoking and coronary heart disease. Combined experience of the Albany and Framingham studies. *New England Journal of Medicine*, vol 266:16 (1962), pp 796-801

Drummond MF, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford University Press, Oxford, 1987

Framingham Study. Kannel WB, Gordon T (eds.). *An epidemiological investigation of cardiovascular disease*. Shurtleff D. Section 30. Some characteristics to the incidence of cardiovascular disease and death: Framingham Study, 18-year follow-up. DHEW Publication no (NIH) 74-599, Washington, 1974

Genugten MLL van (b), Hoogenveen RT, Hulshof JAM, Jager JC. *Een methodologische handleiding voor scenario-onderzoek in de volksgezondheid en de gezondheidszorg. De praktijk bij de STG*. [Methodological guideline for scenario-research in public health and health care. The practice of STG.] Jan van Arkel, 1993, Utrecht

Genugten MLL van (b), Jager JC, Helder JC, Heijden OG van der, Hoogenveen RT, Kramers PGN, Poos MJJC, Postma MJ, Reinders A. *Het conceptuele kader voor de integrale modellering van de volksgezondheid*. [The conceptual framework for the integrative public health modelling.] RIVM-report 431501001. RIVM, Bilthoven, 1993 (in Dutch)

Genugten MLL van, Hoogenveen RT, Mulder I, Smit HA, Jansen J, Hollander AEM de. *The effectiveness of different policies to reduce tobacco smoking in terms of future smoking prevalence and disease burden*. 1998, submitted

Giessen A van der. *Hoofdelementen van een procedure voor modelvalidatie*. [Key elements of a model validation procedure.] Internal memo. RIVM, Bilthoven, 1997

Gijzen R, Verkleij H, Dijksterhuis PH, Lisdonk EH van der, Metsemakers JFM, Velden J van der. *Ziektespecifieke vergelijking van de geregistreerde morbiditeit in vier huisartsen registraties: een analyse ten behoeve van VTV-1997*. [Disease-specific comparison of morbidity figures in four general practice morbidity registration projects: analysis in behalf of VTV-1997.] RIVM-report 431501017. RIVM, Bilthoven, 1997 (in Dutch)

Grinsven JJM van, Haan BJ de, Braat LC. *Ingrediënten voor kwaliteitsborging van modelonderzoek*. [Elements of quality assurance of model development.] RIVM-report 408136001. RIVM, Bilthoven, 1995 (in Dutch)

Gunning-Schepers LJ. *The health benefits of prevention, a simulation approach*. Thesis. iMGZ, Erasmus University Rotterdam, 1988

Habbema JDF, Casparie AC, Mulder JH, Rutten FFH (red.) Medische technology assessment en gezondheidsbeleid. [Medical technology assessment and health policy.] Samson Stafleu, Alphen a/d Rijn, 1989

Hart HE, Wouden JC van der, Höppener P, Schendel GJ, Metsemakers JFM, Knottnerus JA. Huisartsgeneeskundige registraties in Nederland. [General practice registrations in the Netherlands] Instituut Huisartsgeneeskunde Erasmus Universiteit Rotterdam/Vakgroep Huisartsgeneeskunde Rijksuniversiteit Limburg, 1996

Hoogenveen RT, Brouwer JJLM. Scenario's in de volksgezondheid. Inleiding in de methodiek van de STG. [Scenarios in public health. Introduction to the STG methodology.] Jan van Arkel, Utrecht, 1989 (in Dutch)

Hoogenveen RT, Jager JC. Overlevingsduuranalyse en concurrerende doodsoorzaken. Een inleiding. [Survival analysis and competing death risks. An introduction.] RIVM-report 958606001. RIVM, Bilthoven, 1990 (in Dutch)

Hoogenveen RT. Demografische modellen voor scenario-onderzoek. [Demographic models for scenario studies.] RIVM-report 958606005. RIVM, Bilthoven, 1991 (in Dutch)

Hoogenveen RT et al. Resultaten van een analyse van de Zutphen-studie met als uitkomstvariabele morbiditeit. [Results of analyses on the Zutphen-study with outcome morbidity.] CWM-memo 014/93 RIVM, Bilthoven, 1993 (in Dutch)

Hoogenveen RT, Genugten MLL van, Jager JC. Methodological aspects of integral public health and health care modelling. CWM-Memo 004/94. RIVM, Bilthoven, 1994

Hoogenveen RT (a), Smit HA, Genugten MLL van, Mulder I, Jansen J, Hollander EAM den. Estimating transition rates over time and age from series of cross-sectional studies. 1998, submitted

Hoogenveen RT (b), Genugten MLL van, Hollander AEM de. Modelling the hierarchical relationship between social-economic status and other population risk factors. RIVM-report. RIVM, Bilthoven, 1998 (in preparation)

Hoogenveen RT (c), Genugten MLL van, Kommer GJ, Hollander AEM de, Schouten JSAG. Estimating case fatality rates from disease incidence, remission and prevalence rates. RIVM-report. RIVM, Bilthoven, 1998 (in preparation)

Johannesson M, Hedbrant J, Jonsson B. A computer simulation model for cost-effectiveness analysis of cardiovascular disease prevention. Med Inform, vol 16:4 (1991), pp 355-362

Kalbfleisch D, Lawless JF. The analysis of panel data under a Markov assumption. *Journal American Statistical Association*, vol 80 (1985), pp 863-871

Kannel WB, Gordon T (eds.). *The Framingham Study. An epidemiological investigation of cardiovascular disease. Section 30. Some characteristics related to the incidence of cardiovascular disease and death: Framingham Study, 18-year follow-up.* Shurtleff, 1974

Kannel WB, Woolf PA, Garrison RJ (eds.). *The Framingham Study. An epidemiological investigation of cardiovascular disease.* DHEW Publications no (NIH) 87-2284. Public Health Service, Bethesda, Maryland, 1987

Keilman N, Gill R. On the estimation of multidimensional demographic models with population registration data. *Kwantitatieve Methoden*, vol 23 (1987), pp 85-115

Keys A, Aravanis C, Blackburn H, Buchem FSP van, Buzina R, Djordjevic BS, Dontas AS, Fidanza F, Karvonen MJ, Kimura N, Lekos D, Monti M, Puddu V, Taylor HL. Epidemiological studies related to coronary heart disease; characteristics of men aged 40-59 in seven countries. *Acta Med Scand*, vol 460 (suppl) (1967), pp 1-392

Koopmanschap MA, Roijen L van, Bonneux L. *Kosten van ziekten in Nederland. [Costs of diseases in the Netherlands.]* iMGZ, Erasmus University Rotterdam, 1991 (in Dutch)

Land KC, Rogers A. (eds) *Multidimensional mathematical demography.* Academic Press, New York, 1982

Manton KG, Patrick CH, Stallard E. Population impact of mortality reduction: the effects of elimination of major causes of death on the 'saved' population. *International Journal of Epidemiology*, vol 9:2 (1980), pp 111-120

Manton KG, Stallard E. *Chronic disease modelling: measurement and evaluation of the risks of chronic disease processes.* Charles Griffin & Company Ltd, London, 1988

Manton KG, Woodbury MA, Stallard E. Models of the interaction and the evolution of risk factor distribution: a general stochastic process formulation. *Statistics in Medicine*, vol 7 (1988), pp 239-256

Manton KG, Singer BH, Suzman RM. *Forecasting the health of elderly populations.* Springer Verlag, New-York, 1993

Mennen LI, Witteman JCM, Geleijnse JM, Stolk RP, Visser MC, Grobbee DE. Risicofactoren voor hart- en vaatziekten bij ouderen; het ERGO-onderzoek. [Cardiovascular disease risk factors in the elderly: the ERGO project.] *Nederlands Tijdschrift voor Geneeskunde*, vol 139:39 (1995), pp 1983-1988

Mishan EJ. Cost benefit analysis. Unwin Hyman, London, 1990

Mode CJ. Stochastic processes in demography and their computer implementation. *Biomathematics* vol. 14. Springer Verlag, Heidelberg, 1985

Mulder PGH, Hempenius AL. Analysis of total mortality in the Rotterdam sample of the Kaunas-Rotterdam Intervention Study (KRIS). *Statistics in Medicine*, vol 6 (1987), pp 501-516

Murray CJL, Lopez AD. Global health statistics: a compendium of incidence, prevalence and mortality estimates for over 200 conditions. Harvard University Press, Cambridge, 1996

Murray CJL, Lopez AD (eds.). The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Harvard University Press, Cambridge, 1996

NIDI. Workshop on: Forecasting of mortality in developed countries: searching for better methods and realistic assumptions. Reading Materials. NIDI, 's Gravenhage, 1997

RIVM. Public Health Status and Forecasts. The health status of the Dutch population over the period 1950-2010. Eds: D Ruwaard, PGN Kramers, A van den Berg Jeths, PW Achterberg. Sdu Uitgeverij, The Hague, 1994

RIVM. Volksgezondheid Toekomst Verkenning 1997. 0. De som der delen. Eds: D Ruwaard, PGN Kramers. I. De gezondheidstoestand: een actualisering. II. Gezondheidsverschillen. III. Gezondheid en levensverwachting gewogen. IV. Effecten van preventie. V. Effecten van zorg. VI. Zorgbehoefte en zorggebruik. VII. Gezondheid en zorg in de toekomst. [Public Health Status and Forecasts 1997. 0. Overview. I. An update of the health status. II. Health differences. III. Quality adjusted health and life expectancy. IV. Effects of prevention. V. Effects of care. VI. Health care need and use. V. Future health status and health care.] Elsevier/De Tijdstroom, Amsterdam, 1997 (in Dutch)

Rogers A. Parameterized multistate population dynamics and projections. *Journal American Statistical Association*, vol 81:393 (1986), pp 48-61

Rutten FFH, Ineveld BM van, Ommen R van, Hout BA van, Huijsman R. Kostenberekening bij gezondheidszorgonderzoek. Richtlijnen voor de praktijk. [Cost calculation in health care research. Directives for practice.] Jan van Arkel, 1993, Utrecht (in Dutch)

Schatzkin A, Slud E. Competing risks bias arising from an omitted risk factor. *American Journal of Epidemiology*, vol 129:4 (1989), pp 850-856

Schoen R. *Modeling multigroup populations*. Plenum Press, New York & London, 1988

Shopland DR, Eyre HJ, Pechacek TF. Smoking-attributable cancer mortality in 1991: is lung cancer now the leading cause of death among smokers in the United States? *Journal National Cancer Institute*, vol 83 (1991), pp 1142-1148

Spasoff RA, McDowell IW. Estimating the combined effect of several disease precursors in health risk appraisal. *American Journal Preventive Medicine*, vol 3:4 (1987), pp 182-189

StiVoRo. Stichting Volksgezondheid en Roken (Dutch Smoking Foundation). Annual reports. StiVoro, Den Haag (in Dutch)

Tabeau E. Parameterization functions for mortality by cause of death. Working paper No 1995/1. NIDI, 's Gravenhage, 1995

Verdecchia A, Capocaccia R, Egidi V, Golini A. A method for the estimation of chronic disease morbidity and trends from mortality data. *Statistics in Medicine*, vol 8 (1989), pp 201-216

Verschuren WMM, Smit HA, Leer EM van, Berns MPH, Blokstra A, Steenbrink-van Woerden JA, Seidell JC. Prevalentie van risicofactoren voor hart- en vaatziekten en veranderingen daarin in de periode 1987-1991. Eindrapportage Peilstationsproject Hart- en Vaatziekten 1987-1991. [Cardiovascular disease risk factor prevalence rates and trends over 1987-1991. Final report Monitoring Project on Cardiovascular Disease Risk Factors 1987-1991.] RIVM-report 528901011. RIVM, Bilthoven, 1994 (in Dutch)

Vrieze OJ, Boas GM, Janssen JHA. Een simulatiemodel voor toekomstanalyse van coronaire hartziekten. Rapport opgesteld in opdracht van de Stuurgroep Toekomstscenario's Gezondheidszorg. [A simulation model for forecasts on coronary heart diseases. Report in behalf of the Steering Committee on Future Health Scenarios.] Jan van Arkel, Utrecht, 1994

Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the coronary heart disease policy model. *American Journal Public Health*, vol 77:11 (1987), pp 1417-1426

Wijk J van. M - Language Reference Manual Version 3. RIVM, Bilthoven, 1994

Wolfson MC. A system of health statistics toward a new conceptual framework for integrating health data. *Review of Income and Wealth*, vol 37:1 (1991), pp 81-104

Wong O. A competing-risk model based on the life table procedure in epidemiological studies. *International Journal of Epidemiology*, vol 6:2 (1977), pp 153-159

Zahl P-H. Frailty modelling for the excess hazard. *Statistics in Medicine*, vol 16 (1997), pp 1573-1585