

Degenerative disease in an aging population

Models and conjectures

Financial support for the publication of this thesis by the Netherlands Heart Foundation and the Department of Public Health of Erasmus University is gratefully acknowledged.

Financial support for the research reported in this thesis was provided by the Dutch Ministry of Health, the Netherlands Heart Foundation, and the Netherlands Organization for Scientific Research (NWO).

Cover design: Karin Mulder
Layout: Bon Mot, Rotterdam
Printing: Febodruk, Enschede

ISBN 90-72245-81-4

© Jan Barendregt en Luc Bonneux, 1998

Degenerative disease in an aging population
Models and conjectures

Degeneratieve ziekte in een verouderende bevolking
Modellen en hypothesen

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr P.W.C. Akkermans M.A.
en volgens besluit van het college voor promoties.

De openbare verdediging zal plaatsvinden op
woensdag 14 januari 1998 om 14:00 uur

door

Johannes Jacobus Maria Barendregt
geboren te Schijndel

en op
woensdag 14 januari 1998 om 15:15 uur

door

Luc Gustaaf Antoon Bonneux
geboren te Werl

Promotiecommissie Jan Barendregt

Promotor : Prof.dr P.J. van der Maas

Overige leden : Prof.dr E.K.A. van Doorslaer
Prof.dr G.J. Gunning-Schepers
Prof.dr A. Hofman

Promotiecommissie Luc Bonneux

Promotor : Prof.dr P.J. van der Maas

Overige leden : Prof.dr G.J. Gunning-Schepers
Prof.dr A. Hofman
Prof.dr M.L. Simoons

Contents

Part I

Background, research questions and study design

1. General introduction 3
2. The dirty hands of the epidemiologist 23

Part II

Disease models

3. Introduction 31
4. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. The future rise of heart failure 37
5. The new old epidemic of coronary heart disease 59
6. Stroke trends in an aging population 75
7. Regression analysis of recent changes in cardiovascular morbidity and mortality in the Netherlands 93
8. Diverging trends in colorectal cancer morbidity and mortality: earlier diagnosis comes at a price 103
9. Senescence as cause of hip fractures 119
10. The burden of morbidity and mortality from dementia 131

Part III

Multi-disease models

11. Introduction 145
12. Coping with multiple morbidity in a life table 149
13. Consistent estimates of NIDDM incidence, prevalence and mortality 169
14. Changing cardiovascular epidemiology and its impact on diabetes prevalence: a quantification 181

Part IV

Measuring compression and expansion of morbidity

15. Introduction 193
16. Health Expectancy: an indicator for change? 203
17. Changes in incidence and survival of cardiovascular disease and their impact on disease prevalence and health expectancy 215
18. DALYs: the age-weights on balance 225
19. Health Expectancy: from a population health indicator to a tool for policy making 233

20. Disability Adjusted Life Expectancy and comorbidity: exploring uncertainty and sensitivity	247
21. Update and summary	259
<u>Part V</u>	
Population health status, prevention, and health care costs	
22. Introduction	267
23. Preventing fatal diseases increases health care costs: a cause elimination life table approach	271
24. The health care costs of smoking	281
<u>Part VI</u>	
The evolution of senescence	
25. Introduction	295
26. The expiry date of man	299
<u>Part VII</u>	
Discussion and summary	
27. General discussion and conclusions	313
28. Summary	329
29. Samenvatting	339
Dankwoord	351
Curriculum vitae	355

Publications and manuscripts reprinted in this thesis

Chapter 2

Translation of:

Bonneux, L., & Barendregt, J. J. (1993). De vuile handen van de epidemioloog. *Tijdschrift voor Sociale Gezondheidszorg*, 71(5), 292-294.

Chapter 4

Bonneux, L., Barendregt, J. J., Meeter, K., Bonsel, G. J., & van der Maas, P. J. (1994). Estimating clinical morbidity due to ischemic heart disease and congestive heart failure: the future rise of heart failure. *Am J Public Health*, 84(1), 20-8.

A Dutch version was published as:

Bonneux, L., Barendregt, J., Meeter, K., Bonsel, G., & van der Maas, P. (1994). De toekomst van hartziekten in Nederland: Dalende sterfte en stijgende zorgbehoefte. *Ned Tijdschr Geneeskd*, 138, 1612-18.

Chapter 5

Bonneux, L., Barendregt, J. J., & van der Maas, P. J. (submitted). The new old epidemic of coronary heart disease.

Chapter 6

Niessen, L. W., Barendregt, J. J., Bonneux, L., & Koudstaal, P. J. (1993). Stroke trends in an aging population. *Stroke*, 24, 931-939.

Chapter 7

Bonneux, L., Looman, C. W., Barendregt, J. J., & van der Maas, P. J. (1997). Regression analysis of recent changes in cardiovascular mortality and morbidity in the Netherlands. *Brit Med J*, 314, 789-792.

Chapter 8

Bonneux, L., Barendregt, J. J., Looman, C. W. N., & van der Maas, P. J. (1995). Diverging trends in colorectal cancer morbidity and mortality: earlier diagnosis comes at a price. *Eur J Cancer*, 31A(10), 1665-71.

Chapter 9

Translation of:

Reichgelt, B., Bonneux, L., van den Bosch, J. M. H., Barendregt, J. J., & Pols, H. A. P. (1996). Veroudering als oorzaak van heupfracturen; een theoretisch rekenmodel. *Ned Tijdschr Geneeskd*, 140(25), 1312-1317.

Chapter 10

Witthaus, E., Ott, A., Barendregt, J. J., Breteler, M., & Bonneux, L. (submitted). The burden of mortality and morbidity from dementia.

Chapter 12

Barendregt, J. J., van Oortmarssen, G. J., van Hout, B. A., van den Bosch, J. M., & Bonneux, L. (in press). Coping with multiple morbidity in a life table. *Mathematical Population Studies*.

Chapter 13

Barendregt, J. J., Baan, C. A., & Bonneux, L. (submitted). Consistent estimates of diabetes incidence, prevalence and mortality.

Chapter 14

Barendregt, J. J., Baan, C. A., & Bonneux, L. (manuscript). Changing heart disease epidemiology and its impact on diabetes prevalence: a quantification.

Chapter 16

Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1994). Health Expectancy, an indicator for change? *Journal of Epidemiology and Community Health*, 48, 482-487.

Chapter 17

Barendregt, J. J., & Bonneux, L. (1994). Changes in incidence and survival of cardiovascular disease and their impact on disease prevalence and health expectancy. In: C. Mathers, J. McCallom, & J.-M. Robine (Ed.), *Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994* (pp. 345-354). Canberra: Australian Institute of Health and Welfare: AGPS.

A Dutch translation was published as:

Barendregt, J. J., & Bonneux, L. (1994). De invloed van veranderingen in incidentie en overleving van cardiovasculaire ziekten op ziekteprevalentie en gezonde levensverwachting. *Tijdschrift voor Sociale Gezondheidszorg*, 72(8), 429-433.

Chapter 18

Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1996). DALYs: the age-weights on balance. *Bull World Health Organ*, 74(4), 439-443.

Chapter 19

Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (in press). Health Expectancy: from a population health indicator to a tool for policy making. *Journal of Aging and Health*.

Chapter 20

Barendregt, J. J., & Bonneux, L. (submitted). Disability Adjusted Life Expectancy and Comorbidity: Exploring uncertainty and sensitivity.

Chapter 23

Bonneux, L., Barendregt, J. J., & van der Maas, P. J. (in press). Preventing fatal diseases increases health care costs: a cause elimination life table approach. *Brit Med J*.

Chapter 24

Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1997). The health care costs of smoking. *N Eng J Med*, 337, 1052-7.

Chapter 26

Bonneux, L., Barendregt, J. J., & van der Maas, P. J. (submitted). The expiry date of man.

An earlier Dutch version was published as:

Bonneux, L., Barendregt, J. J., & van der Maas, P. J. (1996). Leven in blessuretijd. *Ned Tijdschr Geneeskd*, 140(52), 2636-2639.

PART I

*Background,
research questions
and study design*

1

General introduction

For descriptive epidemiology, ... approximate estimates are better than no estimates. There is a need for a greater infusion of demographic ethos into the epidemiological estimation process in order to better inform health policies and planning.

C.J.L. Murray & A.D. Lopez

Introduction

This PhD thesis is rooted in a multidisciplinary project, called Technology Assessment Methods (TAM). The (ambitious) aim of the TAM project was to develop a comprehensive method of evaluating medical technology in the perspective of multiple risk factors, multiple diseases and multiple causes of death (Bonneux & Barendregt, 1991). The project was an attempt to better understand the dynamics of population health status, in particular in relation to medical interventions, but it was also motivated by the rapidly rising health care costs of the past decades, which fueled the fear that ever expanding costs might become economically unsustainable in the future (van der Maas & Habbema, 1986). The TAM project would provide a better understanding of the consequences for both costs and population health status of a wide array of preventive and therapeutic health care interventions, and through that offer the tools for policy makers to buy the better investments in health with a sustainable health care budget.

Two factors are commonly held responsible for the increase in health care costs: aging and health care technology. "Aging" is the relative increase in the numbers of elderly people in the population, and since the elderly have more diseases and more serious diseases than young people they are more expensive for the health care system. Aging is brought about by two causes. The first is the consequence of the post war baby boom: the large birth cohorts of these generations are presently showing up as large numbers of people in middle age, who will become large numbers of elderly

in the first four decades of the 21st century. The second cause of aging is that an ever larger majority of each successive birth cohort reaches old age: mortality at all ages has been declining, also among the elderly.

The aging of the baby boom can hardly be held responsible for increases in health care costs so far (the eldest members are presently only in their early fifties), but it will be the major cause in the four decades to come. The second kind of aging has been in operation at least since the second part of the 19th century, and shows up as a tremendous increase in life expectancy: between 1850 and 1990 life expectancy increased from about 40 (both sexes) to 74 (men) and 80 (women) (Mackenbach, 1996; Wolleswinkel-van den Bosch, van Poppel, Looman, & Mackenbach, 1997).

While there is considerable debate about the causes of this increase, and the role of medicine was even doubted (McKeown, 1976), it is now accepted that improved medical technology, preventive and curative, is at least in part responsible (Mackenbach, 1988). Many diseases can now be cured or prevented altogether, but one of the causes of lowered mortality in particular among the elderly is undoubtedly longer survival with previously fatal degenerative diseases (chapters 4 and 6). Many diseases thus increasingly turn from fatal into chronic disease, but the disease process is seldomly stopped: those afflicted may survive, but with a high risk of disability.

This thesis addresses the question of the health of an aging population, in interaction with increasingly available and effective health care technologies. More and more health care technologies are available, improving and extending life, but possibly also extending life with disability. What will be the health care needs of this rapidly growing group of elderly?

The scope of this question is very wide, and this thesis offers only partial answers. Much of the thesis is devoted to the problem how one should go about getting even these partial answers, because they require an interdisciplinary effort with methods from epidemiology, demography and economics.

This general introduction provides the background of the research question and delineates the approach. Several consecutive and distinct steps are needed, which we will describe here in general terms, the detail being relegated to the introductions to Part II to VI. We open by explaining the background of the debate on future morbidity. We show then the paramount importance of a disease specific approach to this question and the need for computer models to ensure consistency and track dynamics. We close with summarizing the main questions that will be addressed.

The compression or expansion of morbidity

Background

The question of the future health of the population, and in particular of the elderly population, is in the international context known as “the compression versus expansion of morbidity”. The debate originated with papers by Gruenberg and Kramer, who pointed out that the present advances in medical technology allow us to save the frail and disabled from dying from complications, and that therefore the prevalence of frailty and disability in the population is bound to rise (Gruenberg, 1977; Kramer, 1980). The “failures of success” would lead to a “pandemic of mental disorders and associated chronic diseases and disabilities”.

Their argument is that when the incidence of chronic disease and disability does not change, but survival improves, the stocks of frail patients will increase. They conclude that research should shift emphasis to discover preventable causes for the conditions that will experience expansion.

The antithesis is compression, and was put forward by Fries (Fries, 1980). The main proposition of Fries is that the length of life is fixed. Fries suggests that there is an ideal rectangular survival caused by a quite narrow distribution of so called “natural death” around age 85, and that the Americans (and, presumably, also other western populations like the Dutch) have covered most of the distance to that ideal. Mortality will not decrease anymore, because the genetic potential of the human species is limited, and we have reached the limits. The second proposition of Fries is that chronic disease can be postponed. He points to the decline of cerebrovascular disease mortality, generally attributed to hypertension control, and the decline of tobacco consumption. If incidence (inflow) decreases, and mortality (outflow) remains constant, stocks of prevalent frail patients will decrease.

The expansion or compression theories, as stated above, are built on opposite propositions: expansion assumes a decreasing mortality but fixed incidence, compression assumes fixed mortality but decreasing incidence. Manton points out that both scenarios in fact “posit that chronic disease morbidity and at least certain components of mortality have no necessary connection (Manton, 1982). He argues that for many chronic conditions improved survival will come about by slowing the rate of progression of the primary disease process. “In this (..) case, duration, and hence life expectancy, is increased by reducing disease severity.” From this he went on to develop the concept of “dynamic equilibrium”: an expansion of chronic conditions, with a simultaneous decrease in their severity.

The observed tendency to disconnect morbidity and mortality has largely persisted in studies that followed. On the one hand there is a lively debate on the human life span and future developments in life expectancy (Roush, 1996). Based on the enormous rise of life expectancy in the past some authors predict further increases all the way to 100 years or even higher. Others point out that such increases in life expectancy would re-

quire unprecedented declines in mortality: to achieve a life expectancy of 100 years mortality at all ages would have to decline by 50% from the 1985 level, and to reach 120 a decline of 90% would be needed (Olshansky, Carnes, & Cassel, 1990). In this debate morbidity is largely ignored, but it is easy to see that a compression of morbidity would be very hard to achieve in the case of rapidly declining mortality: major non-fatal causes of disability, such as loss of vision and hearing and locomotor and congenital disease, would have to decline rapidly as well.

On the other hand many studies have tried to find empirical evidence of expansion or compression, usually operationalized as 'disability', with some finding evidence of expansion, but others of compression, and some nothing much of either (Mathers, Robine, & Wilkins, 1994; Perenboom, Boshuizen, & van de Water, 1993; Robine, 1994; Wilkins, Chen, & Ng, 1994). These studies tend to take life expectancy more or less for granted, even to the extent of looking at relative compression and expansion of morbidity, or at 'healthy life percentage' (van de Water, 1991). Several methodological papers have pointed out that there are numerous difficulties with the most employed method, that is based on cross-sectional prevalence data (chapter 16) (Bebbington, 1992; Brouard & Robine, 1992; Rogers, Rogers, & Belanger, 1990). Unfortunately, the longitudinal data that would resolve these problems are extremely scarce.

Another problem concerns the generic morbidity measure like "disability" that is used in most empirical studies. Several authors have argued that compression and expansion depend on the kind of disease and the kind of change in its epidemiology, and that at least distinction should be made between fatal and non-fatal diseases (Manton, 1982; Olshansky, Carnes, & Cassel, 1990). When medical progress successfully lowers mortality from fatal diseases like cardiovascular disease, but fails to postpone the age at onset of non-fatal diseases like osteoarthritis and vision and hearing impairments, the latter will of course expand.

Indeed, it is striking to note that the original proponents of the expansion and compression theses had different diseases in mind. Gruenberg and Kramer rest much of their argument on the increased survival of patients with Down's syndrome and other congenital mental disorders, while Fries based his declining incidence on smoking- and hypertension-related diseases. The developments have been very different indeed for these diseases: patients with Down's syndrome presently live on to develop heart failure and dementia, while the incidence of cardiovascular disease has, in all likelihood, been declining (see chapters 4 and 6).

Three basic questions

The main research question of this thesis is the one that is central to the compression and expansion debate:

Will there be compression or expansion of morbidity?

The observation that compression and expansion depend on the kind of disease led us to the first of two additional questions that determine the methodological basis of this thesis. When overall population health status is the resultant of diverse developments on the disease specific level, a better understanding of the changes in overall health might be gained by studying the disease specific developments, and then derive the consequences for overall health. So the first additional question is:

What can a disease specific approach contribute to the understanding of compression and expansion at the level of overall population health status?

This question puts the basis of our approach to the expansion and compression discussion in the realm of epidemiology, the study of the occurrence of diseases. Changes in overall morbidity and mortality are explained by disease incidence and survival, prevalence and mortality and their changes over time.

The second additional question follows from the observation that incidence, prevalence and mortality are causally linked, and that changes in disease prevalence, and hence morbidity prevalence, are a dynamic phenomenon. Changes in morbidity prevalence are the result of changes in incidence, survival, and/or cure, but as a stock variable prevalence also depends on past values.

The bathtub provides a useful analogy: the current level of the water (the stock) is determined by how long and how wide the tap has been open, and if and when you pulled the drain. The same holds for disease prevalence: it depends on past inflow (incidence) and outflow (cure and mortality). The study of dynamic changes, namely compression and expansion, asks for dynamic models (Manton, 1982).

In combination with the disease specific approach this leads to a need for dynamic epidemiological models. The second additional question becomes:

How can epidemiology and modelling techniques be integrated into public health models fit to answer questions of compression and expansion on the disease specific level, and on the level of overall population health?

These two additional questions determine the methods examining the compression and expansion question in this thesis, and thus a large part of its contents. In part II several dynamic disease specific models are presented which are used to analyze past changes, and, in some cases, to make projections. Parts III and IV present methods to synthesize the disease specific results: part III describes how multi-disease models can be constructed, and part IV how disease specific prevalences can be integrated to a generic indicator of population health status. Because of the central importance of the two additional questions we will first elaborate on them.

Changes in disease epidemiology

'Morbidity' is a general term, referring to a manifestation of ill health. Its main determinants are the prevalence of diseases and the 'severity' for the individual of having that disease.

The WHO introduced the concept of IDH, "Impairment, Disability and Handicap", to describe the consequences of diseases. An impairment is the damage to a system. For example, an elderly woman with hip arthrosis has an impaired hip. However, this impairment may be a purely radiological finding, causing no symptoms. Disability is the loss of function as a consequence of that impairment. The woman may have pain, walk with difficulty and need a stick. Handicap is the social consequence of that disability. If she can not mount the bus anymore, and can not go shopping in town, the woman is handicapped in that respect.

The distinction between disability and handicap is rather subtle, and serves to remind policy makers that handicaps may be created or avoided by the environment: a low instep bus might avoid the mobility handicap of the example without changing the disability. However, the distinction between impairment and disability is very important. The impairment without disability is not an immediate health problem. Hip arthrosis as a radiological finding without symptoms or function loss might cause morbidity in the future, but not now. The health consequence of interest is the disability a disease causes.

The future health of the Netherlands is largely determined by the evolution of degenerative diseases caused by both life style and senescence. There is no sharp distinction between both etiologies. They are both determined by 'age', which is genetic potential, and 'environment', increasing or decreasing the speed of senescence by promoting or diminishing stochastic damage (Jazwinski, 1996). Alzheimer's Dementia would be the prototype of a mostly senescent disease, with hardly any environmental determinants documented (Breteler, Claus, van Duijn, Launer, & Hofman, 1992; Edelberg & Wei, 1996), lung cancer a life style disease, caused predominantly by smoking or occupational exposure (Davila & Williams, 1993).

The combination of changes in life style variables and in the natural history of specific diseases through medical interventions produces a variety of possible changes in disease epidemiology:

- More effective control of environmental damage will decrease disease incidence. Decreasing incidence of degenerative diseases succeeds more often through 'delaying' than through 'abolishing' disease. An example is smoking; damage induced by smoking probably both causes disorders and advances their onset towards younger ages (Bartecchi, MacKenzie, & Schrier, 1994). The decreasing prevalence of smoking, and the ensuing declining incidence of smoking related disorders have consequences for

disease prevalence that may be hard to assess, as better risk profiles also improve the prognosis of fatal diseases, and lower mortality.

- Effective secondary prevention through stepped-up case finding will both increase incidence and improve prognosis of degenerative disorders. Stepped-up case finding increases incidence because both incidence of the degenerative disease and mortality from all other causes increase by age (chapter 8). Occult disease is uncovered which would have passed unnoticed if the patient had died between the moment of early diagnosis and clinical manifestation (Black & Welch, 1993; Cole & Morrison, 1980). Earlier, milder manifestations and consequences of the targeted disease will increase, but partly avoid later, more severe manifestations (de Koning, van Oortmarssen, van Ineveld, & van der Maas, 1990). Again mortality will decrease. An example is breast cancer screening.
- The effective prevention and/or postponement of recurrent disease after a first clinical manifestation (sometimes called tertiary prevention, a less well defined concept) decreases recurrences, and more severe disease states will be postponed. An example is platelet inhibitor and beta-blocking treatment of coronary heart disease. Survival improves, and again patients live longer, and in milder disease states (chapter 4)
- Successful treatment of potentially fatal events will prevent death, but will not necessarily cure the disease process. The survivors of damaging disease attacks are at risk for further chronic disease. An example is thrombolysis in acute myocardial infarction, causing increases of congestive heart failure and probably stroke patients (chapter 4)
- Revalidation and rehabilitation (another intervention called sometimes tertiary prevention) may prevent long term disability. An example is rehabilitation after acute stroke (Dennis & Langhorne, 1994).
- Treatment or palliation of existing symptoms may improve the well-being, with or without improving the prognosis.

Any specific disease may, depending on etiology and natural history, be amenable to one or more of these processes, resulting in changes in disease specific incidence, prevalence, mortality and/or severity. Several of the processes are generated by health care interventions, the application of which depends, among other things, on technological innovation and policy decisions, and hence are difficult to predict.

The idiosyncrasy of these disease specific developments makes statements about compression or expansion of morbidity not very helpful if they are restricted to the generic morbidity level. Therefore we have chosen to analyze epidemiologic changes on the disease specific level, and only then synthesize results to changes in overall morbidity and mortality. We will use quantitative modelling of diseases as a tool, and show the power of such models both in integrating data from various sources as in demonstrating the effects of dynamic changes.

Models of disease

Meta-synthesis of knowledge and figures

The best way to study expansion or compression is to set up an ideal study, obtaining all the desired figures about determinants, incidence, prevalence and mortality of diseases of interest with the best methodology available. However, even if such an ideal study would be available (the Framingham heart study about cardiovascular disease goes a long way), inference to another population is not straightforward, as many determinants may differ in character and distribution.

But comprehensive information on changes over time of a specific disease epidemiology is scarce. There are many papers about incidence, prevalence, survival, mortality or risk factors, but very rarely about patterns in all of these simultaneously. Theories about incidence increases rarely try to explain why mortality of the same disease and at the same time was decreasing, and vice versa (Adami, Bergstrom, Soren, & Baron, 1993; Bonneux, van Oortmarssen, & Barendregt, 1993; Broderick, Phillips, Whisnant, O'Fallon, & Bergstralh, 1989)(chapters 6 and 8). Important determinants, such as hypertension, are said to be increasing without explaining why an important 'target disease', such as stroke mortality, is decreasing sharply (Oei & Erkelens, 1995). Or papers about observed prevalences do not even begin to explain why other studies about the same prevalences in very similar populations show completely different figures (Bonneux, van de Mheen, Gunning-Schepers, & van der Maas, 1996; Van Leer, Seidel, & Kromhout, 1994; Van Leer, Verschuren, & Kromhout, 1994).

Quantitative disease models help making sense of the many sources of information about disease occurrence by combining knowledge from various sources. Such models describe quantitatively the simplified disease history: incidence, survival and death. Complications, such as heart failure after an acute myocardial infarction, are disease stages with their own incidence, survival and death embedded in the disease history. Depending on the research question, and consequently on the disease history as specified, and the available data, time series of secondary event rates can be generated: recurrence, admission and intervention rates, or prevalence of various disease states.

The main use of the model is to force consistency on the various data and assumptions. It can be fitted to existing data, and the congruity of assumptions checked. This type of models might, as an analogue to the technique of meta-analysis, be called meta-synthetic. They aim to explain observations from the past from many different sources by combining all relevant findings in a comprehensive and quantified theory. Examples of meta-synthetic modelling are the disease models in part II.

Time and age

Time and age are often considered synonymous: we grow older in step with the passage of time. When a population is followed through time, however, we have to deal at any point in time with people of all ages. Describing a population through time therefore requires separate age and time dimensions.

The models used in this thesis always have an age dimension: age is such an important determinant of health that the models would lose much of their relevance without it. Changes in incidence and survival at younger ages affect prevalence and mortality at older ages, and such changes in the age pattern are described in particular for cardiovascular disease. This impact of changes at younger ages on the health status of the older population might be called “age dynamics”.

Not all models in the chapters to come have a time dimension: in many cases some variant of the period life table is used. The period life table, while describing the “age dynamics”, is a static model because it ignores dynamic effects of time, and explores only steady states. A period life table states questions like “*What would be life expectancy if mortality remained constant for at least a century at the level of a certain period*”, or, for a multi-state life table: “*What would be the age specific prevalence if incidence and survival remained constant at a certain level*”. Because the incidence and survival in practice do not stay constant, the steady states described by the life table will never occur in reality: they are ideal types.

For many research questions, where dynamic changes over time are not directly the issue, such models have the attraction of greater simplicity. A good example of the power of a simple static model is to be found in chapter 23, where stationary populations with and without a disease are compared. As a disease is eliminated both as source of costs and as cause of death, the age specific population changes and the ensuing financial consequences of eliminating fatal and non-fatal diseases can be calculated. But the static life table does not describe the pathway from the original steady state with the disease to the alternative steady state without it.

A dynamic model can describe this passage of time. It can illustrate explicitly the changes of disease prevalence over time, driven by the various in- and outflows. A dynamic model is more realistic, but pays for that realism with increased complexity. If this complexity oversteps a threshold, the results becomes hard to interpret and validate. The reader can compare chapter 4, *The future rise of heart failure*, and chapter 5, *The ‘new’ old epidemic of coronary heart disease*. Both handle the question of changing coronary heart disease history, but chapter 4 uses a complex dynamic simulation model with both time and age dimensions and chapter 5 a simplified static multi-state model with only an age dimension.

Some research questions, however, require a dynamic model, because the passage of time determines the answer. In chapter 16 we use the complex dynamic heart disease model to show how cross-sectional prevalences used in calculations of health expectancy may bias a trend analysis. And in

chapter 24 we explicitly compare static and dynamic analysis of the health care costs of smoking, and demonstrate that, while the issues of time and age seem straightforward enough, there remains ample scope for confusion.

Simulation of the unobserved

An important application of models is to provide answers to “what if” questions. These ‘what if’ questions can be distinguished in two kinds: *business as usual* and *business not as usual*. *Business as usual* questions simply extrapolate current trends into the future to assess what this future may look like. Given the population structure in the Netherlands, with the large post war birth cohorts, a population projection will preferably be part of the extrapolation, but other trends, like in disease incidence and survival, may be extrapolated too. Such scenarios can be found in chapters 4 and 6, where disease models are used to project future cardiovascular morbidity and mortality.

Business not as usual questions involve interventions. Simulation of the consequences of specific interventions using a model are “best guesses” if no empirical data are available because either the intervention is relatively new, or follow up or sample size of a study needs to be unreasonably large to yield statistically reliable results. In chapter 9, the question is asked “*What are the consequences for hip fractures, if we can manipulate bone mineral density?*”. Because frail bones are but one of the many underlying causes of hip fractures, increasing bone mineral density can only have but a limited impact. Because menopause and decline of estrogen levels start at around 50, and hip fractures occur predominantly after 80, a hormone replacement therapy trial in the general population would have to run for decades before any effect would become measurable.

Another reason to use simulation may be that the intervention is in fact unrealistic: in chapter 23 whole disease categories are eradicated in a cause elimination life table, and in chapters 20 and 24 smoking is entirely abolished. Such simulations are thought experiments, used, for example, to assess the importance of diseases and risk factors for population health status.

Multi-disease models

A consequence of our decision to analyze epidemiologic changes on the disease specific level is that we somehow must ‘add up’ the disease specific results to arrive at a comprehensive picture of the changes in population health status. This synthesis of disease specific results has two aspects. One is a methodology that allows many disease specific models to coexist and together describe the health of a population. The second aspect concerns the development of a meaningful statistic to summarize and measure this multi-disease description of population health status. This second aspect will be introduced in the next section, in this one we turn to the first.

When multiple disease specific models are used to describe the health of a single population they will have to interact. This interaction can occur in a number of ways. For example, when the number of people dying from one disease goes down, numbers of deaths from other causes will, after some time lag, rise. Another example are diseases that share a common risk factor, such as the group of smoking related diseases. Changes in the risk factor exposure will in that case affect a group of diseases simultaneously, and changes in one disease will affect all other diseases that share the risk factor. A further complication occurs when one disease acts as a risk factor for another disease, for example coronary heart disease, which is a risk factor for congestive heart failure, and diabetes, a risk factor for several other diseases with cardiovascular disease among them.

In fact, diseases may interact in very complicated and idiosyncratic ways. The art of modelling is to describe the interactions that are known well enough and are deemed important, given the research question, and to ignore the unimportant and obscure. Part III describes the methodological and mathematical foundations of the disease models and formulates a set of simplifying conditions that allow these models to interact in a multi-disease environment, taking into account substitution and competition of causes of death, comorbidity, and risk factors, while keeping a reign on model complexity.

From disease specific prevalence and mortality to generic morbidity and a measure of population health status

On the disease specific level expansion and compression of morbidity can be regarded as relatively straightforward: it simply is a question of changes in prevalence of diseases. The disease specific models of part II therefore look at prevalence of disease, as a result of changes in incidence, survival, and demography.

Things get more complicated when severity of disease is taken into account: for example, does less ischemic heart disease but more heart failure constitute a compression of heart disease, an expansion, or perhaps neither? And what about comparisons between entirely different diseases, like cancer and heart disease? And, to complicate matters even further, how does being alive (but ill) compare to being dead?

These complications are the consequence of looking at population health status as the result of both mortality and morbidity, and of morbidity from multiple diseases. When only mortality is used the measurement of population health status is relatively simple. Mortality is unambiguous: either you are alive, or you are dead. Cause specific mortality is more ambiguous, because the underlying cause of death registration assumes that there is but one underlying cause; particularly at older ages, death is often the result of a failure of multiple systems. But accepting the 'one dead, one

cause' rule, mortality from one disease can easily be compared to mortality from another.

But if not just mortality but morbidity too is taken into account, disease needs to be compared both to death and to other diseases. For example, thrombolysis saves many acute myocardial patients from death, but some will pay the price of disability from heart failure (chapter 4). To 'value' the effect of thrombolysis on morbidity, the years lived with heart failure have to be compared to the years lost if no thrombolysis had been administered. To value a successful anti-tobacco policy, a wide range of diseases have to be made comparable, from stroke and coronary heart disease over a whole family of cancers to chronic lung disease and even hip fractures.

Part IV is devoted to this quest for a comprehensive measure of population health status that includes both mortality and morbidity, and is fit to indicate quantitatively compression and expansion of morbidity. We argue, with Murray (Murray, 1996), that *time* is the common denominator that allows comparison between diseases and between morbidity and mortality. Time lived (life expectancy), or not lived (years of life lost), and time lived with disability, and thus not lived healthy.

The comprehensive measure of population health status that is still the most used is life expectancy, which expresses average time lived, but since this measure does not include morbidity it obviously cannot address the compression and expansion question. A group of researchers therefore propose Healthy Life Expectancy (or Health Expectancy for short), which stands for a whole family of indicators (Mathers, Robine, & Wilkins, 1994). The predominant approach is to use the prevalence of generic morbidity, such as disability, measured through a cross-sectional survey, and combine it with current mortality to obtain a Health Expectancy by ignoring the years lived with morbidity by the life table cohort (Sullivan, 1971).

Seen from the perspective of our approach this implementation of Health Expectancy raises two problems. The first is to obtain a generic morbidity, starting from disease specific prevalences: the relation between the two is not at all obvious. The second problem concerns the years lived with morbidity: by ignoring these years morbidity is implicitly being equated with mortality, and this leaves hardly any room to deal with severity of morbidity. As is being shown in chapter 17 this may result in counter-intuitive behaviour of the indicator.

Both problems were addressed by the work of Murray and colleagues on the Disability Adjusted Life Years, or DALY (Murray, 1996). The DALY is based on a disease specific approach, and uses disease- and stage-specific disability weights to map diseases to disability: time lived with disease is weighted for the severity of the disease, with weights between 0 (equivalent to full health) and 1 (equivalent to death).

Unfortunately, as we explain in the introduction to part IV, because the DALY includes changes in the population structure, for example the large post World War II baby boom cohorts, it is not very suited for an-

swering questions of compression and expansion that originate from changes in epidemiology. Moreover, the DALY is geared for cost-effectiveness studies, and includes discounting of future life years, and an age-weighting schedule which may produce counter-intuitive results (see chapter 18).

In chapter 19 we therefore propose to combine elements of the Health Expectancy and DALY indicators in the Disability Adjusted Life Expectancy, or DALE. The DALE is a life expectancy, and therefore abstracts from population structure, and includes a disease specific approach with disability weights developed for the DALY. This setup allows to assess the impact on population health status of changes in disease epidemiology, including changes in disease severity.

Given the DALE approach, the question of compression or expansion of morbidity can be operationalized as '*Under which conditions will the life expectancy lived with disability, weighted for severity, increase or decrease?*'. This question is identical to: '*Under which conditions will the difference between life expectancy and disability adjusted life expectancy increase or decrease?*' Population years lived in a prevalent disease state are obtained from incidence-prevalence-mortality disease models nested in a life table approach. These years are weighted for severity of the specified disease state. Combined with a life expectancy they yield a DALE. Overall morbidity is now defined as the difference between life expectancy, and disability adjusted life expectancy: these are the years lost by disability, weighted for severity. If this difference increases, morbidity expands, if this difference decreases, morbidity compresses.

Compression, expansion, and health care costs

In the debate on compression and expansion of morbidity the relation with health care costs has never been far away. Indeed, one could argue that concerns about health care costs fuel much of the debate, with proponents of compression claiming that prevention will decrease health care costs (Fries, et al., 1993), while an expansion would, in all likelihood, cause a relentless rise in costs (chapter 23).

This makes sense: diseases and disability are determinants of health care costs. To investigate the matter we can map, as with disability, health care costs generated by a specific disease to incidence-prevalence-mortality models, nested in a life table approach. The research question of compression or expansion of morbidity is then rephrased as a question of compression or expansion of health care costs, determined by age, sex and diagnosis. However, as we will argue more explicitly in the general discussion (chapter 27), the readers of this thesis must be warned that these parallel research questions do not imply parallel answers. Indeed, the historical demographic and epidemiological transition was typically accompanied by compression of mortality, and very likely by compression of disability, but expansion of health care costs (Barendregt, Nusselder, & Bonneux, 1997;

Murray & Lopez, 1997). Indeed, if we take the large burden of morbidity and mortality from maternal causes in developing countries as an example, the role of health care is obvious. Answering this large health care need will ask for expansion of health care services as cause considerable compression of morbidity and mortality.

As with disability the effect on compression and expansion of costs depends on the changes in disease specific epidemiology. In part V two chapters are devoted to the impact of prevention on costs. In chapter 23, using a cause elimination life table, large categories of diseases are eliminated as causes of costs and of death. The costs of added life years after a fatal disease is eliminated are often higher than the health care costs of the disease before elimination. Only eliminating non-fatal diseases may yield health care savings (ignoring the costs of the prevention program). Chapter 24 looks at the prevention of smoking and the impact of smoking on health care costs. In this chapter we show that smokers are more expensive to health care when alive, that a non-smoking population generates higher costs, but that nevertheless, given suitable assumptions, abolishment of smoking would be financially beneficial.

The evolution of senescence

This general introduction has so far outlined the approach to the question of compression and expansion taken in this thesis: an analysis on the disease specific level, followed by a synthesis to population health status outcome, both in terms of health and health care costs. From there we take one more step: a reflection on the underlying causes of morbidity and mortality and the possible consequences for compression and expansion.

Part VI is devoted to the ultimate cause of age-related morbidity and mortality. Why is aging synonymous to an increasing probability of becoming disabled and dying? Why does frailty increase with age? Why do mortality rates increase with age? Why, in other words, does senescence exist?

Epidemiologists and demographers, let alone economists, may be tempted to dismiss such questions as ontological, or even metaphysical. But the answers to the questions of senescent mortality and morbidity are to be found in the characteristics of senescence. If life span is fixed, as Fries and many others claim, how is it fixed and at what age? Fries argues for around age 85, but others (and better demographers), using basically the same mortality data as Fries, find that there is no evidence at all of a 'ceiling' of life expectancy (Manton, 1982). If our death is genetically 'programmed', 'deprogramming' might be possible, doubling or quadrupling our life span. But if mortality decreases further, what will be the consequences for senescent morbidity? Will mortality be postponed, but not morbidity, giving rise to Kramers' pandemic (Kramer, 1980)? Or will we age 'successfully'?

Epidemiology, an empirical science based on observations among humans, is hardly suited to answer such theoretical speculations, as our life span is long and our actual life expectancy is unprecedented in human history. Evolutionary biologists, however, do ask these questions about the nature of senescence. Senescence is observable in most complex animal and plant species kept in sheltered conditions. The first consequence is that it is highly improbable that the reasons of such an universal phenomenon should be haphazard or coincidental. The second consequence is that questions about senescence may be tested on other species, with a much shorter life span than *homo sapiens*.

The upshot of these modern evolutionary theories of senescence is that after the age of reproduction is reached, energy consuming maintenance systems are neglected, in favour of that reproduction. The ensuing degradation is more or less equally timed by the same life history, the causes of senescence being omnipresent in genes regulating these maintenance systems. Inferences relevant for future population health are many:

1. A population life expectancy beyond 85 year is hardly likely for the *foreseeable* future, as natural selection had no effect whatsoever on the phenotype beyond that age.
2. Genetic manipulation of senescence is unlikely, at least for the foreseeable future, as the numbers of genes and gene interactions involved should be extremely high.
3. Senescent morbidity and mortality are related processes: death and disease are both 'natural'.
4. But slowing down the decay of maintenance might be possible. Those who wish to 'compress morbidity' are then well advised to focus epidemiological research on non-fatal diseases, the major determinants of morbidity in low mortality countries.

Summary and research questions

The central matter of interest in this thesis is the future of population health status in the Netherlands. This is operationalized as the main research question: "Will there be compression or expansion of morbidity?"

In order to answer this question we formulated two additional research questions: "What can a disease specific approach contribute to the understanding of compression and expansion at the level of overall population health status?", and: "How can epidemiology and modelling techniques be integrated into public health models fit to answer questions of compression and expansion on the disease specific level, and on the level of overall population health?"

From the two basic questions follows the road taken: part II presents a number of disease specific models, used to describe and understand disease epidemiology and the changes therein. In part III a methodology is developed to integrate disease specific models to a comprehensive Public Health model. Part IV describes how to summarize disease specific preva-

lences and mortality to an indicator of population health status, permitting an answer to the main research question in terms of health.

Consequences for health care costs of preventive interventions, often advocated as the way to compression, are shown in part V. Part VI presents an entirely different outlook at the same problem of morbidity and mortality, searching for an ultimate cause of age related degenerative disease in evolution theory. It formulates expectations about future mortality and (senescent) morbidity based on the *disposable soma* theory. And finally, in part VII, we discuss our findings and try to provide an answer to the three research questions.

The remainder of part I is formed by chapter 2, *the dirty hands of the epidemiologist*. In this chapter we argue that the use of models in epidemiology is a scientific endeavour, part of epidemiology. Epidemiology has to do with 'dirty' data, and has to make the best of it. Epidemiology tries to answer answerable questions about health and disease, with the aim of improving the health of populations. Models may help, by clarifying and interpreting data in a consistent context.

Obviously, as soon as we venture out from the islands of solid facts, we may make mistakes. But, as Murray and Lopez suggest, epidemiology needs to take some lessons of 'demographers ethos' (Murray & Lopez, 1996). Even if the facts are very scarce, demographers use everything that is available and still construct a 'most plausible' population and a 'most likely' population forecast. The only alternative left to public health epidemiologists is to refuse to dirty their hands, to leave the public health policy makers on their own and to complain about the wrong decisions decades after they have been taken.

Our readers should keep in mind that we borrowed heavily not only from the demographers' tools, but also from their ethos.

References

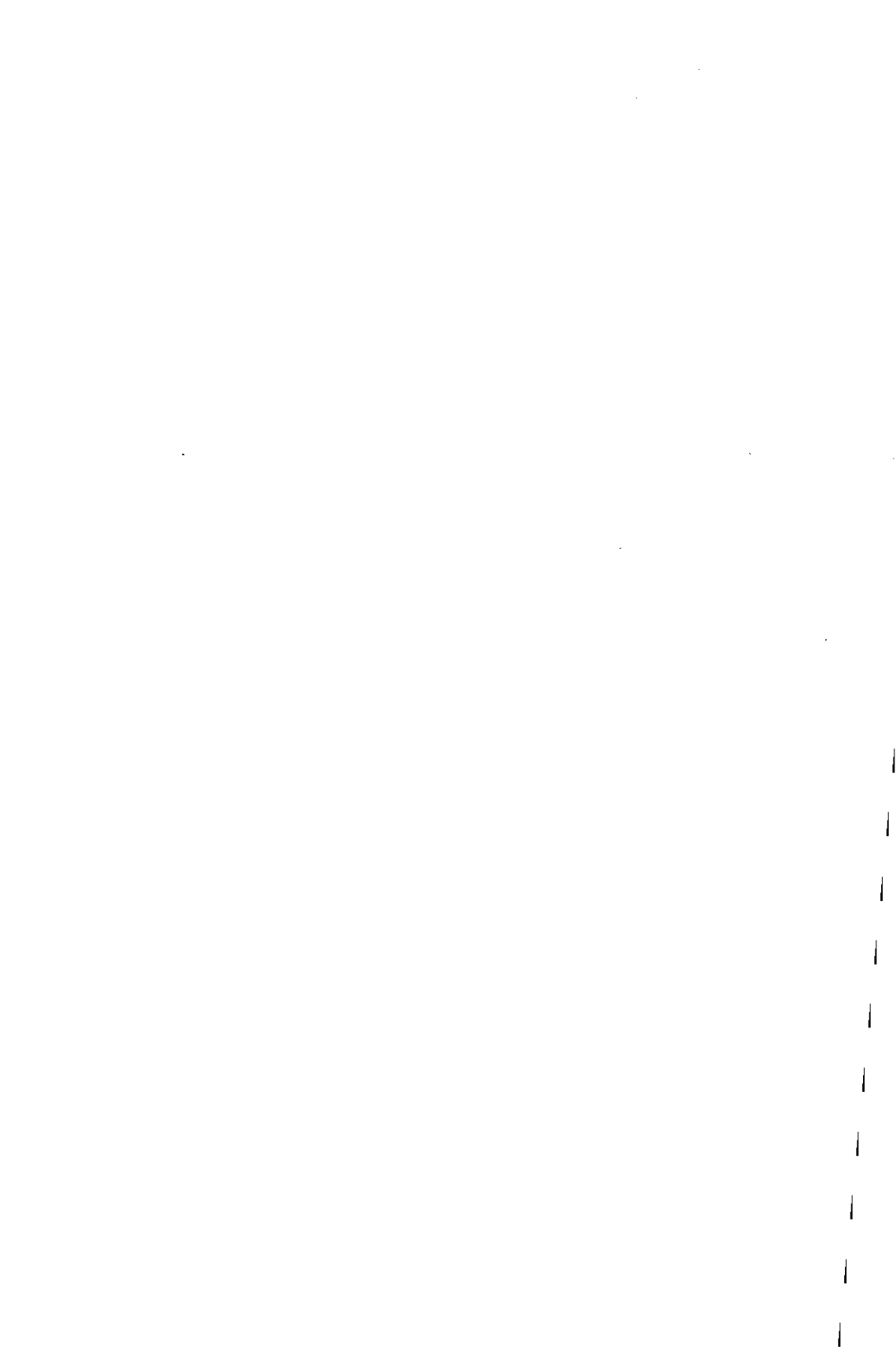
- Adami, H. O., Bergstrom, R., Soren, P., & Baron, J. (1993). Increasing cancer risks in younger birth cohorts in Sweden. *Lancet*, 341, 773-77.
- Barendregt, J. J., Nusselder, W. J., & Bonneux, L. (1997). Global Burden of Disease [letter]. *Lancet*, 350, 142-143.
- Bartecchi, C. E., MacKenzie, T. D., & Schrier, R. W. (1994). The human costs of tobacco use (1) [see comments]. *N Engl J Med*, 330(13), 907-12.
- Bebbington, A. C. (1992). Expectation of life without disability measured from the OPCS disability surveys. In J.-M. Robine, M. Blanchet, & E. J. Dowd (Eds.), *Health Expectancy* (pp. 23-34). London: HMSO.
- Black, W. C., & Welch, H. G. (1993). Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med*, 328(17), 1237-43.
- Bonneux, L., & Barendregt, J. (1991). *Towards a comprehensive health policy model*. Paper presented at the Heidelberg Proceedings.

- Bonneux, L., van Oortmarssen G.J., & Barendregt, J. J. (1993). Increase in cancer incidence in younger birth cohorts? *Lancet*, 341, 1409.
- Bonneux, L., van de Mheen, P., Gunning-Schepers, L. J., & van der Maas, P. J., (1996). Daalde hypertensie in Nederland tussen 1974 en 1993? *Ned Tijdschr Geneesk*, 140(52), 2603-2606.
- Breteler, M. M., Claus, J. J., van Duijn, C. M., Launer, L. J., & Hofman, A. (1992). Epidemiology of Alzheimer's disease. *Epidemiol Rev.* , 14, 59-82.
- Broderick, J. P., Phillips, S. J., Whisnant, J. P., O'Fallon, W. M., & Bergstralh, E. J. (1989). Incidence rates of stroke in the eighties: the end of the decline in stroke? *Stroke.* , 20(5), 577-82.
- Brouard, N., & Robine, J.-M. (1992). A method for calculation of health expectancy applied to longitudinal surveys of the elderly in France. In J.-M. Robine, M. Blanchet, & E. J. Dowd (Ed.), *Health Expectancy*(pp. 87-97). London: HMSO.
- Cole, P. H., & Morrison, A. S. (1980). Basic issues in population screening for cancer. *J Natl Canc Inst*, 64, 1263-72.
- Davila, D. G., & Williams, D. E. (1993). The etiology of lung cancer. *Mayo Clin Proc*, 68(2), 170-82.
- de Koning, H. J., van Oortmarssen, G. J., van Ineveld, B. J., & van der Maas, P. J. (1990). Breast cancer screening: its impact on clinical medicine. *Br J Cancer*, 61, 292-7.
- Dennis, M., & Langhorne, P. (1994). So stroke units save lives: where do we go from here? [see comments]. *Bmj*, 309(6964), 1273-7.
- Edelberg, H. K., & Wei, J. Y. (1996). The biology of Alzheimer's disease. *Mech Ageing Dev*, 91(2), 95-114.
- Fries, J., Koop, C., Beadle, C., Cooper, P., England, M., Greaves, R., Sokolov, J., Wright, D., & Consortium, t. H. P. (1993). Reducing health care costs by reducing the need and demand for medical services. *N Engl J Med*, 329, 321-5.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *N Engl J Med*, 303(3), 130-5.
- Gruenberg, E. M. (1977). The failure of succes. *Milbank Q/Health Society*, 55, 3-24.
- Jazwinski, S. M. (1996). Longevity, genes and aging. *Science*, 273(5 july), 54-9.
- Kramer, M. (1980). The rising pandemic of mental disorders and associated chronic conditions and disabilities. *Acta Psychiatr Scand*, 62 (suppl 285), 282-97.
- Van der Maas, P. J., & Habbema, J. D. F. (1986). Modelontwikkeling en informatie: twee voorwaarden voor volksgezondheidsbeleid. In E. Dekker & B. Wijnberg (Ed.), *Gezondheidsbeleid over de grenzen van de gezondheidszorg.* (pp. 112-27). Den Haag: Nota 2000.
- Mackenbach, J. P. (1988). *Mortality and medical care.* Unpublished Ph.D. thesis, Erasmus University Rotterdam.
- Mackenbach, J. P. (1996). The contribution of medical care to mortality decline: McKeown revisited. *J Clin Epidemiol*, 49, 1207-1213.
- Manton, K. G. (1982). Changing concepts of morbidity and mortality in the elderly population. *Milbank Q/Health Society*, 60, 183-244.

- Mathers, C. D., Robine, J. M., & Wilkins, R. (1994). *Health expectancy indicators: recommendations for terminology*. Paper presented at the Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.
- McKeown, T. (1976). *The role of medicine: dream, mirage of nemesis*. London: Nuffield provincial Hospitals Trust.
- Murray, C. J. L. (1996). Rethinking DALY's. In C. J. L. Murray & A. D. Lopez (Ed.), *The Global Burden of Disease*(Vol. 1, pp. 1-99). Harvard: Harvard University Press.
- Murray, C. J. L., & Lopez, A. D. (1996). Global and regional descriptive epidemiology of disability: incidence, prevalence, health expectancies and years lived with disability. In C. J. L. Murray & A. D. Lopez (Ed.), *The global burden of disease*(Vol. Chapter 4, pp. 201-246). Harvard: Harvard University Press.
- Murray, C. J. L., & Lopez, A. D. (1997). Regional patterns of disability free life expectancy and disability adjusted life expectancy. The global burden of disease study. *Lancet*, 349, 1347-52.
- Oei, L. T., & Erkelens, D. W. (1995). Daling in sterfte door coronaire hartziekten in de periode 1974-1992 grotendeels verklaarbaar door veranderingen in de risicofactoren cholesterol en rookgedrag. *Ned Tijdschr Geneeskd*, 139, 2309-14.
- Olshansky, S., Carnes, B., & Cassel, C. (1990). In search of Methuselah: estimating the upper limits to human longevity. *Science*, 250, 634-640.
- Perenboom, R. J. M., Boshuizen, H. C., & Water, H. P. A. (1993). Trends in health expectancies in the Netherlands 1981-1990. In J. M. Robine, C. D. Mathers, M. R. Bone, & I. Romieu (Ed.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*(pp. 309-320). London: John Libbey.
- Robine, J. M. (1994). *Disability-free life expectancy trends in France: 1981-1991, international comparison*. Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.
- Rogers, A., Rogers, R. G., & Belanger, A. (1990). Longer life but worse health? Measurement and dynamics. *Gerontologist*, 30(5), 640-9.
- Roush, W. (1996). Live long and prosper? *Science*, 273, 42-46.
- Sullivan, D. F. (1971). A single index of mortality and morbidity. *HSMHA Health Rep*, 86(4), 347-54.
- Van Leer, E. M., Seidel, J. C., & Kromhout, D. (1994). Levels and trends in blood pressure and prevalence and treatment of hypertension in the Netherlands, 1987-1991. *Am J Prev Med*, 10, 194-99.
- Van Leer, E. M., Verschuren, W. M. M., & Kromhout, D. (1994). Trends in blood pressure and prevalence and treatment of hypertension in young adults in the Netherlands, 1974-1986. *Eur J Epid*, 10, 151-8.
- Van de Water, H. P. A., (1991). *Health expectancy and changes over time: compression or pandemic?* The fourth meeting of the international network on health expectancy (REVES), Leiden (The Netherlands).
- Wilkins, R., Chen, J., & Ng, E. (1994). *Changes in health expectancy in Canada from 1986-1991*. Advances in health expectancies: proceedings of the 7th

meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.

Wolleswinkel-van den Bosch, J. H., van Poppel, F. W. A., Looman, C. W. N., & Mackenbach, J. P. (1997). Cause-specific mortality trends in The Netherlands: 1875-1992 An analysis of the epidemiological transition. *International Journal of Epidemiology*, 26, 772-81.



2

The dirty hands of the epidemiologist

Abstract

The main object of epidemiology is to provide answers to practical questions in public health and medicine: 'Does factor A elevate or lower the risk of disease B?', 'Is treatment A better than treatment B?'. Designed as a (natural) experiment in selected study populations, the tested hypothesis is not immediately applicable. Public health models therefore take (quantified) theories and apply these to a (simulated) population with due observance of other risk factors, other diseases and other causes of death. The questions they are endeavoring to answer are: 'What will happen if we decrease factor A in the population?', 'What will happen if we introduce treatment A in a population with disease C?'



A good epidemiologist has dirty hands, but a clean mind.

Geoffrey Rose

Epidemiological research studies the health of people with the explicit aim of improving this in the future. In epidemiology, the research question can always be reduced to a carefully specified sub question from the sets of '*How can we treat future patients better?*' or '*How can we enable future populations to live longer and healthier?*'. The sub questions tend to be very practical: 'Is it plausible that treatment A is better than treatment B?' 'Is it plausible that factor A yields a higher or lower risk of disease B?'¹

Tested epidemiological hypotheses yield relative estimates and can not be applied directly for policy purposes. For example, well-founded research

has demonstrated that in a female population aged 50-69 screened with mammography, breast cancer mortality is a quarter lower than in the control population.² This provides information about the measurement of a single effect (death) in a study population. Policymakers, however, are interested in the actual population, in absolute effects (the number of added life years) and in eventual undesirable side effects of the program. The answer will therefore be influenced by the incidence of breast cancer and mortality in that population, by classification into stages in diagnosing the unscreened population (the worse, the greater the effect), by the expected degree of participation, etc. They are naturally also interested in the resources they will need to reserve: people, and money.

Operational research and scenario models take epidemiological theories and apply these to a 'realistic' - albeit simulated - population; the research questions become '*What if we treat patients with therapy A?*' and '*What happens if we perform health intervention B (for example breast cancer screening) in the (Dutch) population?*'.

This assumes the application of epidemiological knowledge in practice. Various Dutch epidemiologists, following in the footsteps of Miettinen, reject such applications of epidemiological theory in the practice of public health.^{3,4} We are defending the proposition that application constitutes an integral part of epidemiology, and that without practical applicability, epidemiology becomes an activity without meaning and hence without a future.

The scientific process

In scientific thinking about reality, a distinction can be made between initial conditions (exposure in epidemiology), the causal relationship (the hypothesis) and a result (the effect).⁵ From two elements, the third follows. The search for the causal relationship is the 'building of the theory': an underlying law is assumed based on relations between the initial conditions and the observed results. A typical statement would be: a causal relationship exists between smoking and lung cancer. In a 'prediction', the theory is applied: the result is predicted based on the initial conditions and the theory. A typical statement is the following: if smoking behavior changes, a change in the incidence of lung cancer may be expected. In an 'explanation' the hypothesis is given and with an explanation of the result being sought on the basis of the initial conditions. The statement in this case would be: the observed rise in lung cancer mortality in women may be attributed to the increase in smoking by women in the past. 'Testing' involves questioning the theory: if I have a new set of initial conditions and results, does my causal relationship still hold up?

Making connections, inferring causality, influencing supposed causal processes and subsequently the inferred causality in the light of the induced changes are all one and the same research process. A problem is posed, an answer is subsequently sought. To return to the lung cancer problem: at the

end of the forties, a strong rise in lung cancer mortality was observed. The causal relation to the increase in smoking habits was satisfactorily demonstrated. Anti-smoking campaigns were initiated during the sixties. Young men smoked less and the prediction that deaths from lung cancer would decrease in these cohorts were confirmed. Thus epidemiology delivered a practical answer to a practical question, from which many men are now reaping the benefits.

Good research starts with a research question: What is the problem? What question am I seeking to answer? This leads to the study design in which the various hypotheses can be tested. Given the answer found, the problem can be approached. Application of a theory, whether as a prediction or policy advice, is inseparable from the development thereof. Seeking answers to questions which do not present themselves seems to us to be a senseless activity for an epidemiologist.

Another practical example is that of the randomized controlled study of breast cancer screening. The problem is: 'Breast cancer mortality is the largest cause of death from cancer in women. Can this be reduced?'. The research question is: 'Does mammography screening reduce breast cancer mortality?'. The policy question is: 'Should screening be implemented or not?'. If it had been impossible for screening to be carried out because, for example, mammography had proven exorbitantly expensive, the bottom would have fallen out of the study: the problem can no longer be answered. These studies then become senseless experiments on over a quarter million women.

Dangerous models or dangerous theories?

Quantitative models, such as, for example, those used in evaluating breast cancer screening⁶, are nothing more than a systematized theory, containing well-tested hypotheses which are to be quantitatively specified and which derive from all kinds of medical and epidemiological research. In view of the plethora of relations and their interrelatedness, nowadays they are entered in a computer, as they can then be calculated far more rapidly. The consequences of interventions, as calculated by the model are subsequently compared with the observed reality, and any necessary adjustments are made to the model.

More complex hypotheses in the form of mathematical models entail a further specification of qualitative theses. I.e. the qualitative statement that 'smoking causes lung cancer' is easily knocked down by the truthful observation that the large majority of smokers does not have lung cancer, nor will ever develop this. This statement becomes more correct when quantified: 'If you smoke, you have a ten times higher chance of dying of lung cancer than if you don't smoke.' Note the inferred causality, the hypothetical model (a simple risk function) and the prediction. Expand this to include more and more stratified risk factors and more causes of death and carefully specified time spans between changes in risk factors and the cause

of death, and the computer model PREVENT is obtained⁷. PREVENT attempts to systematize epidemiological knowledge about risk factors, such that the results are useful to policymakers. To this end, the program combines the prevalence rates measured in the Netherlands (hypertension, smoking behavior, cholesterol levels, etc.) with the relative risks according to the medical literature for all relevant disorders, the expected time lapse between a change in risk and effect, and interventions to be specified by the user in these risk factors. The sole difference between the statement 'Smoking causes cancer' and PREVENT is one of refinement: both make predictions about the incidence of cancer based on causally assumed, empirically measured relations. But PREVENT does more: it provides information about the time span in which these changes are expected. To measure the effects of a changed smoking behavior, at least 20 years are needed. The program offers information about other disorders caused by smoking: cardiovascular disease and chronic lung disease. It demonstrates the overwhelming effect on the aging of the Dutch population on absolute figures, despite the expected age-specific decrease. The ultimate goal is to show policymakers what will *probably* happen if the smoking behavior of a population is altered. With that, PREVENT answers the original problem formulated in the relevant epidemiological research: 'What can we do about the rising lung cancer mortality?'

It is extremely unlikely that PREVENT's predictions are quantitatively all 'correct'. Like all hypotheses, they are partly based on tiny stepping stones of knowledge and partly on giant steps of inferences and extrapolations across oceans of ignorance (with thanks to Geoffrey Rose for the imagery (note)). Hence hypotheses and inherent predictions about the future must be interpreted with caution. Often, an awareness of the fact that computer models are no more than series of specified, more or less hypothetical relations is lacking. The often-used statement that *models offer simplified representations* is clumsily formulated; we propose amending the statement to *models are a representation of a limited number of opinions about reality*.

Yet this limited number of opinions is nevertheless more extensive than can be overseen by the human mind all at once. It is extremely difficult to estimate contradictory consequences. To give another example: breast cancer screening has both desired and undesired effects. On the credit side is the greater number of women who live longer and the smaller chance of severe terminal cancer as a result of the reduction in breast cancer mortality. In the debit column are the funds and people to be deployed, as well as the extra induced morbidity due to the anxiety aroused, early diagnosis and false positive results. The expected effects cannot all be overseen at once. The breast cancer screening model of the Department of Public Health (MGZ) carefully disentangles the advantages from the disadvantages, the cardinal hypothesis being that screening with mammography promotes early diagnosis and improves the prognosis. The model does not 'invent', it assumes this tested supposition from carefully controlled studies. If, as a competent scientist, one should disagree with this, then the model is

discarded. In other words: one rejects the underlying hypotheses. Just as every other hypothesis is discarded with which one should disagree. As a result, one then belongs to a scientific minority, which is fine: there can be no scientific discussion without scientific minorities. But one does not evade responsibility: the price of an error is just as high. Advocates of screening may expose whole populations to fear, anxiety and useless treatments, opponents perhaps cause unnecessary suffering and an early, preventable death. The choice is a thankless one, but unavoidable.

Epidemiologists cannot nor should make this choice; that is the task of those responsible for policymaking. But they *must* respond to the question posed to them, to the problem submitted to their consideration. Policy which is not based on the best possible estimation of the consequences of the choice to be made is irresponsible. The fact that it is a common occurrence is no excuse: mere frequency does not make something automatically right.

Conclusion

Like all hypotheses, models can and will make mistakes. The history of medicine is studded with errors of this kind, some funny, some downright appalling. Yet by building hypotheses, applying these and testing them against the forecasts generated, by adjusting them accordingly, (medical) science can take a tiny step forward. Strict proof of whether or not these are correct will never be obtained: 'If you insist on strict proof (or strict disproof) in the empirical sciences, you will never benefit from experience, and you will never learn from it how wrong you are'⁸. Medicine is an unbelievably slippery field, and the task of epidemiology is to scatter a few pebbles in an effort to reduce the number of people who break their necks.

In this image, the models should be seen as organized attempts to lay a path from the pebbles scattered throughout, a path which can be trodden by doctors, policymakers and patients. It is the nature of the epidemiologist's job which will cause him to get his hands dirty in carrying out these activities. To quote that wise man, Geoffrey Rose, again: "A good epidemiologist has dirty hands, but a clean mind." Epidemiologists work with incomplete, distorted, 'unclean' data which will yield up a part of their underlying secrets only when placed in a clarifying context.

A clinical physician is a pragmatist; based on his (insufficient) knowledge he attempts to address the needs of the individual patient as best he can. An epidemiologist is likewise a pragmatist. Based on his (equally insufficient) knowledge he attempts to address public health care needs as best he can. Attempts to redefine epidemiology as a theoretical science makes no allowance for the problem presented: how to reduce human suffering caused by disease and premature death. Separating the testing of hypotheses in empirical models from applying these hypotheses in simulation studies in no way advances the quest for the right answers and should therefore be rejected as being counterproductive.

Note

Geoffrey Rose was the (charismatic) head of the Epidemiology department at the *London School of Hygiene and Tropical Medicine*. The quotes were taken from his opening lecture of the M.Sc. in Epidemiology training (1987).

References

1. Feinstein AR. Unlearned lessons from clinical trials: a duality of outlook. *J Clin Epidemiol* 1989; 42:497-8.
2. Nyström L, Rutqvist LE, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993; 341:973-8.
3. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons, 1985.
4. Van Limbeek J, van Alem VCM, Rengelink H. Epidemiologie en beleid. *Tijdschr Soc Gezondheidsz* 1992;70:63-7.
5. Popper KR. The poverty of historicism. London: Routledge & Kegan, 1961.
6. Van Oortmarssen GJ, Habbema JDF, van der Maas PJ, et al. A model for breast cancer screening. *Cancer* 1990;66:1601-12.
7. Gunning-Schepers, LJ. The health benefits of prevention: a simulation approach. Amsterdam:Elsevier, 1989.
8. Popper KR. The logic of scientific discovery. London: Hutchinson, 1959.

PART II

Disease models

Introduction

In the general introduction, we observed that compression or expansion of morbidity depends on the consequences of dynamic changes and the type of diseases subject to these changes. This led us to two questions: why is a disease specific approach to the 'generic' question of compression or expansion of morbidity useful, and how can public health models, integrating epidemiology and modeling techniques, contribute to a better understanding of these changes in morbidity?

The chapters of part II demonstrate some of the answers to these questions, and the new problems which arise. The explanatory variables are changes in incidence and mortality of a specific (group of) disease(s), the results are changes in prevalence of that disease. The examples show changes in the two major causes of death (cardiovascular diseases and cancer) and one cause of disability (hip fractures). The last chapter of part II (chapter 10) anticipates part IV, and shows how morbidity and mortality of a specific disease can be integrated in a single generic indicator of health status. The subject is dementia, which is not only a leading cause of morbidity and health care costs in the Netherlands (Polder, Meering, Koopmanschap, Bonneux, & van der Maas, 1997), but which is also associated with considerable excess mortality (also shown in chapter 10).

Synthesizing knowledge

The first step is to identify the changes in the main event rates, incidence and mortality. Time series of disease specific mortality can be abstracted from vital statistics. Because we are interested in *changes*, the main assumption needed is not that the Dutch cause of death register is absolutely reliable (which we know it is not (Mackenbach, van Duyne, & Kelson, 1987)), but that variations in registering and coding practices are relatively constant over time. This assumption holds for younger ages. At older ages, where a single cause of death is hard to identify, changes in the

'underlying cause of death' are more easily biased by changing registration and codification practices of uncertain causes of death (Doll & Peto, 1981).

Time series of disease incidence are harder to come by. For the most frequent causes of death, acute myocardial infarctions and stroke, no incidence registers exist, and we have to rely on health care demand dependent hospital registers. When incidence registers exist, as for cancer, incidence figures are likewise dependent on changing health care technology and increasing patient awareness (Doll & Peto, 1981). To be registered, any patient has to consult the health care system, a potential cancer diagnosis has to be suspected, and the diagnostic technology available at that time has to be able to confirm the diagnosis. However, keeping these drawbacks in mind, health care registers are often the only source of epidemiological information over longer periods of time and/or larger populations. Ignoring such an important source, because it is less than perfect, violates one of the central propositions of public health epidemiology: *approximate estimates are better than no estimates* (Murray & Lopez, 1996).

A true incidence, the first manifestation of a disease, is rarely registered as such. We had to reconstruct changes in incidence retrospectively by using various manifestations of the disease and a model of the disease history. This technique, called 'backcalculation', has been used before to calculate unobserved HIV infections from the incidence of AIDS and assumptions of the incubation time between infection and disease (Rosenberg, Biggar, Goedert, & Gail, 1991). We used predominantly attack rates (the sum of fatal and non-fatal disease events), taken from various health care registers, and calculated an incidence which fitted these events through probabilities of recurrence and death taken from the literature. The obtained incidence rates are hypothetical, but reflect the available time series and figures from the epidemiological literature. As prevalence of a chronic degenerative disease is determined by incidence and death, the prevalence of patients with a history of a clinical manifestation of the studied disease can be calculated by the same backcalculating model. If time series of health care use are fitted to time series of cause specific mortality, the resulting prevalence expands or compresses, depending on the changing in- and outflow caused by the time dependent event series. Examples of such backcalculating incidence/prevalence/mortality models are in chapters 4, 5, 6 and 8. In chapter 4 to 6, prevalence and numbers of patients with heart disease or stroke are the main outcome of interest, while in chapter 8 a range of possible colorectal cancer survival distributions is backcalculated, consistent with admission and mortality trends.

If incidence of a specific disease is available (or can be 'backcalculated'), and the relation between that incidence and potential risk factors is known, changes in incidence can be projected, based on changing risk factor distributions. In this thesis, we present no risk factor/event models based on historical trends in risk factors. The major reason is lack of data: while there is much written about risk factor prevalences, the

available figures about important risk factors such as cholesterol and hypertension are amazingly scarce and remarkably inconsistent (Bonneux, van de Mheen, Gunning-Schepers, & van der Maas, 1996; van de Mheen, Bonneux, & Gunning-Schepers, 1995). "Lack of data" is even true for smoking: to model the relationship between smoking and cancer incidence, the lag time between exposure and disease incidence is potentially that long that at least half a century of detailed information about smoking behavior is needed (Kunst, Looman, & Mackenbach, 1993). However, we used an extremely simple 'prospective' model in chapter 9, where a lifetable-like model is used to adjust age dependent hip fracture rates for (age dependent) bone mineral density. This allows to calculate expected changes in hip fracture rates, 'if' bone loss at a certain (menopausal and postmenopausal) age is stopped or slowed.

Chapter 7, based on statistical models, is a mix of both 'projective' and 'backcalculating' concepts. It shows how the sharply declining coronary heart disease mortality rates (without concomitant changes in admission rates) precedes the trend rupture in stroke admission and mortality rates: the expanding vascular risk, inferred from the decreasing mortality, increases the incidence of competing vascular diseases such as stroke and heart failure.

If there has been no or very little evidence of changes in incidence and mortality, such as in dementia, calculation of changes in prevalence is not very relevant. However, if changes are expected, a single indicator of lost health, combining the life years lost to severe disease and to death is of great interest to health policy. Chapter 10, using dementia as an example, anticipates methods in part IV and describes a method for quantifying the burden of disease, integrating both prevalence and death in a single meaningful indicator. Pharmaceuticals which are able to slow the progression of dementia without too many unpleasant side effects are under development now, and are expected to appear on the drug market soon. In future work, we will use the methods elaborated in chapter 10 to gauge the effects on the burden of disease of dementia of such treatments.

Tracking changes in disease epidemiology

Part II has two aims. The first we discussed in the previous section: how disease models can integrate epidemiological knowledge of various sources. In this section we show how such epidemiological models can track the consequences of changes in event rates for the burden of morbidity of a specific disease. Chapters 4 to 7 describe the most important cardiovascular diseases, namely coronary heart disease, heart failure and stroke. Even within cardiovascular diseases, strongly related to the same type of vascular atherosclerotic processes, very different dynamics are to be observed between the various diseases in the recent past.

In coronary heart disease, incidence was postponed, probably because of improved hypertension control and decreased smoking prevalence. This

should lead to lower disease prevalence, but disease duration was increased too (see chapters 4 and 5); on balance, prevalence increased. Particularly when thrombolytic therapy revolutionized treatment of acute myocardial infarction, many patients survived, but with an impaired heart function, and heart failure incidence and prevalence increased. The effect was that both disease free life expectancy and that life expectancy with disease expanded. As mortality was lowered at all disease stages, prevalence increase cumulated in the most severe disease stages. In sum, the recent history of coronary heart disease was a typical example of expansion of morbidity, both in numbers and in severity. In stroke, incidence decline was prominent, but survival improvements rather limited. Because of the older age of stroke patients, the increased survival was cut short by mortality from other causes, and the sum of remaining morbidity was declining: the history of stroke between the seventies and 1987 was a rather typical example of compression of morbidity (see chapter 6).

The morbidity changes in these cardiovascular diseases that share many common characteristics and risk factors were thus completely opposite, driven predominantly by differences in available health care technology. Disease duration increased rapidly among patients suffering from heart disease, because of increasingly effective therapies. Among stroke patients, incidence was delayed more (partly because of successful hypertension control) and death was delayed less. At the oldest ages, prevalence of stroke survivors tended to increase because of increasing numbers of survivors, but this increase was more than balanced by the decreasing numbers of young stroke patients.

Chapter 8, about colorectal cancer, shows how in solid cancers incidence often is increasing, fueled predominantly by increasing case detection, but mortality is decreasing, hopefully as a consequence of that earlier detection. Here too there are opposite dynamics: numbers of new and old patients increase, numbers of interventions increase and health care needs increase, but the burden of morbidity is the sum of the morbidity caused by diagnosis and (demanding) treatment and the very severe morbidity of terminal disease preceding cancer death. If early detection succeeds in delaying death, both numbers of patients and disease duration are expanding, but the severe end-stage of cancer is compressing.

In chapter 9 we consider hip fractures, whose rates are increasing since decades (Boereboom, De Groot, Raymakers, & Duursma, 1991). Hip fractures are among the most important sources of health care costs among elderly women (Polder et al., 1997), and the burden of morbidity has to be considerable as well. We show that compression by increasing bone mineral density will not be as straightforward as often thought. Hip fractures are a senescent disease with a multifactorial character. Prevention of the fast menopausal bone mass loss by estrogen decrease precedes the age of high hip fracture rates with several decades and will have but moderate effects.

Conclusion

In this part of our thesis we show the importance of considering specific diseases in the debate of future morbidity. Within a large disease group with many shared characteristics such as cardiovascular disease, there may be opposite morbidity trends for specific diseases. Even in a single disease the character of morbidity may change as the prevalence of various disease stages change: in colorectal cancer (characteristic of most solid tumors) the burden of mild morbidity increases because of increasing incidence but the burden of severe morbidity preceding cancer death tends to decrease. These results are obtained by using epidemiological models, reproducing observed event rates by a quantified disease history, the main parameters of the history taken from the literature.

But the integration to overall population health requires more than single disease models. The occurrence of various diseases may depend on the same risks (old age and senescent decay is the most striking example, but also smoking causes a wide variety of diseases), and changes in population health are determined by more than just one disease. Part IV and part V will explain how these disease specific results can be integrated to changes in overall population health status.

References

- Boereboom, F., De Groot, R., Raymakers, J., & Duursma, S. (1991). The incidence of hip fractures in the Netherlands. *Neth J Med*, 38, 51-58.
- Bonneux, L., van de Mheen, P., Gunning-Schepers, L. J., & van der Maas, P. J. (1996). Daalde hypertensie in Nederland tussen 1974 en 1993? *Ned Tijdschr Geneesk*, 140(52), 2603-2606.
- Doll, R., & Peto, R. (1981). The causes of cancer. Sources of bias in estimating trends in cancer mortality, incidence and curability (appendix C). *J Natl Cancer Inst*, 66(6), 1270-81.
- Kunst, A. E., Looman, C. W., & Mackenbach, J. P. (1993). Determinants of regional differences in lung cancer mortality in The Netherlands. *Soc Sci Med*, 37(5), 623-31.
- Mackenbach, J. P., van Duyn, W. M. J., & Kelson, M. C. (1987). Certification and coding of two underlying causes of death in the Netherlands and other countries of the European Community. *J Epid Comm Health*, 41, 156-60.
- Murray, C. J. L., & Lopez, A. D. (Eds.). (1996). *The Global Burden of disease*. Harvard: Harvard University Press.
- Polder, J. J., Meering, J. W., Koopmanschap, M. A., Bonneux, L., & Van der Maas, P. J. (1997). *Costs of Diseases in 1994 (in Dutch)*. Rotterdam: Department of Public Health.

- Rosenberg, P. S., Biggar, R. J., Goedert, J. J., & Gail, M. H. (1991). Backcalculation of the number with human immunodeficiency virus infection in the United States. *Am J Epidemiol*, 133(3), 276-85.
- van de Mheen, P. J., Bonneux, L., & Gunning-Schepers, L. J. (1995). Variation in reported prevalences of hypertension in The Netherlands: the impact of methodological variables. *J Epidemiol Community Health*, 49(3), 277-80.

Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. The future rise of heart failure.

Abstract

Objectives: Many developed countries have seen declining mortality rates for heart disease, together with an alleged decline in incidence and a seemingly paradoxical increase in health care demands. The paper presents a model for forecasting the plausible evolution of heart disease morbidity.

Methods: The simulation model combines data from different sources. It generates acute coronary event and mortality rates from published data on incidences, recurrences and lethalties of different heart disease conditions and interventions.

Forecasts are based on plausible scenarios for declining incidence and increasing survival.

Results: Mortality is postponed more than incidence. Prevalence rates of morbidity will decrease among the young and middle aged, but will increase among the elderly. As the milder disease states act as risk factors for the more severe, effects will cumulate in the most severe disease states, causing a disproportionate increase in severe morbidity in older people.

Conclusion: Increasing health care needs in the face of declining mortality rates are no contradiction, but reflect a trade off of mortality for morbidity. The aging of the population will accentuate this morbidity increase.

Introduction

Heart disease, comprising ischemic heart disease and congestive heart failure, is not only the leading cause of death and lost life expectancy in most Western countries, but also one of the most important sources of morbidity and health care costs.¹ After a period of increasing mortality during the fifties and sixties, the same countries have witnessed a dramatic decline of ischemic heart disease mortality (age adjusted), with the United States taking the lead.² Though the dynamics are not completely understood, time trends in known risk factors for ischemic heart disease (hypertension, cholesterol, smoking) in Western countries indicate a clear decline.^{3,4,5} Therefore it is reasonable to attribute at least part of this decrease in mortality to a decrease in incidence.

The observed sharp decline in mortality, together with the probable decline in incidence leads one to expect a concomitant decrease in health care utilization. In reality, however, health care providers all over the Western world are facing increases in demand. After its introduction, the utilization of percutaneous transluminal coronary angioplasties has increased almost exponentially, while the growth in utilization of coronary artery bypass graft surgery has merely slowed.⁶ Viewed superficially, the observed increases in demand are incongruent with the supposed decrease in need.

But is need truly decreasing? Basic epidemiologic theory teaches us that if mortality slows down faster than incidence, prevalence will increase. Although the Framingham Heart Study has shown a small decline in incidence, a strong increase in prevalence has been observed, indicative of a survival gains.³ This improved prognosis has been partly caused by improved risk factor distributions, but medical therapy has also contributed.^{7,8} Recently, beta-blocking agents, platelet aggregation inhibitors, coronary artery bypass grafting surgery and Angiotensine Converting Enzyme inhibitors have been applied in chronic therapy, and thrombolytic therapy in emergencies. All have proven to be effective in randomised controlled trials of high standards.^{9,10,11} If trends in risk factors and survival remain favourable, the further decrease in incidence and mortality of ischemic heart disease is likely to be accompanied by a continuing increase in prevalence. In Western countries, the increased prevalence will be accentuated by the ageing of the baby boom generation, born after World War II. This cohort will reach middle age with its concomitant heart disease risks: the increase in age-specific prevalence will be accompanied by an increase in sheer size of the older age groups.

To make a quantitative analysis of the dynamics of the heart disease epidemic, we developed a simulation model for the most important manifestations of ischemic heart disease and congestive heart failure. Available information from the medical literature was combined with data obtained from mortality and morbidity registers. We explored the future of heart disease morbidity in the Netherlands by different scenarios.

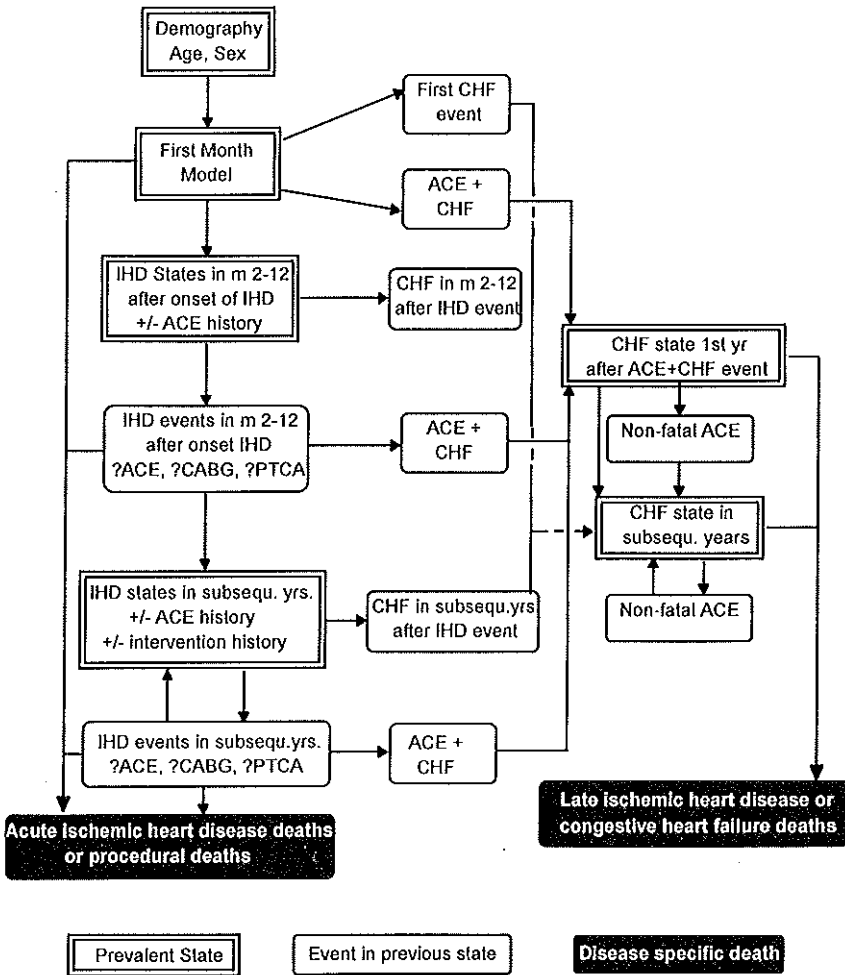
Methods

Model structure

The deterministic state transition model used divides the population under study (here the entire Dutch population) into homogeneous subpopulations, called states, defined by age, sex and disease history. Each year, the population "ages" by 1 year and the states are updated. Transitions from one state to another are governed by probabilities, extracted from registries and literature. Transition probabilities are independent from the preceding states and depend only on the current state (Markov chain assumption).¹² The structure of the model resembles the Coronary Heart Disease Policy Model of Weinstein *c.s.*, but we have excluded cardiac arrest as a prevalent state and included congestive heart failure.¹³

The morbidity model consists of three main sectors: the first month model, the ischemic heart disease model, and the congestive heart failure model. Persons developing heart disease, the rate of which is dependent on age and sex, move into the first month model, which characterises the initial event and its immediate sequelae during the first 30 days following the onset of disease. Persons in whom congestive heart failure forms the first heart disease event move directly into the appropriate model. Survivors of the first 30 days enter the ischemic heart disease model or the congestive heart failure model that describes the remaining 11 months of the first year, and all subsequent years. Persons with ischemic heart disease can move into the congestive heart failure model, but not vice versa. Death from heart disease is modeled through the specific transition probabilities of the model. In all prevalent states persons are subjected to an additional risk of death from all other causes, corrected for heart disease, based on national life tables.¹⁴

The first month model covers the month immediately after the initial cardiac disease event (Figure 1). This event may be angina pectoris, an acute coronary event (an acute myocardial infarction and/or cardiac arrest) or congestive heart failure (congestive heart failure). Persons suffering a first acute coronary event may die before reaching a health care facility ("deaths before admission"). Once taken into care, they are subjected to a further death risk ("deaths in hospital"). If the persons survive the coronary event, the pump function of the heart may be permanently impaired. Patients with heart failure after an acute myocardial infarction enter the congestive heart failure model, the others move into the ischemic heart disease model. Persons with a first diagnosis of congestive heart failure or uncomplicated angina proceed to the ischemic heart disease and congestive heart failure model respectively.



(Death of other causes occurs in all prevalent states, at any time)

Note:

CHF is congestive heart failure; ACE is acute coronary event; IHD is ischemic heart disease; CABG is coronary artery bypass graft; PTCA is percutaneous transluminal coronary angioplasty.

Figure 1 - The heart disease model

On entry, the ischemic heart disease model distinguishes between angina pectoris and acute coronary event states. Conditional on these states, patients face possible new events in the 2 to 12 months after their first incident, ranging from an acute coronary event, a coronary artery bypass grafting, a percutaneous transluminal coronary angioplasty to all possible

combinations of these. Each combination of events in a state has a probability of sudden death from ischemic heart disease dependent on the current event, the current state (and therefore the ischemic heart disease history) and the patient's age and sex. The patients also risk heart failure as a sequel of an acute coronary event ("acute coronary event + congestive heart failure" in the figure) or as a consequence of chronic ischemic damage; these patients move into the congestive heart failure model. If none of the above occurs, patients will remain in the ischemic heart disease model.

After this first year, patients remaining in the ischemic heart disease model are reclassified into 4 new states, conditional on previous disease history. These states are described by the presence or absence of an acute coronary event or of an intervention. These patients face the same events as the previous year. At the end of each subsequent model year, persons 'age' one year and are reassigned to states that reflect their updated history.

Persons with heart failure after an acute myocardial infarction face a high risk of dying during the first year.¹⁵ On surviving this first year, they move to the congestive heart failure state describing the subsequent years, where they will stay until their death; this state was called "chronic", to distinguish it from the first, critical year following such a complicated infarction. In these "chronic" years the model assumes that mortality rates are lower. All other patients developing heart failure immediately enter the chronic congestive heart failure state.

Although non-fatal acute coronary events are possible in the congestive heart failure state, such events no longer effect any change in the assigned state. Patients in the congestive heart failure state face an annual probability of death due to cardiovascular causes which may be classified as acute myocardial infarction (ICD 410-411), chronic ischemic heart disease (ICD 412-414), congestive heart failure (ICD 428-429) or stroke (ICD 430-438). We have incorporated stroke as another specific cause of death, because the Framingham Heart Study showed that, due to common risk factors (hypertension and diabetes), congestive heart failure is associated with increased rates of stroke and death of stroke.¹⁶

Incidence and recurrence rates

Reliable and recent incidence data on heart disease are not available in the Netherlands, as in many other countries. Incidence figures have been estimated by an iterative non linear procedure from observed numbers of events (acute myocardial infarctions, ischemic heart disease deaths, interventions) and externally estimated transition probabilities.¹⁷ In this section we give a brief account how they were derived. The transition probability matrix for a typical age group (60 - 64 year) is furnished in table 1. The complete transition matrices with a more detailed account of references are available from the authors on request.

Table 1 - Transition probabilities for the age group 60-64.

	M	F
first incidence in the total population		
acute coronary event	0,80%	0,26%
angina pectoris	0,47%	0,30%
congestive heart failure	0,20%	0,09%
yearly probabilities of an acute coronary event in the diseased population		
<i>• in the population with a history of acute coronary event</i>		
Before no intervention in 1st y	5,3%	5,3%
Before percutaneous transluminal coronary angioplasty in 1st y	7,1%	7,1%
Before coronary artery bypass grafting in 1st y	10,7%	10,7%
After congestive heart failure in 1st y after acute coronary event	20,5%	19,4%
After no intervention in y > 1	2,8%	2,8%
After any intervention in y > 1	4,3%	4,3%
After congestive heart failure in y > 1	7,3%	6,2%
<i>• in the population with a history of uncomplicated angina pectoris</i>		
Before no intervention in 1st y	1,5%	0,6%
Before percutaneous transluminal coronary angioplasty in 1st y	3,7%	3,7%
Before coronary artery bypass grafting in 1st y	5,5%	5,5%
After no intervention in y > 1	1,5%	0,7%
After any intervention in y > 1	2,2%	2,2%
probabilities of congestive heart failure in the diseased population		
within 1 month after acute coronary event	8,5%	11,1%
after 1 month after acute coronary event	2,5%	2,5%
After angina pectoris	1,0%	0,5%
intervention rates		
<i>coronary artery bypass grafting</i>		
in 1st y after acute coronary event	11,3%	5,9%
in y > 1 after acute coronary event	2,3%	1,2%
in 1st y after angina pectoris	15,9%	8,8%
in y > 1 after angina pectoris	3,2%	0,6%

Table 1 - continued

	M	F
<i>percutaneous transluminal coronary angioplasty</i>		
in 1st y after acute coronary event	2,6%	1,9%
in y > 1 after acute coronary event	0,5%	0,4%
in 1st y after angina pectoris	3,9%	2,9%
in y > 1 after angina pectoris	0,8%	0,8%
mortality rates		
<i>heart disease:</i>		
within the 1st month after an acute coronary event		
first acute coronary event, before admission	17,4%	18,0%
first acute coronary event, inside hospital	8,4%	11,0%
recurrent acute coronary event, before hospital	24,0%	24,8%
recurrent acute coronary event, inside hospital	12,7%	16,5%
recurrent acute coronary event after congestive heart failure, before hospital	47,1%	44,0%
recurrent acute coronary event after congestive heart failure, inside hospital	18,9%	22,0%
operative mortality after 1st coronary artery bypass grafting	1,4%	3,0%
operative mortality after recurrent coronary artery bypass grafting	3,4%	3,4%
procedural mortality after percutaneous transluminal coronary angioplasty	0,7%	0,7%
1st-12th m after acute coronary event with congestive heart failure	19,9%	17,9%
in y > 1 after congestive heart failure	13,0%	11,0%
<i>All other causes of death in all states</i>	1,14%	0,64%

acute coronary event: Acute Coronary Event (including acute myocardial infarction, unstable angina pectoris and cardiac arrest)

angina pectoris: Uncomplicated Angina Pectoris

congestive heart failure: Congestive Heart Failure

coronary artery bypass grafting: Coronary Artery Bypass Grafting

percutaneous transluminal coronary angioplasty: Percutaneous Transluminal Coronary Angiography

in y > 1: In the years subsequent to the first year after an event.

Total numbers of admissions for acute myocardial infarction (ICD 410¹⁸), total numbers of hospital deaths, and total numbers of interventions by age and sex are available from the nationwide hospital register^{19,20,21,22}. Patients with symptomatic infarctions who are not admitted to hospital (6%, in the Netherlands limited to the oldest age groups) are added.²³ A constant ratio for the probability of dying before hospital admission versus dying inside the hospital from an acute coronary event is assumed.^{20,21,24,25}

Reported probabilities of recurrent acute coronary events vary widely, from 2 % to 13 % and more per year, depending on age, case-mix and selection criteria.^{13,26,27,28,29,30} We have assumed that ischemic heart disease patients without congestive heart failure have a first annual probability of a recurrent acute coronary event of 0.062. This is consistent with the estimate of Weinstein et al of 0.078, in which patients with congestive heart failure are included.^{13,31} Recurrence rates are strongly age dependent. We assume the age distribution to be identical to the known age distribution of a first acute coronary event in the reference population.

First acute coronary events after an history of angina pectoris were modeled based on data derived from the Framingham study, showing that male patients with angina pectoris, who have suffered no previous acute myocardial infarction, have a relative risk of 0.54 compared to those with a previous acute myocardial infarction.³² Female patients with angina pectoris have a relative risk of 0.45 compared to male patients.³²

Patients eligible for coronary artery bypass grafting surgery run a risk for recurrent events that is twice that for patients who are not eligible.^{31,33,35} Coronary artery bypass grafting reduces hazard rates by 20 %.^{36,37} Lacking data on the effectiveness of percutaneous transluminal coronary angioplasty, we assumed the relative risk for recurrent acute coronary event to be 1.5, with a rate reduction of 10 %.

Intervention rates

Intervention rates are known to vary widely from country to country and even from hospital to hospital.^{38,39} We have used age and sex specific nationwide output data and assumed constant proportional relations between states to estimate the numbers of interventions in each state. The risk of an intervention in uncomplicated angina pectoris versus the risk of an intervention after suffering an acute coronary event is 1.4: 35 % of patients with a history of an acute coronary event are potentially eligible for an intervention compared to 50 % in uncomplicated angina.^{40,41} The risk of undergoing an intervention the first year after an ischemic heart disease event versus the subsequent years was taken to be 5.⁴² The annual probability of having a subsequent intervention after a first intervention is 1 % (coronary artery bypass grafting) and 0,5 % (percutaneous transluminal coronary angioplasty).⁴³ For every 10 successful percutaneous transluminal coronary angioplasties, another 4 are considered to be not successful due to restenosis.⁴⁴ If both coronary artery bypass grafting and percutaneous

transluminal coronary angioplasty occurred in the same year, percutaneous transluminal coronary angioplasties are ignored.

Heart failure

Heart failure due to ischemic heart disease is caused by acute events or by chronic ischemic damage. 18 % of the ischemic heart disease patients survive an acute myocardial infarction with irreversible heart failure as a sequel.^{33,34} The distribution by age, sex and previous ischemic heart disease history is assumed to be the same as the distribution of in hospital mortality by age, sex and previous ischemic heart disease history after an acute myocardial infarction.

The annual probabilities of developing congestive heart failure, given an ischemic heart disease history, are estimated from the Framingham Heart Study.¹⁷ We obtained the published annual event rates of 1.5 % given an uncomplicated angina pectoris history and 3.0 % given a history of an acute coronary event and adjusted to allow for the number of acute cases caused by an acute coronary event. The age distribution was calculated by applying a relative risk of 3.0 (angina pectoris) and 5.0 (acute myocardial infarction) to the age specific incidence data taken from Framingham.¹⁷

The incidence rates of congestive heart failure without a history of ischemic heart disease are calculated by assuming that persons in whom congestive heart failure is the certified cause of death survived their first diagnosis of congestive heart failure for 5 (males) to 7 years (females).^{11,12,15,17}

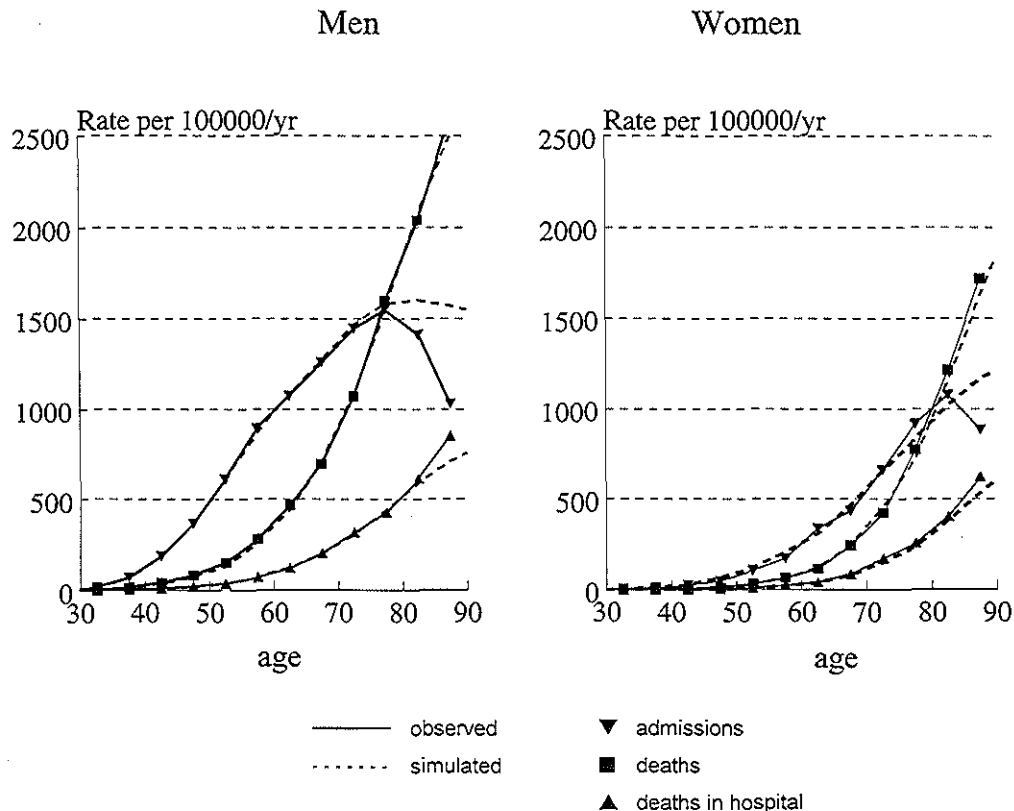
Heart disease mortality

The total age- and sex-specific in-hospital mortality rates following acute coronary events and interventions are known in the Netherlands.^{20,22} The relative risk of dying after a subsequent versus a first acute coronary event was taken to be 1.5.^{13,28,45} Publications are available in which the figures on mortality due to a subsequent versus a first coronary artery bypass grafting operation are given.⁴⁶ The relative risk of death due to percutaneous transluminal coronary angioplasty versus coronary artery bypass grafting is 0.5.^{47,48} Procedural mortality due to a subsequent or a first percutaneous transluminal coronary angioplasty is taken to be the same.

Risk of death during the first year after acute myocardial infarction followed by congestive heart failure was assumed to be 25 %; the distribution by age was constructed by exponential inter- and extrapolation of data taken from Ahnve et al.³⁴ The annual mortality risk of heart failure during the subsequent years is assumed to be constant, i.e. 10 % (females) and 13 % (males).^{15,17,33,45} Of all deaths caused directly or indirectly by heart failure 25 % (female) and 37 % (male) are attributable to recurring ischemic heart events, and another 25 % (males) and 35 % (females) to stroke (and hence to non-cardiac disease).¹⁷ The figures for non fatal acute coronary events in persons with heart failure are estimated by assuming a case fatality of 60 %.

Figure 2 - Observed and simulated events in 1985

Admissions for acute myocardial infarction are the rates of patients discharged with as primary diagnosis acute myocardial infarction (ICD 410), deaths in hospital are the rates of patients with this same diagnosis dying in hospital.²⁰ Deaths from acute myocardial infarction are the total rates of death from acute myocardial infarction (ICD 410).²¹ The solid lines with symbols show observed numbers, the dotted line represents the simulation by the heart disease model. The simulated points given for the group aged 85-89 apply solely to these 5 years, while the observed points are for the open ended category "85 +".



All persons in all states are subjected to the risk of dying of other diseases, adjusted to allow for the causes of death under study and based on Dutch life tables.

Results

Calculated estimates for 1985

Figure 2 shows the observed and simulated events of acute myocardial infarction: total admissions and deaths in and outside the hospital by age and sex. The model closely reproduces the observed data. The gap between simulated and observed rates in the older age groups represents recognized acute myocardial infarction, which are not admitted to a hospital.

Calculated prevalences of heart disease by age and sex are shown in figure 3. Heart disease is more concentrated in the oldest age groups in females, while the burden of morbidity in the male population is already high at middle age. The levelling off and decrease by age in the prevalence of clinical ischemic heart disease in men shown in the figure is caused by the steep rise in the incidence of congestive heart failure and sudden cardiac death, reducing clinical ischemic heart disease incidence. Prevalence of all heart disease reaches up to 25 % in the age groups of 70-74 (males) and 80-84 (females). Women suffer relatively more often than men from angina pectoris and congestive heart failure, but less often from severe ischemic heart disease. This has been confirmed time and again in many surveys.^{32,49,50}

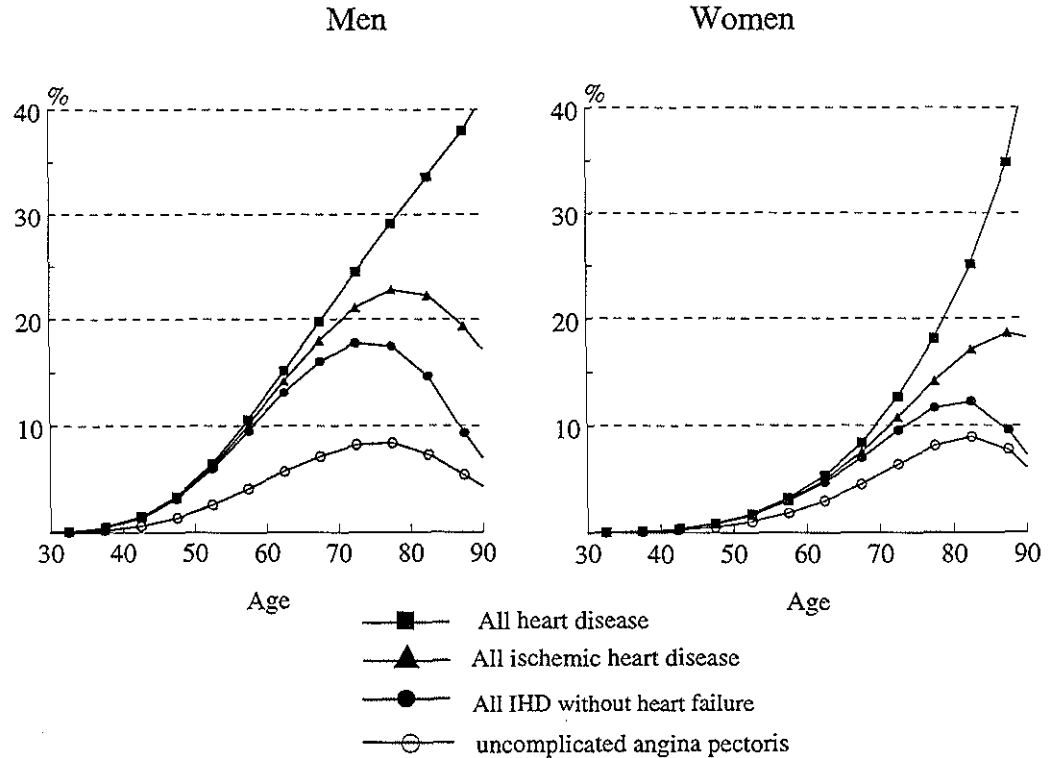
Projections for 1985-2010

Disease specific mortality is the direct result of incidence and survival. If two of these elements are known, the third can be calculated. As the decline in ischemic heart disease mortality, which has been observed since 1975, is indisputable, we postulated an ongoing decline of this mortality into the near future. We provide three plausible scenarios, based on varying assumptions of further decreasing incidence rates and improving prognoses (see table 2).

The reference scenario, based on the Framingham Heart Study, supposes that one third of the mortality decline is explained by a decline in incidence (-1.0 %/year) and that two thirds is explained by an increase in survival.^{3,4,8,9,26,51} In an extreme "low incidence change scenario" we assume that the incidence decrease is even less: - 0.5 % /year. This is the magnitude of change predicted by a risk factor model, based on changes in risk factor prevalences in the Netherlands.⁵ In an extreme "high incidence change scenario", we assume a high decrease in incidence (-1.7 % per year).

Figure 3 - Cumulated prevalences of heart disease by age and sex, the Netherlands, 1985

Uncomplicated angina pectoris is defined as angina pectoris only, without a history of an acute myocardial infarction. *Ischemic heart disease without congestive heart failure* describes all patients with a history of ischemic heart disease but without congestive heart failure. *All ischemic heart disease* comprises all ischemic heart disease, including patients who developed congestive heart failure after ischemic heart disease. *All Heart disease* summarizes all three conditions. The difference between all heart disease and all ischemic heart disease shows the prevalence of congestive heart failure.



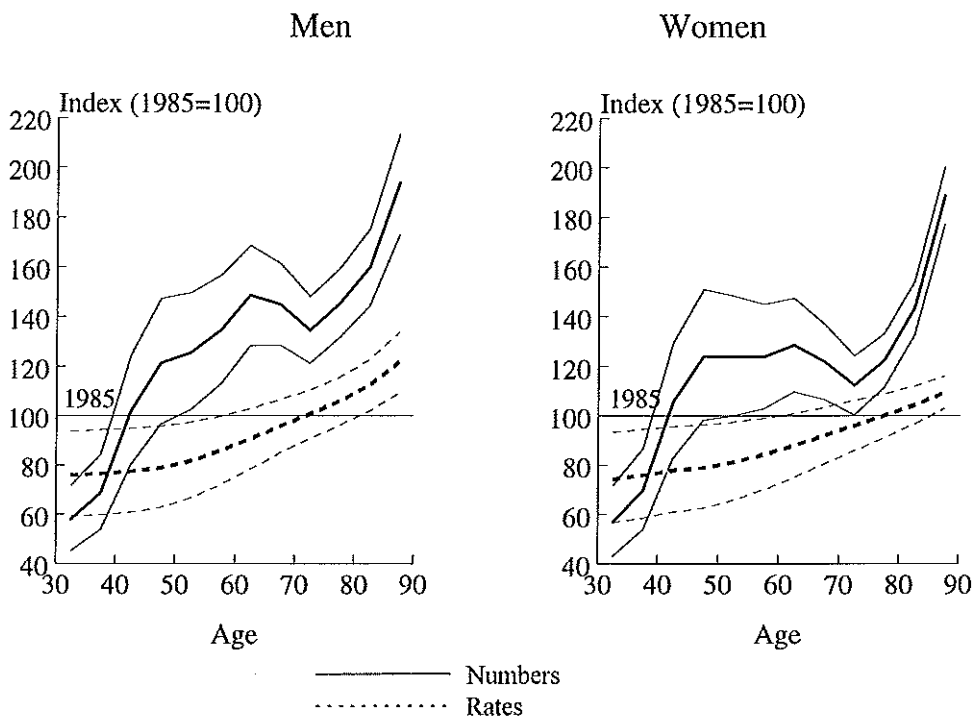


Figure 4 - Expected change in absolute and relative prevalences of patients with ischemic heart disease, the Netherlands, 2010, compared with 1985.

The dashed lines show the expected index of change in prevalence rates between 1985 and 2010 (prevalence in 2010 divided by prevalence in 1985), with 1985 = 100. The solid lines indicate the change in absolute numbers, showing the force of demography. The heavier lines indicate the reference scenario, the thinner the both extreme scenarios.

Table 2 - Annual changes postulated in the scenarios. The extreme scenarios figure between brackets. The mortality decline (observed between 1975-1990) is kept equal for all scenarios.

	Men	Women
Incidence	-1,1% [-1,7%; -0,6%]	-0,9% [-1,4%; -0,5%]
mortality before admission	-1,2% [-0,9%; -1,4%]	-0,7% [-0,5%; -0,8%]
mortality after adm. within 1st month	-2,1% [-1,6%; -2,5%]	-1,3% [-1,0%; -1,6%]
late mortality	-1,8% [-1,4%; -2,1%]	-1,8% [-1,4%; -2,1%]
Recurrence	-3,1% [-2,3%; -4,7%]	-1,9% [-1,4%; -2,9%]
ischemic heart disease mortality	-2,4% †	-2,0% †

† It is assumed that the mortality decline remains equal in all scenarios.

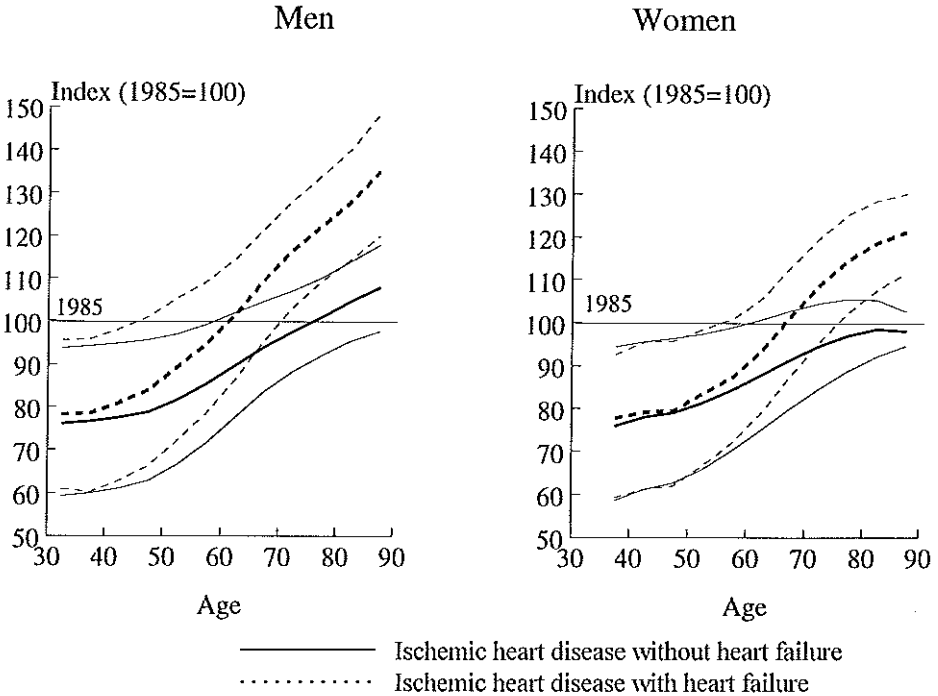


Figure 5a - Expected relative change in prevalence rates of ischemic heart disease, the Netherlands, 2010 compared with 1985.

The solid lines indicate the relative changes in number of ischemic heart disease patients without heart failure, the dashed those with heart failure (age specific prevalences in 2010 divided by the same prevalences in 1985). The heavier lines indicate the reference scenario, the thinner the two extreme scenarios. The relative change of all ischemic heart disease patients is shown in figure 4.

We have projected these trends over the period 1985-2010 to juxtapose the effects of demographic with epidemiologic change. The results are summarized in figures 4 and 5 and table 1. Figure 4 shows how the decreasing incidence together with the increase in survival results in lower prevalence rates in the young and increasing prevalence rates in older age. In age group 35 to 39 the decrease is about 25%, while in age group 85-89, prevalence increases by 20% (males) and 10% (females).

On superimposing the population projections on these figures, a completely different picture is obtained, which is shown in the same figure. The movement of the post-war baby boom generation into middle age and the sharp increase in the number of the very elderly is clearly revealed. Together with the projected increase in prevalence rates, this results in a steep climb in the number of heart disease patients in the highest age groups. Hence, while epidemiological changes in cardiovascular disease result in an

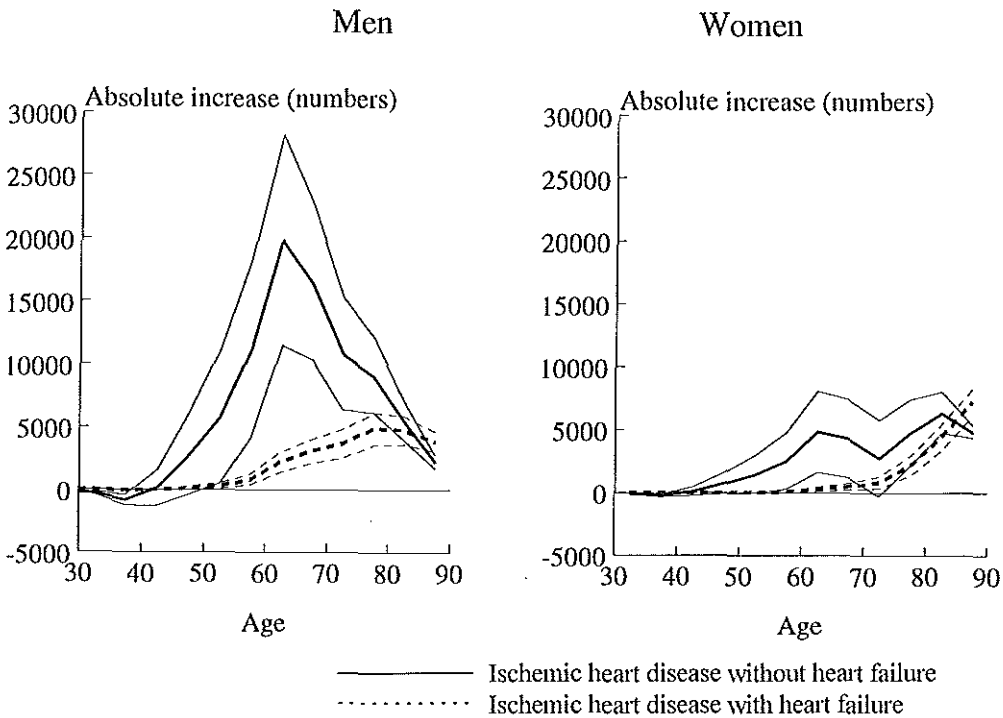


Figure 5b - Expected absolute change in numbers of patients with ischemic heart disease, the Netherlands, 2010 compared with 1985

The solid lines indicate the absolute changes in number of ischemic heart disease patients without heart failure, the dashed those with heart failure (age specific numbers of patients in 2010 minus the same numbers in 1985). The heavier lines indicate the reference scenario, the thinner the two extreme scenarios.

increase of prevalence rates of 10 to 20% for age group 85-89, the absolute number of patients will more than double.

Figure 5 shows the changes in age specific prevalences for different types of heart disease; 5a shows the relative changes of the rates and 5b the absolute changes in number. The change in age-specific rates for all ischemic heart disease combined is the same as in figure 4.

Within the category ischemic heart disease, figure 5a shows a relative increase of severe ischemic heart disease. Prevalences of milder stages of disease act as a risk factor for the more severe. If survival increases in all disease stages, effects will cumulate in the most serious stages: as more patients stay alive with ischemic heart disease, more will have the opportunity to develop congestive heart failure.

Figure 5 b shows the changes in absolute terms: the relative decrease of prevalence at younger ages has hardly any impact due to the low number of

patients with ischemic heart disease at an early age. However, the combination of ageing and increasing survival of heart disease leads to a flood of middle-aged men with ischemic heart disease followed by a slower wave of elderly persons with heart failure.

Table 3 shows expected changes in age standardized rates, using the European Standard Population as standard, and in absolute numbers, combining the rates with the expected population dynamics in the Netherlands. The mortality decline (observed between 1975 and '90 and projected over the next 20 years) is so strong, that this will be evident even when expressed in absolute numbers. Although the assumed incidence decline of first myocardial infarctions, which accounts for one third of the mortality decline is too small to counteract the ageing of the Dutch population, the attack rate of all myocardial infarctions together will nevertheless decrease substantially. This is the result of the combined decline in incidence and recurrence; the better survival, on the other hand, through which the number of patients at higher risk of recurrent infarctions is increased, is not strong enough to counteract this decline. Because more patients will stay alive and at risk for interventions and reinterventions, and because of the ageing of the population the need for coronary artery bypass grafting and percutaneous transluminal coronary angioplasty will increase by 1 % per year [range: 0.4 - 1.6]. The age standardized prevalence ratios for milder ischemic heart disease tend to decrease, but those for ischemic heart disease complicated with congestive heart failure increase strongly: in absolute numbers, a growth of more than 70 % is to be expected.

Discussion

Simulation modelling provides quantitative insight in epidemiological and public health dynamics and in potential future developments. The information needed for these models can never be supplied by a single data source. Combining data from different sources is therefore appropriate, while the use of state-transition models provided a simple and comprehensible means of doing so. The methods used are relatively easy to understand. The parameters of the model are meaningful to physicians and the results interpretable by health care providers, who are the principal users.

Nevertheless, in order to interpret results correctly, the limitations of this kind of modelling should be understood. A number of simplifying assumptions are always necessary to keep the model, and more particularly the number of parameters to be estimated, manageable.

We included certain disease manifestations, and excluded others. Unlike Weinstein *c.s.*, we excluded cardiac arrest as a prevalent state, but included congestive heart failure. Cardiac arrest is highly lethal, and no major error is introduced by mixing the history of those surviving an acute myocardial infarction with that of a few survivors of a cardiac arrest. On the other hand, the inclusion of congestive heart failure takes into account an important

Table 3 - Age standardized rates (per 100 000, using the European Standard Population as direct standard) and absolute numbers of admissions, interventions, patients and deaths due to ischemic heart disease (ischemic heart disease) and congestive heart failure (congestive heart failure) in the Netherlands in 1985 and the projected index of change (1985 = 100) in 2010.

The expected rates and numbers can be calculated by multiplying the figures for 1985 by the index of 2010 divided by 100. The upper and lower bounds present the results of scenarios assuming extremely high or low incidence changes, but leading to the same mortality changes as a result of extremely low or high survival changes (see text).

	MEN				WOMEN			
	1985		Index of change in 2010 (1985 = 100)		1985		Index of change in 2010 (1985 = 100)	
	Rates	Numbers	Rate	Numbers	Rates	Numbers	Rate	Numbers
acute coronary event incidence	267	16,438	76 [65 - 87]	114 [97 - 130]	109	9,507	80 [71 - 89]	115 [105 - 126]
First admissions	214	13,150	80 [67 - 93]	119 [100 - 139]	81	6,863	83 [73 - 94]	118 [105 - 131]
Admissions for acute coronary event	352	21,687	67 [60 - 73]	100 [90 - 108]	122	10,529	73 [66 - 78]	105 [96 - 111]
coronary artery bypass grafting	96	5,797	84 [72 - 90]	128 [110 - 147]	18	1,302	84 [72 - 97]	115 [98 - 132]
History of angina pectoris	1726	106,968	91 [79 -104]	136 [119 - 156]	1088	94,635	89 [78 - 99]	123 [110 - 136]
History of acute coronary event	2087	129,149	90 [79 - 102]	136 [118 - 153]	547	45,195	94 [82 - 106]	130 [114 - 146]
congestive heart failure with history of ischemic heart disease	575	35,761	118 [104 - 131]	170 [149 - 189]	284	29,251	115 [103 - 125]	169 [153 - 183]
congestive heart failure without history of ischemic heart disease	697	41,323	94 [88 - 126]	144 [133 - 163]	509	50,729	84 [77 - 104]	130 [120 - 146]
Deaths coded acute myocardial infarction	194	12,045	54 [†]	80 [†]	82	7,889	63 [†]	96 [†]
Deaths coded chronic ischemic heart disease or congestive heart failure	106	6,422	65 [†]	96 [†]	60	6,036	59 [†]	90 [†]

† In the calculation of upper and lower bounds, the mortality decline is assumed to remain equal.

cause of chronic and severe morbidity. By modelling the dynamics of and the relations between these different types and levels of morbidity a more realistic estimate of future health care needs should be achieved.

We used registered admissions for acute myocardial infarction (ICD18 410), deaths from ischemic heart disease (ICD 410-414) and congestive heart failure (ICD 428-429) and transition probabilities extracted from multiple sources in order to estimate incidences. This complicates the assessment of validity, which is the most important problem of this or any other model with such complexity. The model output reflects 'state of the art' clinical evidence, closely simulating present hospitalisation, intervention and mortality rates (see figure 2). Its estimated prevalences of clinical disease coincide closely with empirical observations.^{49,50,52}

Our forecasts are based on mortality projections. For health care planning purposes projections of incidence based on past risk factor prevalences are less reliable. Firstly, prevalences of all known risk factors for heart disease are not available; prevalence numbers which are available often show highly inconsistent results between different studies. For example, estimates of the prevalence of hypertension in Dutchmen from 50-59 y old vary between 14.4 and 67 %.^{5,52} Secondly, not all risk factors are known or observable: projections only based on known risk factors are bound to underestimate future trends.

The assumption that mortality will continue to decline is uncertain, but not implausible. It supposes a progressive change in risk factor distributions, a more widespread use of effective health care interventions in the future and ongoing technological advances.^{3,4,8,9} The level of ischemic heart disease mortality expected in 2010 is comparable to that of 1950 (making allowances for changes in codification). If the absence of effective therapies in 1950 is taken into account, this is not an extreme result. The reference scenario supposes a decrease of incidence, explaining one third of the mortality decline, based on the Framingham study.⁴ Other assumptions about incidence decreases and survival increases within extreme ranges are shown in the alternative scenarios.^{3,4,8} These influence the absolute magnitude of the expected changes, but not the direction.

Declining incidence rates mean that the onset of disease is postponed but not halted. The delay in occurrence causes a reduction of morbidity in the young, which, however, is cancelled out with age. Mortality is deferred to an even greater extent, owing to a combination of the decrease in incidence and improved prognosis. This mortality decrease will be traded off for a steep increase of more severe morbidity in the old.

Changes in demography will prevent a decrease of new cases of heart disease.¹³ However, while age specific incidence rates show a decline, acute myocardial infarction attack rates will drop more strongly, due to a combined reduction of incidence and recurrence rates; the increasing prevalences are insufficient to cancel these reductions. The reference scenario predicts that a decline in attack rates which will be sufficiently strong to offset

population ageing; this development has been observed in the Netherlands since 1985.^{19,20,21} But although disease incidence rates are declining, intervention rates will remain unchanged, even assuming that intervention rates per group of patients in a certain disease state remain constant. The reason is that more patients will survive and that, on average, survival is longer, resulting in more people in disease states eligible for invasive coronary revascularisation procedures.

The overall picture emerging from the heart disease scenarios is one of transition from an acute to a more chronic disease: event rates are declining strongly, but the postponed mortality is traded off for increased morbidity and severity of disease among the elderly. This can be observed in hospital admission rates in the Netherlands, in data from the Health Interview Survey in the USA and from the National Health Survey of Japan.^{19,53} All of these data show increasing prevalences of ischemic heart disease. As a consequence, overall health care needs will continue to rise, regardless of declining mortality and incidence.

These calculations, based on empirical observations, shed a new light on Fries' compression of morbidity theory.⁵⁴ The compression of morbidity theory hinges on the proposition that the incidence of disease will be postponed more rapidly than mortality. For heart disease, the evidence suggests that the decline of lethality expands the number of years spent with heart disease morbidity, regardless of the incidence postponement. Healthy life expectancy is extended as expected by Fries and assumed by the scenarios, but contrary to his prediction total life expectancy is extended even further. Among the middle aged and elderly, acute heart disease lethality in the Netherlands is high, increasing sharply with age. Obviously, here would appear to be a serious potential for reduction. In 1985, for example, the model calculated that 41 % of the 70 year old men died within 4 weeks of their first acute coronary event.²⁰⁻²⁴ In our scenarios 32 % [range: 30% - 34%] are predicted to die within the same period in the year 2010, which in our view remains a conservative estimate. Because heart diseases is concentrated in the older age groups, even small differences in the rates of change in incidence and mortality can have major consequences, particularly now that pushing back mortality has been shown to be counterbalanced by an increase in the more severe stages of heart disease. With short term case fatality subsequent to an acute event as high as it is, lowering this will remain an important goal of health care intervention. Should this aim be successful, the resulting increase in the number of survivors will boost heart disease morbidity and lengthen life expectancy with disease. Only if health care interventions aimed at the reduction of mortality subsequent to the event should prove a waste of time and money would there be no expansion of morbidity expected.

References

1. Koopmanschap MA, Roijen van L, Bonneux L. Costs of Illness in the Netherlands. Report MGZ 91.03. Rotterdam: Department of Public Health and Social Medicine, Institute of Medical Technology Assessment, 1991 (*in Dutch*).
2. Kaazuo Uemura, Zbynek Pisa. Trends in cardiovascular disease mortality in industrialised countries. *Wld Health Statist Quart* 1988; 41: 155-68.
3. Sytkowsky PA, Kannel WB, D'Agostinho RB. Changes in risk factors and the decline in mortality from cardiovascular disease. *N Engl J Med* 1990; 322:1635-41.
4. Sigfusson N, Sigvaldason H, Steingrimsdottir L, et al. Decline in ischemic heart disease in Iceland and change in risk factor levels. *Br Med J* 1991; 79: 366-71.
5. Gunning-Schepers L. The health benefits of prevention: a simulation approach. *Health Policy* 1989; 12/1-2 (special issue); 93-129.
6. Graboys TB, Biegelsen B, Lampert S, Blatt CM, Lown B. Results of a second-opinion trial among patients recommended for coronary angiography. *JAMA* 1992; 268: 2537-40.
7. Mackenbach JP, Looman CWM, Kunst AE, Habbema JDF, Maas PJ van der. Post-1950 mortality trends and medical care: gains in life expectancy due to declines in mortality from conditions amenable to medical intervention in the Netherlands. *Soc Sci Med* 1988; 27: 889-94.
8. Goldman L, Cook EF. The decline in Ischemic Heart Disease mortality rates. *Ann Intern Med* 1984; 101: 825-836
9. Moss AJ, Benhorin J. Prognosis and management after a first myocardial infarction. *N Engl J Med* 1990; 322: 743-53.
10. SOLVD investigators. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med* 1991; 325: 293-302.
11. Cohn JN, Johnson G, Ziesche S, et al. A comparison of enalapril with hydralazine-isosorbide dinitrate in the treatment of chronic congestive heart failure. *N Engl J Med* 1991; 325: 303-10.
12. Manton KG, Stallard E. *Chronic Disease Modelling*. London: Charles Griffins & Co., 1988: 207-210.
13. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality and cost: The coronary heart policy model. *Am J Public Health* 1987; 77: 1417-26.
14. Central Bureau of Statistics. Monthly population statistics. Voorburg (Netherlands), Central Bureau of Statistics. 1986/3; pp 20-28.
15. McFate Smith W. Epidemiology of congestive heart failure. *Am J Cardiol* 1985; 55: 3A-8A
16. World Health Organisation. *International Classification of diseases, injuries and causes of death, 9th revision*. Geneva: World Health Organisation, 1977.
17. Kannel WB. Epidemiological aspects of heart failure. *Cardiology clinics* 1989; 7: 1-9.
18. Press, W., Flannery, B.P., Teukolsky, S.A., & Vetterling, T. *The art of scientific computing*. Cambridge University Press. Cambridge: 1990.

19. Hoogendoorn DE. Atherosclerotic diseases of the heart in the hospital and in the vital statistics. *Ned Tijdschr Geneesk* 1985; 129: 1827-33 (*in Dutch*).
20. Hoogendoorn D. Some notes on the current situation regarding the epidemic of acute myocardial infarction. *Ned Tijdschr Geneesk* 1990; 134: 592-595 (*in Dutch*).
21. LMR (Information Center for Health Care). Files from the National Medical Registration concerning admissions and interventions in 1988. Utrecht (Netherlands) 1990.
22. Hoogendoorn D. Frequencies and operative mortality of interventions on the heart and the large intrathoracic vessels. I. Direct revascularisation of the coronary arteries. *Ned Tijdschr Geneesk* 1989; 133: 2445-2448 (*in Dutch*).
23. Fracheboud J. Coronary Care or Home Care? A descriptive study of Home Care for patients with an acute myocardial infarction. NIVEL, Utrecht, 1987. (*in Dutch*)
24. Lubsen J, Does van der E, Pool J. Incidence of acute myocardial infarction and sudden death. *Hart Bulletin* 1976; 7: 107-113.
25. Goldman L, Cook F, Hashimoto B, Stone P, Muller J, Loscalzo A. Evidence that hospital care for acute myocardial infarction has not contributed to the decline in coronary mortality between 1973-74 and 1978-79. *Circulation* 1982; 65: 936-42.
26. Ericsson CG, Erhardt L, Rehnqvist N. Two-year survival after myocardial infarction. *J Intern Med* 1990; 227: 195-9.
27. Marcus FI, Friday K, McCans J, et al. Age related prognosis after acute myocardial infarction (the multicenter diltiazem postinfarction trial research group). *Am J Cardiol* 1990; 65: 559-66.
28. Benhorin J, Moss AJ, Oakes D, and the multicenter diltiazem postinfarction trial research group. Prognostic significance of nonfatal myocardial reinfarction. *J Am Coll Cardiol* 1990; 15: 253-8.
29. Multicenter Diltiazem Postinfarction Trial Research Group. The effect of Diltiazem on mortality and reinfarction after myocardial infarction. *N Engl J Med* 1988; 319: 385-82.
30. Norwegian Multicenter Study Group. Timolol induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med* 1981; 304: 801-7
31. DeBusk RF, Kraemer HC, Nash E, Berger WE III, Lew H. Stepwise risk stratification soon after acute myocardial infarction. *Am J Cardiol* 1983; 52: 1161-6.
32. Kannel WB, Feinleib M. Natural history of angina pectoris in the Framingham study. *Am J Cardiol* 1972; 29: 154-63.
33. DeBusk RF, Blomqvist G, Kouchoukos NT, et al. Identification and treatment of low-risk patients after acute myocardial infarction and coronary artery bypass graft surgery. *N Engl J Med* 1986; 314: 161-6.
34. Ahnve S, Gilpin E, Dittrich H, et al. First myocardial infarction: age and ejection fraction identify a low-risk group. *Am Heart J* 1988; 116: 925-32.
35. Sanz G, Castañer A, Betriu A, et al. Determinants of prognosis in survivors of myocardial infarction. *N Engl J Med* 1982; 306: 1065-70.
36. European Coronary Surgery Group: Long term results of a prospective randomised study of CABG in stable angina pectoris. *Lancet* 1982; ii: 1173-1180

37. CASS principal investigators and their associates. Coronary Artery Surgery Study: A randomized trial of coronary artery bypass surgery. *Circulation* 1983; 68: 939-50.
38. Brook RH, Kosecoff JB, Park RE, Chassin MR, Winslow CM, Hampton JR. Diagnosis and treatment of coronary artery disease: comparisons of doctors attitudes in the USA and the UK. *Lancet* 1988; ii: 750-753.
39. Winslow CM, Kosecoff JB, Chassin M, Kanouse DE, Brook RH. The appropriateness of performing coronary artery bypass surgery. *JAMA* 1988; 260: 505-509.
40. Weintraub WS, Jones EL, King SB, et al. Changing use of coronary angioplasty and coronary bypass surgery in the treatment of chronic coronary artery disease. *Am J Cardiol* 1990; 65: 183-188.
41. Kaul S, Lilly DR, Gascho JA, et al. Prognostic utility of the exercise thallium-201 test in ambulatory patients with chest pain: comparison with cardiac catheterization. *Circulation* 1988; 77: 745-758.
42. Parisi AF, Peduzzi P, Detre K, et al. Characteristics and outcome of medical nonadherers in the veterans administration cooperative study of coronary artery surgery. *Am J Cardiol* 1984; 53: 23-8.
43. Veldkamp RF, Baartman GJ, Domburg van R, Thijssen JGP, Bos E, Meeter K. Survival 11 year after an Aortocoronary bypass operation. *Ned Tijdschr Geneesk* 1991; 135: 1229-33 (*in Dutch*).
44. Glazier JJ, Varicicchione TR, Ryan TJ, et al. Outcome in patients with recurrent stenosis after percutaneous transluminal balloon angioplasty. *Br Heart J* 1989; 61: 485-488.
45. Fioretti P, Tijssen JGP, Azar AJ, et al. Prognostic value of predischarge 12 lead ECG after myocardial infarction compared with other routine variables. *Br Heart J* 1987; 57: 306-12.
46. Laird-Meeter K, Domburg van R, Brand van den MJBm, Lubsen J, Bos E, Hugenholtz PG. Repeat interventions after initial aorto-coronary bypass surgery and its results. *Ned Tijdschr Geneesk* 1988; 132: 2316-20.
47. Detre K, Holubkow R, Kelsey S, et al. One year follow-up results of the 1985-1986 National Heart, Lung, and Blood Institute's PTCA registry. *Circulation* 1989; 80: 421-428.
48. Epstein SE, Palmeri ST, Patterson RE. Evaluation of patients after acute myocardial infarction. *N Engl J Med* 1982; 307: 1487-92.
49. Anonymous. Epidemiology of chest pain and angina pectoris, with special references to treatment needs. *Acta Med Scand*; 1984; Suppl 682: 1-120.
50. Reunanen A, Aroma A, Pyörälä K, Punsar S, Maatela J, Knekt P. The social insurance institution's coronary heart disease study. *Acta Med Scand* 1983; Suppl 673: 1-120.
51. Cock de CC, Visser FC, Fenige van MJ. Prognosis after an infarction: short and long term follow-up data. *Ned Tijdschr Cardiol* 1991; 5: 96-104. (*in Dutch*)
52. NUHI (Nijmeegs Universitair Huisartsen Instituut). Morbidity figures from general practice. NUHI, Nijmegen, 1985.
53. Uemura K. International trends in cardiovascular diseases in the elderly. *Eur Heart J* 1988; 9 (Suppl D): 1-8.
54. Fries J.F. Aging, natural death and the compression of morbidity. *N Engl J Medicine* 1980; 303: 130-135

The new old epidemic of coronary heart disease

Abstract

Objectives: To quantify the changing prevalence of coronary heart disease as a consequence of declining mortality and increasing hospital admission rates in the Netherlands in 1980-93.

Methods: A multi-state life table describes the Dutch population in three disease states (healthy, one or more acute coronary events). The life table uses parsimonious assumptions about disease history to fit observed mortality and registration rates from the nationwide hospital register. The outcome is prevalence by age, sex, time and disease state. Ranges show the results of varying sensitive assumptions.

Results: The prevalence in 1980-93 was 4.4% (range 4.0-4.9) and 1.4% (range 1.3-1.6) for men and women of 25-84 y. Between 1980-83 and 1990-93, the clinical incidence changed little, but age adjusted prevalence increased by 19% (range 10-29) and 59% (range 47-73) for men and women respectively. This increase was more pronounced among the elderly.

Conclusion: In the Netherlands, the sharply decreasing mortality but near constant attack rates of coronary heart disease has caused distinct increases of prevalence. The number of cardiovascular frail elderly is expanding rapidly.

Introduction

Coronary heart disease is the leading cause of death in the Netherlands, as in the rest of the world.¹ As the population is aging, the costs for coronary heart disease are increasing fastly. In 1988, coronary heart disease caused 2.7% of all allocable health care costs in the Netherlands, in 1994 this was 3.1%.²⁻⁴ This relative increase makes coronary heart disease one of the

faster growing causes of medical costs in the Netherlands.⁴ Part of this increase seems to be caused by increasing health care needs. In the Netherlands, mortality rates are decreasing steeply, but hospital admission rates are increasing.⁵ As less patients die suddenly out in the street or at home, more are admitted to the hospital. As more patients survive an acute myocardial infarction, more are at risk of a recurrent infarction or congestive heart failure.⁶

This paper estimates the changing prevalence of coronary heart disease as a consequence of the changing disease history. An incidence-prevalence-mortality model describes a simplified coronary heart disease. In such a model, inflow (incidence) and outflow (mortality) determine pools of surviving patients.⁷ We used available time series of hospital admission rates and mortality from the period 1980-1993 as flows, and calculated the prevalence of patients with a history of an admission for a myocardial infarction by age, sex and calendar period. The incidence-prevalence-mortality model is a reduced, stationary version of a dynamic simulation model published before.⁶ We show that the decreasing coronary heart disease mortality causes the increasing hospital admission rates and we quantify the resulting increase of prevalence.

Methods

Data

The model uses as input numbers of patients discharged with the diagnosis of acute coronary event (ICD 410-411), by age, sex, calendar period (1980-93), and outcome (fatal or non-fatal) and numbers of death from vital statistics by age, sex, calendar period and cause of death.^{5,8,9} Fatal outcomes of hospital admissions are counted as acute coronary deaths in hospital. The hospital register is nationwide. Nearly all recognized acute coronary events are hospitalized^{10,11}, but silent or not recognized myocardial infarctions can not be counted. The causes of death considered are acute coronary heart disease (ICD 410-411), chronic coronary heart disease (ICD 412-414), other cardiovascular disease (ICD 390-409; 415-459) and all other causes of death. By definition, all acute coronary deaths are the sum of all acute in and out hospital coronary deaths; attack rates are the sum of fatal and non-fatal acute coronary events. Age and sex specific rates are calculated by using the midyear population of the Netherlands as person-years.

Model

Figure 1 illustrates the concept. A life table population without heart disease ('healthy') is subjected to an incidence of acute coronary events and two mortality hazards: a risk of dying from cardiovascular disease other than coronary heart disease and a risk of dying from all other causes of death. These (and all other) transition probabilities are age (5 year age

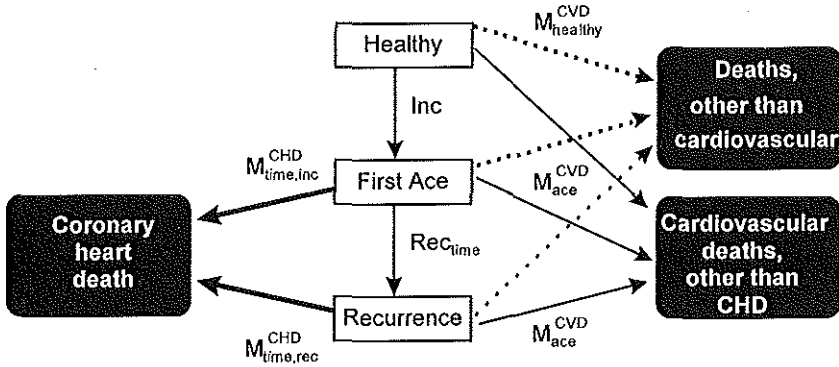


Figure 1 - Coronary heart disease multi-state life table model.

In superscript is the cause of death (all other causes not mentioned). All transition probabilities are age and sex specific. The subscript "time" refers to the period after the event (immediate, first year or subsequent years), the subscript 'healthy', 'inc', 'rec', 'ace' to the status (healthy, first event, recurrent event, or one of both).

groups) and sex specific. At the moment of incidence, patients may die before admission to the hospital, to be counted as a coronary death, or they may die after admission and are counted as a fatal outcome. If they survive, they move to a second life table state, the population with a history of an acute coronary event. Here they are at risk of recurrent acute coronary events and three mortality hazards: a risk of dying from late coronary heart disease, an increased risk of dying from other cardiovascular diseases and a risk of dying from all other causes of death. Risks of recurrence or chronic coronary heart death are higher during the first year after the event. If persons suffer a recurrent acute coronary event, their risk of dying is higher, as compared to an incident event. If they survive, they move to a third life table state, the population with a history of multiple infarctions. Here, they are subjected to the same risks of dying from other cardiovascular and non-cardiovascular causes, but at increased risk of recurrent events and death from acute and chronic coronary heart disease.

Assumptions and sensitivity analysis

There are four dimensions in the history of an acute coronary event. The first is disease history. A history of coronary disease predisposes to new coronary heart disease events and other atherosclerotic disease.¹² The prognosis of such a recurrent event is worse than the first event.¹³⁻¹⁵ The second is age: the prognosis is heavily age dependent, the older the worse.^{13,16} The third is duration: prognosis, compared to a reference population, improves over time.^{12,15,17} The fourth is gender: women have (age adjusted) less coronary heart disease than men, while the fatality rates may be both higher or lower.^{11,12,17,18}

Table 1 shows how these dimensions are quantified. Data about the distribution between first and recurrent acute coronary event at hospital are lacking in the Netherlands, and are reconstructed by using clinical and epidemiological literature. We made three assumptions to estimate recurrence and incidence rates. First, that the proportion of recurrences was 30% (range 25%-35%; upper and lower figures between brackets are alternative assumptions used in sensitivity analysis),^{11,19-23} second, that the risk of a recurrence increased linearly by age^{21,23-25}, third that the risk of a recurrence was 2.5 (range 2-3.5) times higher the first year than the subsequent years.^{12,22} This yields an average risk the first year after the myocardial infarction of 6.5% (4.9%-8.1%) at age 40 and 11.2% (8.5%-14.0%) at age 70, the subsequent years of 2.6% per year (2.0%-3.2%) at age 40 and 4.5% (3.4%-5.6%) at age 70 (see also table 1, recurrence rates). These figures are comparable to estimates previously used in simulation models in the Netherlands and in the USA^{6,26,27}, based on data of the Framingham Heart Study^{12,25}, and other studies, also from the Netherlands.^{28,29}

The prognosis after a recurrent event is worse; we assumed that the odds of death after a recurrent event was 1.8 (range 1.5-2.2) times higher than after a first event.^{14,15,21,23,24,30,31} Survivors may die from chronic coronary heart disease. We assumed that the risk of dying of late coronary heart disease is 5.0 times higher during the first year than during all subsequent years.^{12,25} This figure is relatively high not only because the risk is increased the first year, but also because the risk of being coded 'late coro-

Table 1 - Assumptions used to fit the observed series.

1. Asymptomatic events are not counted.
2. All persons with a symptomatic acute coronary event die or are admitted. ^{10,11}
3. The vital statistics' registration codes reliably acute and chronic coronary heart disease as a cause of death. ^{11,48}
4. Of all discharges for an acute coronary event, 30% (range 25%-35%) are discharges for a recurrent event. ^{11,19-23}
5. The risk of a recurrence increases linearly by age. ^{21,23-25}
6. The risk of a recurrence is 2.5 times higher in the first year after the event than in a subsequent year. ^{12,22}
7. The odds of dying after a recurrent event are 1.8 (range 1.5-2.2) higher than after a first event. ^{14,15,21,23,24,30,31}
8. The risk of dying of late coronary heart disease, and being coded as such, is 5.0 times higher in the first year than in a subsequent year. ^{12,25}
9. The risk of dying from other cardiovascular disease than coronary heart disease is 2.2 (range 1.6-3.3) higher among patients with a history of an admission for acute coronary heart event than a healthy population. ^{25,32}
10. There are no differences among men and women, except the observed differences in the time series of admission, mortality and in-hospital survival
11. The annual risk of a recurrence decreased by 20% (range 0-40%) between 1980-83 and 1990-93. ^{15,27}

nary death' decreases over time. The longer the period between hospital admission and death, the more likely a patient will be coded 'other heart death' (particularly congestive heart failure). This increased relative risk of dying from other cardiovascular diseases is assumed to be 2.2 (range 1.6-3.3) higher than among individuals without CHD history.^{25,32}

With these assumptions and data, the model is fully specified. The incidence of a first acute coronary event, the acute fatality before and after reaching the hospital and the 'late' mortality (coded as chronic coronary heart disease mortality) can now be computed by fitting the calculated mortality and admission numbers on the observed mortality and admissions.

All calculated rates and ratios are age and sex specific. Most assumptions are based on male disease history. Some state that coronary heart disease among females occurs later, but has a worse prognosis (even after correcting for older ages), other studies show the opposite.^{11,12,16,23} Lacking reliable gender specific information, we used identical assumptions for disease history in women.

The period 1980-93 is estimated first, considered as one homogeneous time period. Then the changes in that period are calculated by estimating the periods 1980-83 and 1990-93 separately, using the same assumptions, but the attack and mortality rates of the relevant period only. As it is not very plausible that the probability of recurrence remained constant in the 80s - the use of beta-blocking agents and platelet inhibitors became widespread - we assumed that recurrence rates declined by 20% (0%-40%) between the two periods.^{15,27}

Results

Table 2 and figures 2 and 3 present input and results. All rates and ratios are age adjusted to the Dutch population of 1980-93. Admission rates increased between 1980-83 and 1990-93 by 30% among women and 10% among men, but decreased slightly among younger men (-6%). Mortality rates declined steeply, with an even more pronounced decline among younger men.

Attack rates, the sum of fatal and non-fatal acute coronary events, declined slightly among men (-3%), and increased among women (+7%). Among younger men, the decline of attack rates was more pronounced (-12%). Among older persons, both men and women, attack rates increased by 3 and 6%. This implies that improved survival explains the increase of admission rates at older ages (see figure 2. Nearly all patients which survive such an acute coronary event will be admitted¹¹: lower death rates mean more admissions.

Table 2 - Overview of input data, assumptions and results.

The bold figures are observed, the italic figures are assumptions, based on the literature. Incidence and prevalence are results. Between brackets are assumptions and results from the sensitivity analysis. All estimates are age adjusted to the population of 1980-93.

<i>Mean rates and ratios for 1980-93 (per 1000 persons per year)†</i>								
	age	admission	death	attack	recurrence†	incidence	prevalence	prevalence of recurrence
Men	25-59	2.87	0.66	3.29	<i>83 (63; 104)</i>	2.78 (2.63; 2.92)	17.7 (16.8 ; 18.6)	3.0 (3.5; 2.5)
	60-84	13.56	10.34	19.47	<i>112 (85; 140)</i>	15.13 (13.90; 16.66)	139.1 (122.9; 157.0)	39.5 (41.5 ; 35.5)
	25-84	5.22	2.79	6.85	<i>103 (79; 128)</i>	5.49 (5.11; 5.94)	44.3 (40.1; 49.0)	11.0 (11.8; 9.7)
Women	25-59	0.61	0.15	0.70	<i>91 (64; 105)</i>	0.57 (0.55; 0.61)	3.4 (3.2; 3.6)	0.6 (0.6; 0.4)
	60-84	5.95	5.06	8.77	<i>125 (87; 144)</i>	6.83 (6.60; 7.39)	41.9 (38.9; 48.6)	9.8 (10.1; 8.4)
	25-84	2.10	1.52	2.95	<i>119 (84; 137)</i>	2.32 (2.24; 2.50)	14.1 (13.2; 16.2)	3.1 (3.3; 2.7)
<i>Relative change between 1980-83 and 1990-93 (%)</i>								
	age	admission	death	attack	recurrence	incidence	prevalence	prevalence of recurrence
Men	25-59	-6.0	-42.3	-12.0	<i>-20 (0; -40)</i>	-10.1 (-13.9; - 6.2)	0.0 (-4.1; +4.5)	-9.0 (+4.4; -20.9)
	60-84	22.9	-25.6	3.0	<i>-20 (0; -40)</i>	+5.9 (-1.5; +14.0)	28.7 (+16.7;+42.0)	27.2 (+35.2;+19.3)
	25-84	9.5	-28.9	-2.9	<i>-20 (0; -40)</i>	-0.6 (- 6.5; +5.7)	19.0 (+9.7;+29.0)	18.4 (+27.9; +9.2)
Women	25-59	22.5	-30.3	13.3	<i>-20 (0; -40)</i>	+15.0 (+10.8;+19.4)	34.2(+28.7;+39.9)	27.2 (+46.2;+10.3)
	60-84	32.1	-27.5	6.3	<i>-20 (0; -40)</i>	+1.4 (-3.1; +6.2)	64.8 (+50.6;+80.2)	90.0 (+104.0;+76.7)
	25-84	30.0	-27.7	7.4	<i>-20 (0; -40)</i>	+3.7 (-0.8; +8.4)	59.1 (+46.6;+72.5)	80.3 (+95.3;+66.0)

† Recurrence rates shown are rates the first year after a coronary heart disease event

The calculated incidence and recurrence rates are based on the assumption that 30% (25%-35%) of all admissions are recurrent, and the risk increases by age. The resulting recurrence rates among coronary heart disease patients are consistent with previously used estimates and figures observed in the Netherlands.^{6,26,29} The seemingly higher recurrence rates among women is because women are older.

Clinical incidence changed probably little between '80-83 and '90-93. Only among younger men, the decrease seems quite certain. Among younger women, clinical incidence increases.

The prevalence of patients ever admitted for an acute coronary heart event is three times higher among men than among women; 1 out of 7 Dutch men and 1 out of 24 Dutch women between 60 and 84 have been admitted for an acute coronary event. These prevalences have been increasing recently, predominantly among elderly, caused by the increased survival both of out of hospital and in hospital (see table 1b and figure 3). Over one decade, there are - age standardized - 19% more surviving men and 59% more surviving women with acute coronary heart disease patients to care for (see figure 3); among persons aged 60 and more, prevalence increased by 29% (men) and 65% (women). Among women, the relative changes are more pronounced, but we have to keep in mind that the absolute levels are far lower than among men. The prevalence of recurrence increases in the same order of magnitude, more among the elderly. These will be at the highest risk to develop congestive heart failure.

Sensitivity analysis

The main assumptions in this model concern prognosis of an acute coronary event and the distribution of attack rates between incidence and recurrence. The model is little sensitive to varying assumptions about prognosis, such as the probability of dying of a first versus a recurrent event or the probability of dying from other cardiovascular disease. Because the model fits observed admission and death rates, varying assumptions about prognosis will 'win some and lose some'. If the hazard ratio of death after a recurrent versus an incident event is high, the hazard of death after a first event is relatively low, but more will be at risk of a recurrent event, which will have a worse prognosis. Over all, the duration of disease will not change much, and neither will the calculated prevalence.

The model is sensitive to assumptions about proportions of recurrence. These will redistribute incidences and recurrences: low recurrence means high incidence, and consequently a shorter disease duration to obtain the same death rates. The uncertainty about incidence and recurrence rates dominates the uncertainty of the calculated results.

Validity

Figure 3 shows observed prevalence in the Rotterdam Study of more or less the same period (1990-92).³³ We did not use any figures of the Rotterdam

Figure 2 - Observed admission (acute coronary heart disease) and mortality (coronary heart disease) by gender, age and two periods.

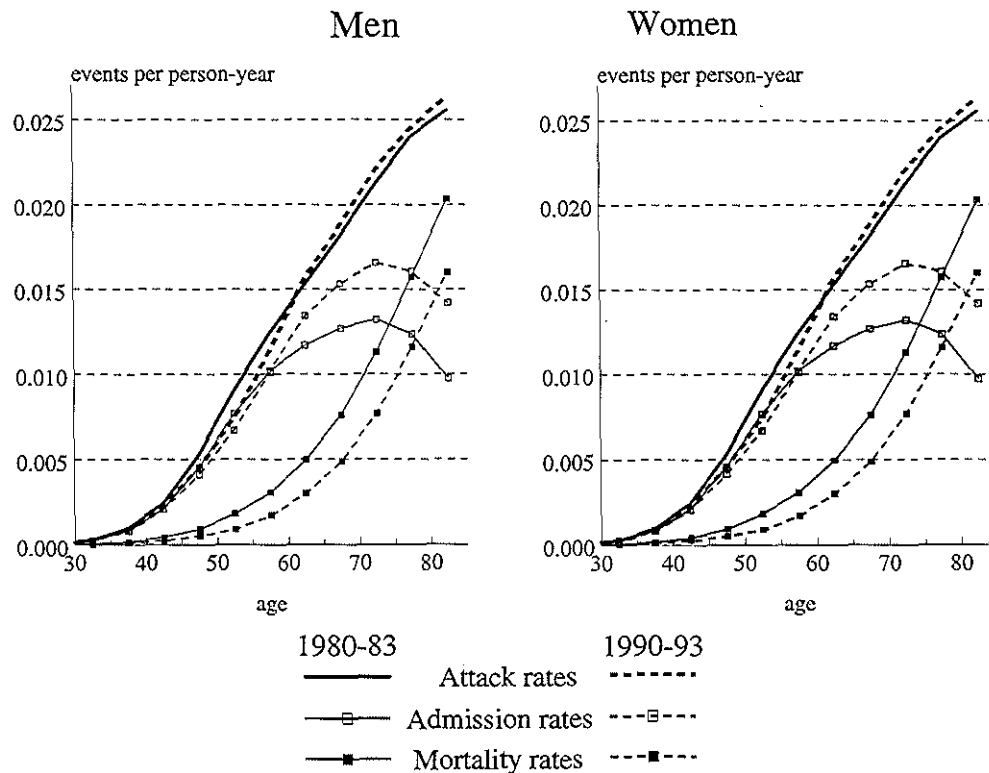


Figure 3 - Calculated prevalence by gender, age and status (history of myocardial infarction or history of recurrent myocardial infarction).

The symbols are observed prevalence and 95% confidence intervals from 1990-93, taken from the Rotterdam Study.³³



Study to specify the model. The close agreement is noteworthy; the somewhat lower prevalence corresponds with the expectations: observed prevalence will lag behind expected figures as a consequence of dynamic effects (see discussion).³⁴ The disappearing decline of prevalence at oldest ages is a characteristic consequence of diminishing mortality pressure: more persons with heart disease survive until older ages. This depends on changes in past flows, and the intermediary position of prevalence in the Rotterdam study between the stationary estimates for 1980-83 and 1990-93 is therefore consistent. The assumption of linearity in the calculation of recurrence rates might cause the lower prevalence in the youngest age group: if the recurrence rates at younger ages are too low, incidence and consequently prevalence is overestimated. Among women the observed age gradient is steeper than expected, with lower prevalence at younger ages. This might be due to somewhat lower coronary heart disease in Rotterdam than in the whole of the Netherlands³⁵, or to a less well specified female coronary heart disease epidemiology.

Discussion

The used model combines available administrative data about admission and death rates with parsimonious assumptions in a stationary multi-state model to calculate expected health care needs for patients with a history of a single or a recurrent acute coronary event. The results suggest expansion of heart disease morbidity in the recent past in the Netherlands. Expected incidence remains nearly constant, mortality declines rapidly. Prevalent patients fill the gap between constant incidence and decreasing mortality, surviving till older ages. These are at risk for further coronary and other cardiovascular disease, and boost health care needs.

These results might be biased by the nature of the data or the assumptions made to specify the model. We used administrative data, not designed for studying the epidemiology of coronary heart disease. But there are no financial incentives to register a particular diagnosis, and trends in myocardial infarction rates of such administrative hospital register data compare well to trends in epidemiologic studies.³⁶ To be registered as an acute coronary event depends from the patients' awareness and doctors' diagnostic acuity and decision between competing choices. Indeed, it is likely that increased awareness of women's risks for coronary heart disease partly causes the increase of clinical incidence among women, particularly among younger women.^{37,38} Doctors (and patients) rarely expect coronary heart disease among younger women. This is certainly not unique for the Netherlands. In the Framingham Heart Study, the prevalence of unrecognized infarction is higher among women than among men.³⁹ The MONICA centers signal a definite correlation between lower admission rates and higher case fatalities among women, suggesting that less severe cases are missed in those centers with low admission rates.¹¹ Among men, however, awareness

has been high, and admission rates correspond well with epidemiological event rates.¹¹

Is there ancillary evidence that the true incidence changed? Surveys of risk factors in the Netherlands show little change over this period.⁴⁰ There is little evidence that the hypertension prevalence decreased recently, no more than that prevalence of treated hypertensives increased in the most recent period.^{41,42} Neither is there evidence that cholesterol increased or decreased in the population in that period.⁴⁰ Smoking prevalence stopped decreasing since the beginning of the eighties.⁴⁰ Indeed, the post world-war female birth cohorts, who took up smoking in large numbers in the sixties and are now arriving in middle age, might cause part of the suspected increase in coronary heart disease among younger women.⁴³ Lung cancer rates are increasing rapidly among Dutch women.⁴⁴ Denmark signaled the same rapid increases of female lung cancer and a relative increase of coronary heart disease among women.^{45,46} While part of the increase in incidence among women might be caused by increasing detection of more benign cases, part of it seems real, caused by the evolving smoking epidemic among women of the baby boom generation.

The mortality register is reasonably valid for coronary heart disease from a population perspective.¹¹ Many studies, both observational and theoretical, have observed the coronary heart disease mortality in the developed world, and linked to improving treatment.^{5,6,15,17,22,23,27,47} Only changes in coding practice might have introduced artefactual changes over time periods. While they can't be excluded, there is no evidence of such changes.

With a constant incidence, and a decreasing mortality prevalence has to expand inevitably. Our life table calculations show impressive increases, particularly among the elderly. These changes are calculated by stationary life tables, assuming a steady state. Prevalence is a 'stock' variable, built up by incidence and mortality flows from the past.³⁴ The calculated prevalence assumes that the in- and outflows remained constant in the past. In a dynamic population, the true prevalence will lag behind the calculated prevalence from a steady state model, as it takes time to build up prevalence. The true prevalence will be reached only if the present rates would remain unchanged long enough (and the condition of a steady state is fulfilled).

How do these results compare with previous results?²⁶ Qualitatively, the results of the previous paper remain intact, but quantitatively the present results, showing the prevalence increases in the recent past, are more precise, building on more data and fewer assumptions in a simpler model. Because the incidence changes were lower in the Netherlands and the mortality changes higher than previously expected, the prevalence increase is higher than previously expected. Particularly the increase among women of the large post world war birth cohorts, consistent with a lower mortality decline and high historical smoking uptake rates, is an ominous sign, which may indicate that the gender gap is closing, but from the wrong end.

We conclude that in the Netherlands, the recent history of coronary heart disease shows expansion of morbidity. Life expectancy free of coronary disease does not change, only the life expectancy with disease increases. The recent, steep mortality decline dominates the disease epidemiology. While there is circumstantial evidence of declining incidence among younger men, the same evidence suggests increasing incidence among younger women. The increasing prevalence is probably bought by successful health care interventions. As many survivors will lead satisfactory lives, this is a great success of modern (heart) health care. However, the lack of a recent decline in acute coronary incidence, attributable to unchanging smoking prevalence in the eighties, and the recent increase of smoking among youth⁴⁰, show that there is little place for complacency. The marked decline of mortality has resulted in large pools of prevalent survivors. As soon as the therapeutic progress will halt, even temporarily, we suspect that the coronary heart disease mortality will start increasing again, particularly among the elderly, as a result of 'postponed' death.

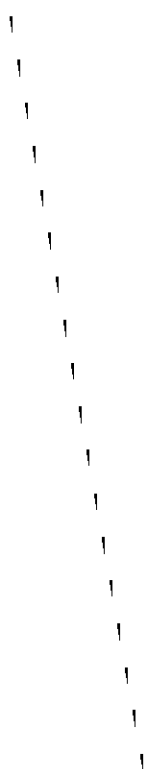
References

1. Murray CJL, Lopez AD. Mortality by cause for eight regions of the World: Global burden of disease study. *Lancet* 1997;349:1269-76.
2. Koopmanschap M, L van Roijen, Bonneux L. *Costs of diseases in the Netherlands [in Dutch]*. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Instituut voor Medische Technology Assessment, 1991.
3. Koopmanschap M, van Roijen L, Bonneux L, Bonsel G, Rutten F, van der Maas P. Costs of diseases in international perspective. *Eur J Public Health* 1994;4:258-64.
4. Polder J, Meerding JW, Koopmanschap MA, Bonneux L, Van der Maas P. *Costs of Diseases in 1994 (in Dutch)*. Rotterdam: Department of Public Health, 1997.
5. Bonneux L, Looman CW, Barendregt JJ, Van der Maas PJ. Regression analysis of recent changes in cardiovascular mortality and morbidity in the Netherlands. *Br Med J* 1997;314:789-92.
6. Bonneux L, Barendregt J, Meeter K, Bonsel G, van der Maas P. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. The future rise of heart failure. *Am J Publ Health* 1994;84:20-29.
7. Murray CJL, Lopez AD. Global and regional descriptive epidemiology of disability (chapter 4). In: Murray CJL, Lopez AD, editors. *The global burden of disease and injury series*. Harvard: Harvard University Press, 1996:201-246.
8. Reitsma JB. *Cardiovascular diseases in the Netherlands [In Dutch]*: Netherlands Heart Foundation, 1995.
9. Statistics Netherlands. *Deaths by causes of deaths, age and gender*. Voorburg (Netherlands): Central Bureau of Statistics, 1994.

10. Fracheboud J. *Coronary Center Care of Home Care? A descriptive study of home care of patients with an acute myocardial infarction in the Netherlands [in Dutch]*. Utrecht: NIVEL, 1987.
11. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 1994;90:583-612.
12. Kannel W, Sorlie P, McNamara P. Prognosis after initial myocardial infarction: The Framingham study. *Am J Cardiol* 1979;44:53-9.
13. Ahnve S, Gilpin E, Dittrich H, et al. First myocardial infarction: age and ejection fraction identify a low-risk group. *Am Heart J* 1988;116:925-32.
14. Benhorin J, Moss A, Oakes D, the multicenter diltiazem postinfarction trial research group. Prognostic significance of nonfatal myocardial reinfarction. *J Am Coll Cardiol* 1990;15:253-8.
15. McGovern PG, Pankow JS, Shahar E, Doliszny KM, Folsom AR, Blackburn H, et al. Recent trends in acute coronary heart disease--mortality, morbidity, medical care, and risk factors. The Minnesota Heart Survey Investigators. *N Engl J Med* 1996;334:884-90.
16. Kober L, Torp-Pedersen C, Ottesen M, Rasmussen S, Lessing M, Skagen K. Influence of gender on short- and long-term mortality after acute myocardial infarction. TRACE study group. *Am J Cardiol* 1996;77:1052-6.
17. Torp-Pedersen C, Hildebrandt P, Kober L, Nielsen FE, Jensen G, Melchior T, et al. Improving long-term survival of patients with acute myocardial infarction from 1977-1988 in a region of Denmark. *Eur Heart J* 1995;16:14-20.
18. Kober L, Torp-Pedersen C, Ottesen M, Burchardt H, Korup E, Lyngborg K. Influence of age on the prognostic importance of left ventricular dysfunction and congestive heart failure on long-term survival after acute myocardial infarction. TRACE Study Group. *Am J Cardiol* 1996;78:158-62.
19. Fiorretti P, Brower R, Simoons, et al. Relative value of clinical variables, bicycle ergometry, rest radionuclide ventriculography and 24 hour ambulatory ECG monitoring at discharge to predict 1 year survival after myocardial infarction. *J Am Coll Cardiol* 1986;8:40-9.
20. Meeter K, Honkoop P, Verhage A, Boersma H, Fioretti P, Deckers J. Treatment of myocardial infarction during the hospitalization phase and a short period afterward: now and 10 years ago [in Dutch]. *Ned Tijdschr Geneesk* 1993;137:1922-6.
21. Boersma H, van der Vlugt MJ, Arnold AER, Deckers JW, Simoons ML. Estimated gain in life expectancy. A simple tool to select optimal reperfusion therapy in individual patients with evolving myocardial infarction. *Eur Heart J* 1996;17:64-75.
22. De Vreede JM, Gorgels APM, Verstraaten GMP, Vermeer F, Dassen WRM, Wellens HJJ. Did prognosis after acute myocardial infarction change during the past 30 years? A meta-analysis. *J Am Coll Cardiol* 1991;18:698-706.
23. Stevenson R, Ranjadayalan K, Wilkinson P, Roberts R, Timmis AD. Short and long term prognosis of acute myocardial infarction since introduction of thrombolysis. *Br Med J* 1993;307:349-53.

24. Marcus F, Friday K, McCans J, et al. Age related prognosis after acute myocardial infarction. The multicenter diltiazem postinfarction trial research group. *Am J Cardiol* 1990;65:559-66.
25. Califf RM, Armstrong PW, Carver JR, RB DA, Strauss WE. 27th Bethesda Conference: matching the intensity of risk factor management with the hazard for coronary disease events. Task Force 5. Stratification of patients into high, medium and low risk subgroups for purposes of risk factor management. *J Am Coll Cardiol* 1996;27:1007-19.
26. Weinstein M, Coxson P, Williams L, Pass T, Stason W, Goldman L. Forecasting coronary heart disease incidence, mortality and cost. The coronary heart disease policy model. *Am J Public Health* 1987;77:1417-26.
27. Hunink MGM, Goldman L, Tosteson ANA, Mittleman MA, Goldman PA, Williams LW, et al. The recent decline in mortality from Coronary Heart Disease, 1980-1990. *JAMA* 1997;277:534-42.
28. Sanz G, Castañer A, Betriu A, et al. Determinants of prognosis in survivors of myocardial infarction. *N Engl J Med* 1982;306:1065-70.
29. Simoons L, Vos J, Tijssen J, et al. Long term benefit of early thrombolytic therapy in patients with acute myocardial infarction. *J Am Col Cardiol* 1989;14:1609-15.
30. Davis H, DeCamilla J, Bayer L, Moss A. Survivorship patterns in the post-hospital phase of myocardial infarction. *Circulation* 1979;60:1252-8.
31. Fioretti P, Tijssen J, Azar A, et al. Prognostic value of predischARGE 12 lead ECG after myocardial infarction compared with other routine variables. *Br Heart J* 1987;57:306-12.
32. Wong N, Cupples A, Ostfeld A, Levy D, Kannel W. Risk Factors for long term coronary prognosis after initial infarction. The Framingham Study. *Am J Epid* 1989;469-80.
33. Grobbee DE, Van der Bom JG, Bots ML, De Bruijne MC, Mosterd A, Hoes AW. Cardiovascular disease among elderly people. The "Rotterdam elderly study" [in Dutch]. *Ned Tijdschr Geneesk* 1995;139:1987-82.
34. Barendregt J, Bonneux L, van der Maas PJ. Health expectancy, an indicator for change? *J Epid Com Health* 1994;482-86.
35. Mackenbach J, Looman C, Kunst A. Geographic variation in the onset of decline of male ischemic heart disease mortality in the Netherlands. *Am J Public Health* 1989;79:1621-7.
36. Mahonen M, Miettinen H, Pyorala K, Molarius A, Salomaa V, Kuulasmaa K. Hospital discharge register data in the assessment of trends in acute myocardial infarction. FINMONICA AMI Register Study Team. *Ann Med* 1995;27:547-54.
37. Khaw KT. Where are the women in coronary heart disease? *Br Med J* 1993;306:1145-46.
38. Healy B. The Yentl syndrome. *N Engl J Med* 1991;325:274-76.
39. Lerner D, Kannel W. Patterns of coronary heart disease morbidity and mortality in the sexes: A 26-year follow-up of the Framingham population. *Am Heart J* 1986;111:383-9.
40. Ruwaard D, Kramers PGN, Van den Bergh Jeths A, Achterberg PW, editors. *Public health status of the Dutch Population over the period 1950-2010*. The Hague: SDU, 1994.

41. Van de Mheen PJ, Bonneux L, Gunning-Schepers LJ. Variation in reported prevalences of hypertension in The Netherlands: the impact of methodological variables. *J Epidemiol Community Health* 1995;49:277-80.
42. Bonneux L, Van de Mheen P, Gunning-Schepers LJ, van der Maas PJ. Hypertension in the Netherlands, did it decline between 1974 and 1993? [In Dutch]. *Ned Tijdschr Geneesk* 1996;140:2603-2606.
43. Gunning-Schepers L. The health benefits of prevention: a simulation approach. *J Health Policy*. 1989;12/1-2 (special issue):93-129.
44. Janssen-Heijnen ML, Nab HW, van Reek J, van der Heijden LH, Schipper R, Coebergh JW. Striking changes in smoking behaviour and lung cancer incidence by histological type in south-east Netherlands, 1960-1991. *Eur J Cancer* 1995;31A:949-52.
45. Hildebrandt P, Jensen G, Kober L, Torp-Pedersen C, Joen T, Ege M, et al. Myocardial infarction 1979-1988 in Denmark: secular trends in age-related incidence, in-hospital mortality and complications. *Eur Heart J* 1994;15:877-81.
46. Engholm G, Palmgren F, Lynge E. Lung cancer, smoking, and environment: a cohort study of the Danish population. *BMJ* 1996;312:1259-63.
47. Sytkowski P, Kannel W, D'Agostinho R. Changes in risk factors and the decline in mortality from cardiovascular disease. *N Engl J Med* 1990;322:1635-41.
48. Mackenbach JP, van Duyne WMJ, Kelson MC. Certification and coding of two underlying causes of death in the Netherlands and other countries of the European Community. *J Epid Comm Health* 1987;41:156-60.



Stroke trends in an aging population

Abstract

Background and Purpose: Trends in stroke incidence and survival determine changes in stroke morbidity and mortality. This study examines the extent of the incidence decline and survival improvement in the Netherlands from 1979-1989. In addition, it projects future changes in stroke morbidity over the period 1985-2005 when the country's population will be aging.

Methods: A state-event transition model is used, which combines Dutch population projections and existing data on stroke epidemiology. Based on the clinical course of stroke, the model describes historical national age- and gender-specific hospital admission and mortality rates for stroke. It extrapolates observed trends and projects future changes in stroke morbidity rates.

Results: There is evidence of a continuing incidence decline. The most plausible rate of change is an annual decline of -1.9% (range: -1.7 to -2.1) for men and -2.4% (range: -2.3 to -2.8) for women. Projecting a constant mortality decline, the model shows a 35% decrease of the stroke incidence rate over a period of twenty years. Prevalence rates for major stroke will decline among the younger age groups but increase among the oldest due to increased survival in the latter. In absolute numbers this results in an 18% decrease of acute stroke episodes and an 11% increase of major stroke cases.

Conclusions: The increase in survival cannot fully explain the observed mortality decline and, therefore, a concomitant incidence decline has to be assumed. Aging of the population partially outweighs the effect of an incidence decline on the total burden of stroke. Increase in cardiovascular survival leads to a further increase in major stroke prevalence among the oldest age groups.

Introduction

The dynamics of stroke morbidity and mortality are of major interest for clinicians as well as for epidemiologists and health policy makers. A changing stroke epidemiology results from changes in incidence and survival. The balance between these trends determines the numbers of short and long term stroke survivors within a population. Recently, the debate on the relative contribution of trends in incidence and survival to stroke mortality decline has intensified, complicated by different study methods and inconsistent results¹⁻⁴.

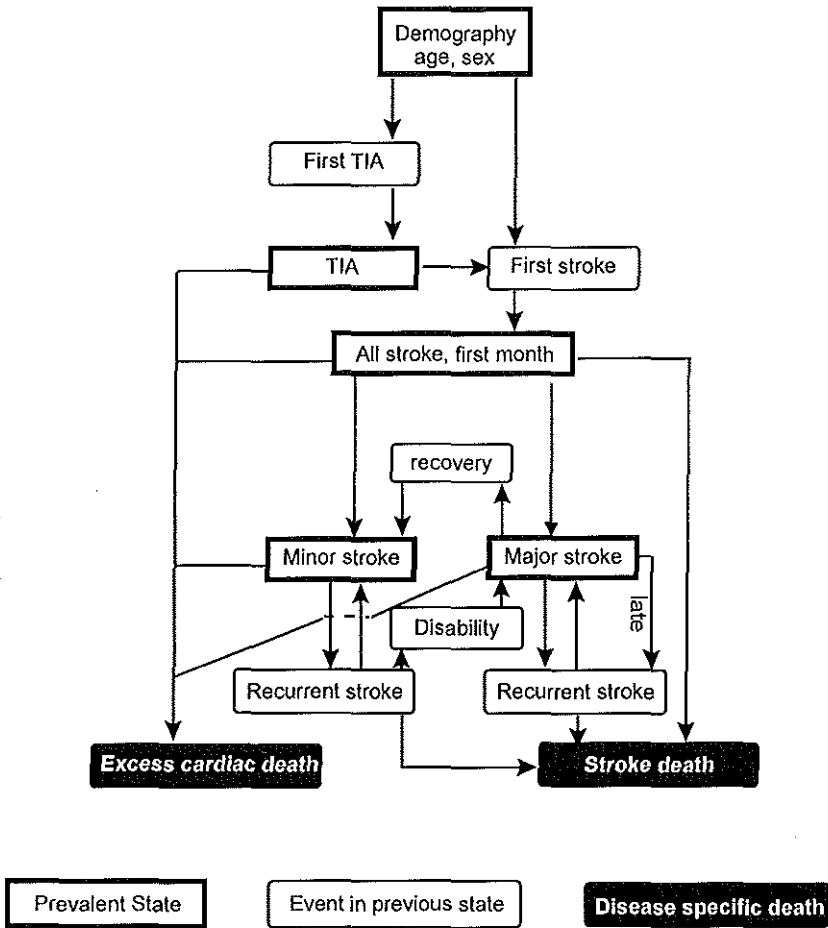
In the Netherlands as well as in the USA, stroke mortality has been declining for all age groups since the early sixties^{1,3-6}. In the Netherlands, the age-adjusted decline from 1979-1989 has been a constant 3.1 % per year for men and 4.0% for women, while in the USA mortality decline has been 5.7% and 5.2%, respectively^{3,6}. An incidence decline has also been observed^{7,9}, ascribed to better hypertension control and a decline in smoking prevalence^{7,8,10}. Observed incidence trends are confounded by the introduction of computerized tomography, improving the specificity of the diagnosis but also increasing case finding. Declines in short- and long-term case fatality have been documented for the last decades^{7,9,11-13}, explained mainly by a better prognosis after intra-cerebral hemorrhage, by increased hypertension control^{7,10} and by a better prevention and treatment of complications, especially of cardiac disease. However, the observed mortality decline has started long ago and cannot be fully accounted for by observed changes in risk factors. Most likely, both incidence and case-fatality decline will remain largely unexplained^{3,13}.

This study determines the most plausible range of incidence and fatality decline that explains the impressive observed reduction in stroke mortality in the Netherlands by means of a state-event model¹⁴. In addition the model is applied to project future changes in stroke incidence and prevalence using the calculated trend values.

Material and methods

Stroke mortality trends are determined by changes in stroke incidence, survival, recovery, recurrence, and mortality from other diseases. Given this complexity, a mathematical model is indispensable¹⁴. We developed a state-event model that is based on clinical course of stroke (figure 1). Combining data from various sources, the model describes the epidemiology of stroke in the Netherlands.

The basic principle of a state-event model is that patients move from one particular state to another after experiencing a particular event.^{14,16} The likelihood to move from one state to another, a transition probability, is independent from the preceding states or events and depends only on the current state defined by disease stage, age and gender during the event. All probabilities are age and gender specific. There are five year age groups ranging from 25 years to 90 and over. The model combines a demographic component,



Note: Death from other causes occurs in all prevalent states, at any time
TIA is transient ischemic attack

Figure 1 - Diagram presenting the state-event model for stroke.

First year states and states for all subsequent years have been combined to facilitate the depiction of the model.

containing the most likely national population projections¹⁶, and a stroke specific component. Risk of death from other causes is accounted for in all states.

During computation the model annually generates first incident cases from the demographic component. These enter the respective states within the stroke specific component and follow the various flows with the model. Simultaneously, the model also annually updates all existing prevalent states for recurrences and their consequences.

Within the stroke disease component of the model two events can occur: a transient ischemic attack (TIA) or a stroke, both defined as in the Oxford Community Stroke Project (OCSP).^{17,18} We distinguished different states for the first year and for all subsequent years together. Because the flows in the model during the first year and the subsequent years are almost identical, these two year states for each condition are depicted together to clarify the presentation. In both TIA states the patient has an increased risk of stroke. After the first year, patients with a history of a TIA enter the "subsequent years" state and run a lower stroke risk. A separate first month state after a first stroke allows for the acute phase with a high risk of disability and death. After a first stroke patients enter this state. Patients surviving the first month are left either with a minor or major stroke as defined by a Rankin grade 0-2 and 3-5¹⁹ respectively, and are divided up between the two separate states. Patients with a history of stroke run a risk of recurrence. If this occurs, there is an excess risk of dying or having a major stroke. In the first year state for major stroke, some patients recover, defined by Rankin grade 0-2. The recovered patients enter the minor stroke state. The remaining fraction moves to the major stroke state for all subsequent years. In both major stroke states a patient suffers a delayed death due to a first and disabling stroke.

One model assumption is that in the acute phase almost all deaths can be contributed to the first stroke and only a few to the other causes of death. In this phase, the risk of recurrence and the excess risk of death from heart disease are not accounted for due to an absence of recorded data. Also, a single state for all subsequent years together implies that the recurrence risk in subsequent years state is the same for all following years. This is supported by recent Dutch data.²⁰ In addition, we do not distinguish strokes caused by intracerebral hemorrhage and those caused by cerebral infarction. From a patient-based view, these are different. At the aggregated population level a distinction is less useful as thrombotic infarctions make up more the 80% of all the stroke cases.^{17,24} Moreover, the survival after a hemorrhagic strokes is reaching the level of survival after an infarction as a result of improved prognosis and increased detection of smaller and less harmful bleedings.²²

The origin of the crude data used to calculate the base-line input is summarized in table 1. We have calculated the age-specific transition probabilities using the results of Dutch population-based studies, if available. If incomplete, they have been used to check selected comparable figures of other white populations as listed in the table. Relative risks are used when comparing risks for one patient category with another. Ratios are used as transition probabilities without further calculations. The choice of measure depends on the way the data have been made available.

The risk of a first TIA is calculated using the incidence figures from the OCSP¹⁸. Age-specific probabilities are calculated by exponential interpolation and are comparable to data from Dutch primary care practices (NUHI)²³ and also the Rochester study²⁴. The relative risks of stroke after a first TIA reported by the OCSP³⁰ have been interpolated and are multiplied by the population risks from the Tilburg Epidemiological Study of Stroke (TESS)

Table 1

Crude literature data used to calculate base-line transition probabilities within the stroke model and resulting transition probabilities for the group of patients aged 70-74 years.

	Crude data		Probability (70-74 y)	Source
	Measure	Figure		
First TIA	Rate	0.42/1000	0.0037/0.0026	NUHI ²¹ ; OCSP ¹⁸
First stroke	Rate	1.62/1000	0.012/0.010	TESS ²¹ ; OCSP ¹⁷
Death from first stroke	Ratio	0.20	0.21/0.21	TESS ²⁴ ; OCSP ¹⁷
Major disability after first stroke	Ratio	0.39	0.39	DGP ²⁵ ; OCSP ¹⁷
Recovery from major stroke	Ratio	0.76	0.22	DGP ²⁵ ; New Zealand ²⁶
Stroke after TIA/minor stroke	RR			NUHI ²³ /OCSP ³⁰ / Dutch TIA Trial ²⁰
first year <75 years		13.2	0.09/0.07	
first year >75 years		4.5		
subsequent years		5.0	0.05/0.04	
Late death from major stroke	Ratio MHR*	0.17/0.42 1.261	0.15/0.11	Dutch Registry ⁶ N Carolina ^{27,28}
Cardiac death after TIA or stroke	RR MHR**	3.2 0.665	0.038/0.025	Dutch Tia Trial ²⁰ N.Carolina ^{27,28}
Cardiac death after major stroke	RR#	1.57	0.06/0.04	N.Carolina ^{27,28}

TIA, transient ischemic attack; RR, relative risk; MHR, multivariate hazard ratio.

* risk comparison: age (X+10)/X

** risk comparison: women/men

risk comparison: major/minor stroke

and the OCSP (see below)^{17,21} to calculate age-specific absolute stroke risks after a TIA.

Probabilities of a first stroke are calculated by averaging the results of the incidence studies by the OCSP and the TESS that have produced similar figures^{17,21}. Because the TESS included fewer age groups, the results of both studies are combined to have reliable incidence figures for as many age groups as possible. Results are consistent with the NUHI data^{14,21}. Age-adjusted all stroke case-fatality rates from the TESS study have been corrected for recurrences by assuming a double recurrence death hazard. The resulting case-fatality ratios of the OCSP and TESS are nearly identical (19% and 20%). The probability of residual major disability after one month is assumed to be constant for age and gender.²⁶

The recurrence risk for minor stroke patients is assumed to be the same as the first stroke risk for TIA patients. The ratio of the risk of dying after a recurrence versus the risk after a first ever stroke is 1.5.²⁷ For the risk of a major stroke after a recurrence the same ratio is used. Recurrence risk in the

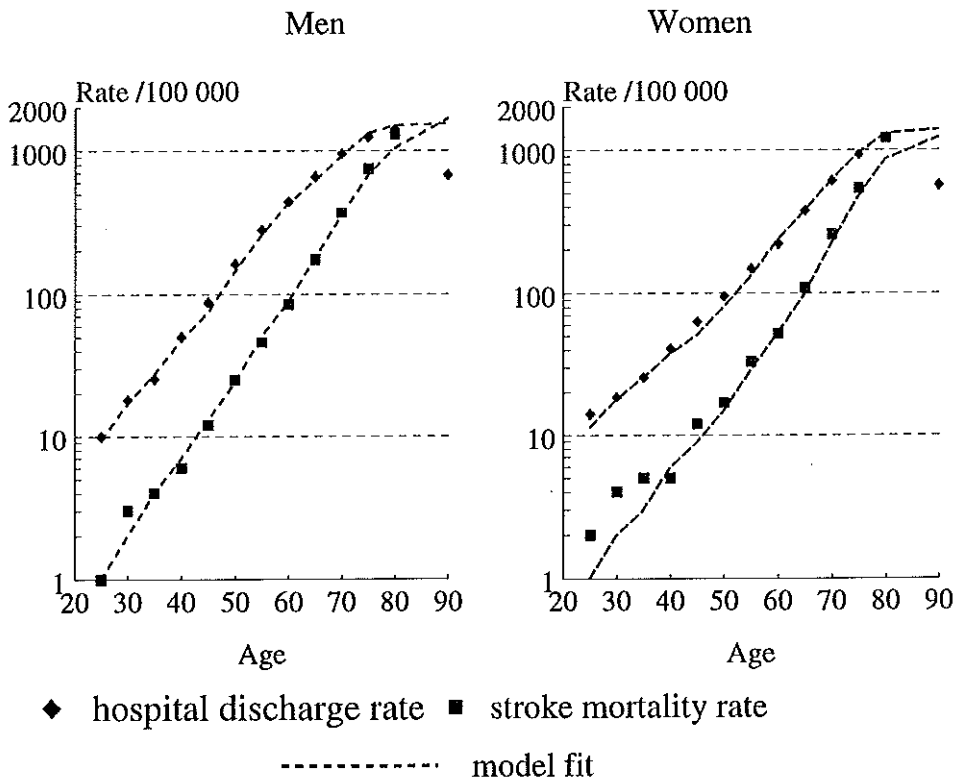


Figure 2

Logarithmic graph showing age-specific admission rates and death rates in the Netherlands in 1985 for all cerebrovascular accidents (ICD codes 430-438, except 435) according to the national registries^{9,28} and as calculated by the state-event model for stroke. Model admission rates are based on the model events rates for all strokes together multiplied by age-specific chances of hospital admission found in a Dutch study.²⁸

subsequent years state after a minor stroke is estimated half the first year risk,²⁰ with the same risk of dying and of major stroke, given the recurrence, as during the first year.²⁸

The probability of recovery from major stroke has been calculated by age group and is the same for both sexes.²¹ In agreement with various studies^{9,21,27,29} recurrence risk and subsequent death after major stroke has been estimated by doubling the hazard ratio of the same parameter as found by Howard et al²⁷ for the minor stroke group. Late stroke mortality during subsequent years is half the risk in the first year state. The recurrence risk in the subsequent years state after a disabling stroke is half the first year risk^{8,10} as assumed for minor stroke.

The excess age-specific risk of death from ischemic heart disease (ICD 410-414) in the prevalent states has been calculated by multiplying the hazard ratios for cardiac death^{27,28} with the age-specific risk of death from ischemic heart disease for the general Dutch population.⁶

Risk of death from other causes is calculated using the all causes death rates from the national death registry corrected for the stroke related figures (ICD codes 430-438).

The stroke disease component of the model in combination with the demographic component describes the stroke epidemiology for the Dutch population in steady state. So far it has been presented with fixed transition probabilities. The model estimates stroke prevalence for the baseline year (figure 2) after assuming the same transition probabilities during the preceding years.

The model allows for age-specific time trends for all the transition probabilities depicted in figure 1. This allows for plausible projections over longer periods. The value of these trends are calculated by time series analysis of available figures from comparable populations. A trend is defined as the annual percent change, which means an exponential change. To calculate this kind of trend, first, a log-transformation is applied. The regression line, through the log-transformed figures of each time interval is determined by a least squares fit. The regression coefficient, or slope, of this line, is the annual percent change of the time series figures.⁵

During computation two sets of trends are used. One set consists of all "attack" parameters: the risks of a TIA, first stroke or recurrence. The other set includes all parameters regarding acute and late case-fatality of stroke accounting for the decreasing severity of stroke. This study focuses on the first stroke incidence and case-fatality trends.

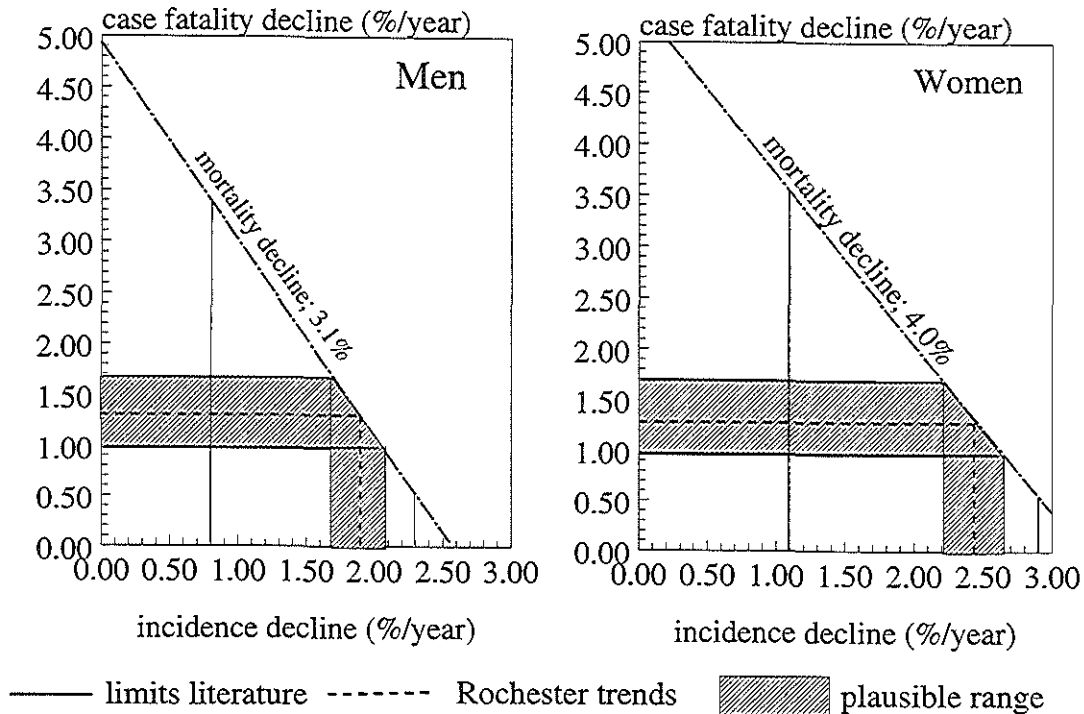
These trends in transition probabilities are not very well documented. The Rochester study⁷ is the only study that produces age and gender specific data on the secular changes both in stroke incidence and fatality. We applied the Rochester incidence and case-fatality trends to the respective attack and case-fatality sets of trends within the model. This can be done as the Dutch incidence figures for TIA and stroke as well as case-fatality ratios over a single two-year period agree with comparable Rochester data.^{21,23,24} We ignored the recent, most likely temporary, incidence increase for Rochester caused by increased case-finding because of the introduction of computed tomography.⁷ Trends in the two remaining transition probabilities, the risk of major stroke after stroke and the chance of recovery from a major stroke are not known and are assumed constant. The application of the Rochester incidence and

Table 2 - Annual incidence and case-fatality trends applied in the stroke model and goodness-of-fit of the corresponding model mortality trends with the empirical Dutch national mortality trend for 1979-1989. All χ^2 values indicate a good fit ($P > 0.5$, $df=10$).

Scenario	Men			Women		
	Annual trend (%)		Fit	Annual trend (%)		Fit
	Incidence	Case fatality	χ^2	Incidence	Case fatality	χ^2
Upper	-1.7	-1.7	1.02	-2.3	-1.7	1.08
Rochester	-1.9	-1.3	0.46	-2.4	-1.3	0.55
Lower	-2.1	-1.0	0.71	-2.8	-1.0	0.72

Figure 3

Plots of two-way sensitivity analysis of two variables: the stroke incidence trend and the case-fatality trend. The iso-mortality lines shown are for Dutch men (3.1%) and women (4.0%) during the period 1979-1989. Rochester trends for men and women are plotted as well as the most extreme values for both trends found in the literature. The plausible range of incidence decline (the grey area) falls within the range of values reported in the literature.^{7,10,30,33}



survival trends appear to reproduce almost identical sex-specific mortality trends as observed in the Netherlands from 1979 to 1989 (see table 2).^{5,6} This seems logical as both populations are mainly white and comparable in most other aspects. Other combinations of trend values for incidence and case-fatality, however, can also account for the observed mortality decline.^{8,12,16,31-33} The difficulty is that trend values are reported mostly separately from each other and without corresponding overall mortality trend.

To solve this we used a two-way sensitivity analysis of incidence and case-fatality trends to determine a plausible ranges of values (fig 3). In the analysis, for each value of one set of trends we calculated the corresponding value of the other set that, in combination, leads to the same mortality decline. Under the condition of a fixed mortality decline, the incidence and case-fatality trends are inversely proportional. When the incidence rates decrease, the case fatality rates have to increase five times as rapidly to outweigh the decreasing mortality decline. The main reason for this is that the case-fatality after a first stroke is 20%. This inverse relation leads to the isomortality lines in figure 3. The area above the lines includes those values of trends that lead to a larger mortality decline. The area below includes the values that lead to an lesser mortality decline. The range of reported values for case-fatality trends is small.^{9,12} As a result, it defines a much narrower

Table 3

Estimated standardized stroke rates (European population as direct standard) and absolute numbers of stroke cases in the Netherlands for 1985 and projected future changes for the year 2005 of standardized stroke rates and absolute numbers of stroke cases relative to 1985, assuming continuing Rochester-like trends and population projections by Dutch Central Bureau for Statistics¹⁶.

	1985		2005 (% change)	
	Rates	Numbers	Rates	Numbers
<i>Men:</i>				
First strokes	171	10600	-35	-16
All strokes	224	14000	-37	-19
First major stroke	67	4210	-34	-14
Prevalence major strokes	275	17300	-15	+11
Prevalence minor strokes	700	43900	-12	+15
Stroke deaths	68	4170	-46	-30
<i>Women:</i>				
First strokes	128	11200	-36	-9
All strokes	159	15000	-38	-17
First major stroke	46	4260	-35	-15
Prevalence major strokes	200	18400	-17	+8
Prevalence minor strokes	466	41300	-13	+11
Stroke deaths	51	5110	-45	-23

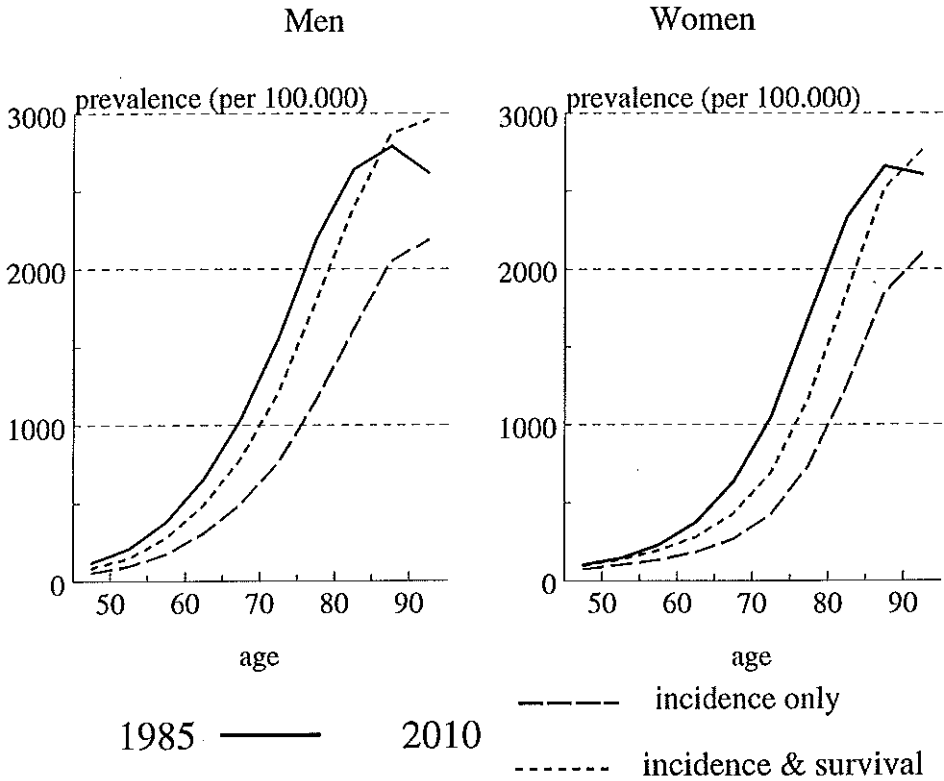


Figure 4

Graphs showing the age-specific prevalence rates of major stroke for 1985 and 2010, after computation for a period of 20 years. For the 2010 projection, the results of two runs are shown. During the first run, only the Rochester incidence trend (a decline) is applied to all "attack" parameters. During the second run, the Rochester case-fatality trend (also a decline) on all stroke case-fatality parameters is added.

range of values for a possible incidence decline than is reported in the literature.

Statistical testing of the results of the computations is done by comparing model results with the observed national data. A χ^2 goodness-of-fit test is used (see appendix).³⁴ A value of $P > 0.5$ for this test indicates a good fit of the computed results and the observed data. The goodness-of-fit of the model age-specific stroke mortality rates and stroke admission rates (first and recurrent strokes together) is given as well that of the trends in the model mortality rates.

Results

The stroke model produces nation-wide projections regarding the stroke epidemiology including first stroke incidence, stroke recurrences as well as stroke mortality and prevalence. Figure 2 shows the model outcome regarding national stroke hospital rates, including both fatal and non-fatal cases, and national stroke mortality rates for the 1985 baseline year. There is a good fit for the age groups up to 80 years between model outcomes and the national data of the same year⁶: the χ^2 for men is 8.10 and 6.37, respectively and for women is 6.21 and 8.15 (all $P > 0.5$, $df = 11$). These χ^2 values support the validity of the model. The national admission rate²⁴ for the 85 year old and over from the hospital registry is lower than the model admission rate, as the latter refers to the age group 85-89 only. The lower model stroke deaths rates of the same age group are most probably due to an overregistration of stroke deaths for this age group in general practice and/or a possible underregistration of incidence in the OCSP and TESS.

Table 3 lists the aggregated model results for 1985, and the relative change of output results for year 2005. Here trend values are used assuming a continuing Rochester-like scenario and therefore a continuing mortality decline. All rates are decreasing and there are no major sex differences. As all attack rates, including risk of recurrence, are assumed to decrease, all stroke rates decrease more than first stroke rates. Because of improved survival the drop in prevalence rates is considerably less. The projected decline in stroke death rate, is the same as during the past twenty year in the Netherlands.^{5,6} This has been the basic assumption of this projection.

The effect of the aging of the post-war baby boom is shown in the shifts in absolute numbers: less decline in all-stroke cases and an increase in prevalent cases. The longer life expectancy of Dutch women is reflected in a smaller decrease of first stroke cases and stroke deaths. In general, the incidence decline outweighs the expected increase in absolute numbers of acute stroke episodes with the aging of the post-war generation. Nevertheless, increased survival increases the absolute number of prevalent cases considerably.

Resulting prevalence rates for major stroke are given in figure 4 and agree with the population-based rates found in Finland and Rochester.³⁰ The same figure also demonstrates the trend dynamics by a step-wise inclusion of the two "attack" and "case-fatality" sets of trends that are based on the time-series analysis of the Rochester data. A decreasing incidence of stroke results in a decreasing major stroke prevalence and also a decreasing minor stroke prevalence (data not shown). This decreasing effect on prevalence is nearly halved by an increase in survival, especially of major stroke patients and, to a lesser extent, of minor stroke and TIA patients who live longer with the risk of suffering a debilitating stroke. The effect of survival improvement on stroke prevalence increases with age. This results in an increase in stroke prevalence among the older age groups.

The two-way sensitivity analysis determines the plausible range of values for the incidence trend that explains Dutch stroke mortality decline, given the reported survival improvement. Figure 4 shows that the values for the incidence trend range between 1.7% and 2.1% per year for men, and between 2.3% and 2.8% per year for women. The figure also demonstrates that, if one supposes no incidence decline⁷ an unreported annual improvement of survival of over 5% would be necessary to effect the observed mortality decline. On the other hand, an incidence decline of more than 3% for men would imply an unreported absence of survival improvement or even a deterioration. In Table 2, the results of a goodness-of-fit test of the model mortality trends and the empirical trend are given for three plausible scenario's: one with the highest case-fatality improvement, one with the lowest case-fatality improvement and a Rochester-like scenario that turns out to be in between these two case-fatality declines. Assuming no incidence decline and the highest known annual case-fatality improvement results in a mortality decline that doesn't fit to the national figures.

Figure 5 shows the projected age-specific absolute changes in prevalence rate during the 20-year period. The decrease of age specific stroke mortality is largest for persons in their late 60s and 70s.⁶ Case-fatality decline within these age groups has also been limited also,⁷ so for these persons a rather large incidence decline has to be assumed. Consequently, stroke morbidity among these groups is decreasing remarkably, for men at a younger age than for women. Later in life, the increase in survival results in a large increase in stroke prevalence, offsetting a relatively small incidence decline. In Figure 5, an upper and lower limit of the age specific prevalence changes is given. These limits are determined by the extreme values of the plausible ranges for incidence decline and case-fatality decline as reported in Figure 3. A smaller incidence decline results in a smaller prevalence decline among the younger patients and a higher morbidity among the older groups. The resulting prevalence changes in these alternatives scenario's, again assuming a constant mortality decline, however, are only slightly different.

Discussion

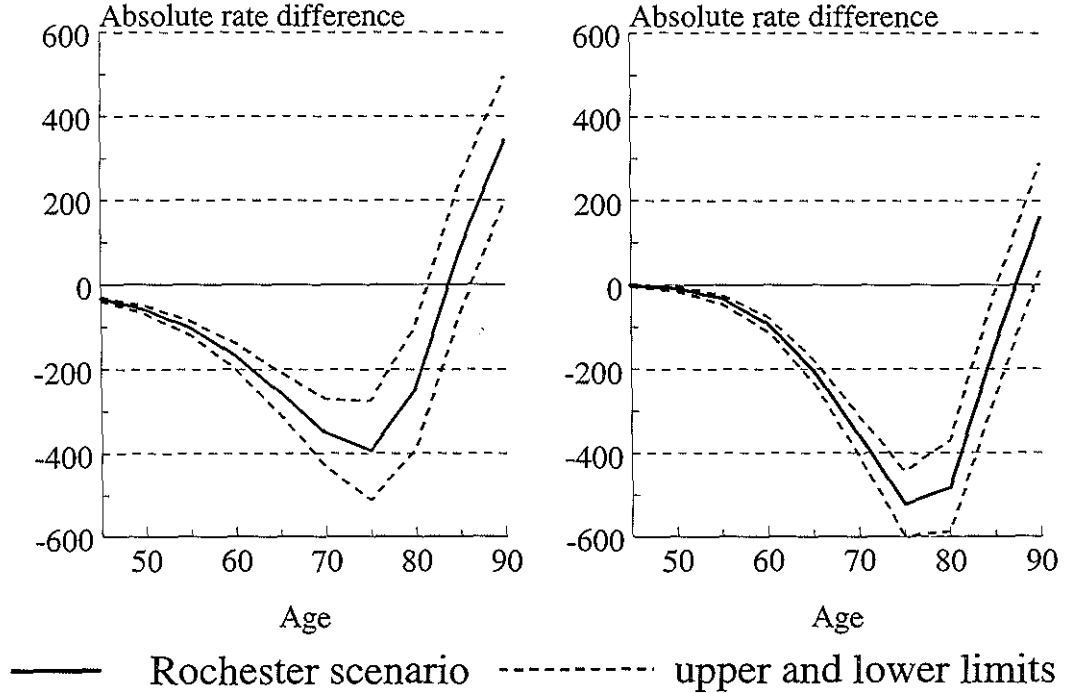
Downward trends in the occurrence of ischemic heart disease and stroke characterize changes in health within the aging population: reduced disease-specific mortality results in a relatively limited increase in life expectancy but might cause a longer period of severe disability from the same disease.¹⁵ In the case of stroke, the major question is whether declining mortality rates are resulting in a paradoxically increasing burden of disease, especially among the oldest. The answer depends on whether one supposes mortality and morbidity to be compressed against an alleged fixed biological upper limit to the life span, or whether one supposes a mortality decrease in the oldest age groups and a parallel expansion of morbidity. In the former scenario health care provision results in a decrease of morbidity but in the latter it may well result in an increase of chronic morbidity.

Men

Women

Figure 5

Graph showing the projected age-specific absolute change in prevalence rate in the Netherlands over the period 1985-2010. The most plausible trends in incidence and case-fatality are assumed, resulting in continuation of a Rochester-like scenario. Upper and lower limits have been calculated using the most extreme values of the range of plausible incidence and case-fatality trends found through sensitivity analysis.



Our analysis is based on empirical data from different sources. None of the large population-based studies, or clinical trials has been comprehensive enough to be able to assess the extent of incidence decline and survival improvement in relation to stroke mortality decline. The results show a plausible range of a considerable incidence decline for the Netherlands. As mortality decline in the USA has been much higher,⁵ most likely incidence decline has also been higher. For the Netherlands, consequently, Figure 5 confirms a most likely scenario for stroke with a compression of morbidity in the near future, but with an increase of major stroke prevalence among the very old. At a younger age dominant incidence decline results in a decrease in morbidity. For the oldest age groups, however, the decrease in case fatality is larger than the calculated incidence decline and therefore the resulting mortality decrease is small i.e. approximately 1% annually.⁶ At these ages the result is, indeed, a trade-off of stroke mortality for morbidity. The projected changes in morbidity are supported by recent observations: age-adjusted admission rates for stroke are decreasing in the Netherlands²⁵ and definitely also in the USA among whites.³⁵ In both countries the average age of stroke patients is increasing.^{29,35} In the Netherlands, the average age of patients admitted to long-term care institutions, which is indicative for the prevalence of major stroke, is increasing as is the average disability score.²⁹ This confirms the changes towards a higher major stroke prevalence among higher age groups as reported in Figure 5. Similarly the average stay of severely disabled in nursing institutions is increasing and consequently Dutch mortality statistics are showing a parallel increase of late cerebrovascular deaths (ICD code 438) among the oldest.⁶ The increase in institutionalization is not explained by social factors, as the intensive home care programme has been expanding the last five years to cope with waiting lists of chronic patients with major stroke.⁶

In addition to incidence and survival, stroke morbidity rates are determined by the risk of residual disability after stroke and the chance of recovery. In this respect, some groups are running larger risks after a stroke due to concomitant debilitating diseases such as atherosclerotic heart disease or other risk factors, such as hypertension. It is hoped that ongoing empirical studies may be able to answer questions regarding stroke trends in these patient groups now that their survival is improving.¹⁵ Also, empirical studies will have to answer related questions on co-morbidity and disability from other diseases among the aged.

An important question is: will stroke mortality continue to decline? Because most of the mortality decline is unexplained, no one can be sure of the answer, nor does the stroke model answer this question. In its projections it assumes the same continuing mortality decline, as this decline has been very constant in the Netherlands. Our model demonstrates the dynamics of stroke morbidity change. Both in the Netherlands and the USA there are still benefits to be gained from large-scale hypertension control and reduction of smoking.^{1,3,35,37,38} Better intervention possibilities might further improve prognosis. Population benefits from recurrence prevention are limited, because of the relatively high first stroke fatality and relatively low recurrence risk. The

effects of increasing cardiac disease prevention and treatment are already evident and will further increase survival.

The influence of demographic changes differs between the Netherlands and the USA. Stroke prevalence will increase less in the USA, because of the less extreme aging of the population. Stroke incidence and morbidity rates are higher for blacks. However, for this group mortality decline parallels the decline for whites.⁴ With a mortality trend of the same magnitude, similar dynamics in stroke morbidity might be taking place. These issues can only be dealt with after including population-specific transition probabilities and demographic and epidemiological trends in the state-event model, which is possible.

In conclusion, this study supports evidence of a further decline in stroke incidence.³ It also supports the observation³⁹ that a further improvement of survival of the older age groups as a result of therapeutic interventions may result in a longer period of severe disability before death. The findings are of importance for setting health care priorities for the aged, especially in regard to the nursing needs of stroke patients during the acute, rehabilitative, and chronic phase of their illness.

Appendix

Testing the goodness-of-fit

Age-specific computed stroke figures are compared with the national registry figures. This is done by using the standard formula for the chi-square test for larger tables: $\chi^2 = \Sigma((O-E)^2/E)$, *d.f.* = (C-1).³⁴ Here O represents observed figure in both groups of data and E the expected figure. E is based on the calculation $(R \cdot N)/T$ where R is the sum of the computed and the registry figure for the age group involved, N is the total of all age groups, and T the total for all age groups of the computed and registry figures together. C is the number of age groups. The number of degrees of freedom is the product of the number of age groups minus 1 and the number of categories (i.e. computed and observed) minus 1.

References

1. Casper M, Wing S, Strogatz D, Davis CE and Tyroler HA. Antihypertensive treatment and US trends in stroke mortality, 1962 to 1980. *Am J Public Health* 1992;12:1600-08.
2. Harmsen P, Tsiopogianni A and Wilhelmsen L. Stroke incidence rates were unchanged, while fatality rates declined, during 1971-1987 in Goeteborg, Sweden. *Stroke* 1992;23:1410-15.
3. McGovern PG, Burke GL, Sprafka JM, Xue S, Folsom AR and Blackburn H. Trends in mortality, morbidity and risk-factor levels for stroke from 1960 through 1990 - the Minnesota Heart Survey. *JAMA* 1992;268:753-59.

4. Modan B and Wagener DK. Some epidemiological aspects of stroke: mortality/morbidity trends, age, sex, race, socioeconomic status. *Stroke* 1992;23:1230-36.
5. Bonita R, Stewart A. and Beaglehole R. International Trends in Stroke Mortality: 1970-1985. *Stroke* 1990;21:989-92.
6. Central Bureau of Statistics. *Deaths by causes of deaths, age and gender*. Voorburg, The Netherlands: 1985. Annual Report.
7. Broderick JP, Philips SJ, Whisnant JP, O'Fallon WM and Bergstralh EJ. Incidence rates of stroke in the eighties: the end of the decline in stroke? *Stroke* 1989;20:577-82.
8. Kotila M. Decline in the incidence of stroke. *Stroke* 1988;19:1572-73.
9. Terent A. Survival after stroke and transient ischemic attacks during the 1970s and 1980s. *Stroke* 1989;20:1320-26.
10. Tuomilehto J, Bonita R, Stewart A, Nissinen A and Salonen JT. Hypertension, cigarette smoking and the decline in stroke incidence in Eastern Finland. *Stroke* 1991;22:7-11.
11. Sacco RL, Wolf PA, Kannel WB, McNamara PM. Survival and recurrence following stroke; the Framingham Study. *Stroke* 1982;13:290-295.
12. Howard G, Toole JF, Becker C, Lefkowitz DS, Truscott L et al. Changes in survival following stroke in five North Carolina counties observed during two different periods. *Stroke* 1989;20:345-50.
13. Klag MJ, Whelton PK and AJ Seidler. Decline in US Stroke Mortality; demographic trends and antihypertensive treatment. *Stroke* 1989;20:14-21.
14. Weinstein MC. Methodologic Issues in policy modeling for cardiovascular disease. *J Am Coll Cardiol* 1989;14:38-43A.
15. Tsevat J, Weinstein MC, Williams LW, Tosteson ANA and Goldman L. Expected gains in life expectancy from various coronary heart disease risk factor modifications. *Circulation* 1991;83:1194-1201
16. Central Bureau of Statistics. *Population projections* [in Dutch]. Monthly Population Statistics 1986;3:20-28.
17. Bamford J, Sandercock P, Dennis M, Warlow C, Jones L, McPherson K, Vessey M, Fowler G, Molyneux A, Hughes T, Burn J and Wade D. A prospective study of acute cerebrovascular disease in the community: the Oxfordshire Community Stroke Project 1981-1986; 1. Methodology, demography and incident cases of first-ever stroke. *J Neurol, Neurosurg and Psych*, 1988;51:1373-80.
18. Dennis M, Bamford JM, Sandercock PAG and Warlow CP. Incidence of transient ischemic attacks in Oxfordshire, England. *Stroke* 1989;20:333-39.
19. Van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJA and Van Gijn J. Inter-observer agreement for the assessment of handicap in stroke patients. *Stroke* 1988;19:604-07.
20. Dutch TIA Trial Study Group. A comparison of two doses of aspirin (30 mg vs. 283 mg a day) in patients after a transient ischemic attack or minor ischemic stroke. *N Engl J Med* 1991;325:1261-66.
21. Herman B, Van Luijk JH, and Frenken CWGM, Op de Coul AAW and Schulte BPM. Epidemiology of Stroke in Tilburg, The Netherlands; the population-based stroke incidence register: 2. Incidence, initial clinical picture and medical care and three-week case-fatality. *Stroke* 1982;13:629-634

22. Giroud M, Milan C, Beuriat P, Gras P, Essayagh E et al. Incidence and survival rates during a two-year period of intracerebral and subarachnoid hemorrhages, cortical infarcts, lacunas and transient ischemic attacks. The Stroke Registry of Dijon: 1985-1989. *Int J Epidemiol* 1991;20:892-99.
23. Vanden Hoogen HJM, Huygen FJA, Schellekens JWG, Straat JM and Van der Velden HGM. Morbidity figures from general practice; data from four general practices (1978-1982). Nijmegen University Department of General Practice, Nijmegen, The Netherlands; 1985
24. Whisnant JP. Comparison of case ascertainment by medical record linkage and cohort follow-up to determine incidence rates for transient ischemic attacks and stroke. *J Clin Epidemiology* 1990;43:791-97.
25. Vander Meer K and Smith RJA. Stroke patient in general practice. A study in 1% of the Dutch population [in Dutch]. *Huisarts Wet* 1990;33:141-44.
26. Bonita R and Beaglehole R. Recovery of motor function after stroke. *Stroke* 1988;19:1497-1500.
27. Howard G, Walker MD, Becker C, Couli B, Feibel J, McLeroy K, Toole JF, Yatsu F. Community hospital-based stroke programs: North Carolina, Oregon and New York III. Factors influencing survival after stroke: proportional hazard analysis of 4219 patients. *Stroke* 1986;17:294-99.
28. Howard G, Evans GW, Murros KE, Toole JF, Lefkowitz D and Truscott BL: Cause specific mortality following cerebral infarction, *J Clin Epidemiol* 1991;42:45-51.
29. Information Center for Health Care. Cerebrovascular accidents and cerebral infarctions. In: *Diseases in the picture* (in Dutch). Utrecht, The Netherlands: Information Center for Health Care; 1990, part 3.
30. Dennis M, Bamford J, Sandercock P and Warlow C. Prognosis of transient ischemic attacks in the Oxfordshire Community Stroke Project. *Stroke* 1990;21:848-53.
31. Garraway WM, Elveback LR, Connolly DC and Whisnant JP. The dichotomy of myocardial and cerebral infarction. *Lancet* 1983;ii:1332-35.
32. Garraway W. M. The changing pattern of survival following stroke. *Stroke* 1983;14:699-702.
33. Malmgren R. Geographical and secular trends in stroke incidence. *Lancet* 1987;iii:1196-1200.
34. Alo K. Prevalence of stroke in Finland. *Stroke* 1986;17:681-86.
35. Kirkwood BR. *Essentials of medical statistics*. Blackwell Scientific Publications, Boston; 1988.
36. Centers for Disease Control. Cerebrovascular disease mortality and Medicare Hospitalization - United States (1980-1990). *MMWR* 1992;41:476-77.
37. Gunning-Schepers LJ. The health benefits of prevention, a simulation approach. *Health Policy* 1989;12/1-2:93-129A.
38. Dyken ML, Wolf PA, Barnett HJM, Bergan JJ, Hass WK, Kannel WB, Kuller L, Kurtzke JF and Sundt TM. Risk factors in stroke. *Stroke* 1984;15:1105-10.
39. Malmgren R, Bamford J, Warlow C, Sandercock P and Slattery J. Projecting the number of patients with first ever stroke and patients newly handicapped by stroke in England and Wales. *BMJ* 1989;298:656-660.

Regression analysis of recent changes in cardiovascular morbidity and mortality in the Netherlands

Abstract

Objectives: To test whether recent declines in mortality from coronary heart disease were associated with increased mortality from other cardiovascular diseases.

Design: Poisson regression analysis of national data on causes of death and hospital discharges.

Setting and subjects: The population of the Netherlands, 1969-93.

Main outcome measures: Annual changes in mortality from coronary heart disease, stroke and other cardiovascular diseases and annual changes in hospital discharge rates for acute coronary events, stroke and congestive heart failures.

Main results: Patterns of cardiovascular mortality changed abruptly in 1987-93. Annual decline in mortality from coronary heart disease increased sharply for women and men, from -1.9% (95% confidence interval: -2.2% to -1.6%) and -1.7% (-1.9% to -1.4%) respectively in 1979-86 to -3.1% (-3.5% to -2.6%) and -4.2% (-4.6% to -3.9%) in 1987-93. The long standing decline in mortality from stroke levelled off: from annual change of -3.3% (-3.7% to -2.8%) and -3.2% (-3.7% to -2.8%) in 1979-86 to -0.1% (-0.7% to +0.4%) and -1.1% (-1.7% to -0.5%) in 1987-93. Mortality from other causes of cardiovascular death started to increase, from -2.0% (-2.4% to -1.6%) and -0.2% (-0.5% to +0.2%) in 1979-86 to +1.5% (+1.0% to +2.0%) and +1.9% (+1.5% to +2.3%) in 1987-93. Hospital discharge rates for acute coronary heart disease, congestive heart failure and stroke increased

during 1980-86. During 1987-93 discharge rates for stroke and coronary heart disease stabilised, but rates for congestive heart failure increased.

Conclusion: Improved management of coronary heart disease seems to have reduced mortality, but some of the gains are lost to deaths from stroke and other chronic vascular diseases. The increasing numbers of patients with coronary heart disease who survive will increase demands on health services for long term care.

Introduction

In the early 1970's mortality from cardiovascular diseases started to decline in many industrialised countries.¹ Despite considerable debate, most observers would agree that reductions in risk factors, particularly smoking and hypertension, was the more effective than improvements in treatments in achieving this decline in the 1970s and 1980s.²⁻⁷ In the mid 1980s, however, management of acute myocardial infarction was revolutionised, particularly by thrombolytic therapy, causing steep decreases in mortality from coronary heart disease.⁸⁻¹⁰

Coronary heart disease is not the only cardiovascular disease, however, and other cardiovascular diseases, such as stroke and congestive heart failure, share many of the same risks.^{5 11 12} The improving prognosis of coronary heart disease caused by improved management should increase the number of surviving patients at high vascular risk. We present a time series analysis of Dutch nationwide statistics to illustrate the relation between mortality from coronary heart disease and other cardiovascular diseases.

Methods

Source of data

For our mortality analysis, we used the registered numbers of death by cause, age (from 25 till 84 y), sex and calendar year from Statistics Netherlands.¹³ We took account of only primary causes of death and considered three causes of cardiovascular related death: coronary heart disease, stroke and all other causes, including unknown causes of sudden death.¹⁴ Table 1 shows the ICD codes (international classification of diseases) that we searched for.

The second database we used was the hospital register. This provides nationwide coverage and is complete since 1980. The register includes hospital patient's diagnosis at discharge as classified by the treating physician and codified by local staff. For every discharge, the vital status of the patient is registered. Patients dying during the ambulance ride or at entry in the emergency room are not admitted, and are considered 'dead out of hospital'. Again, we considered only primary diagnoses.

Statistical analysis

We estimated trends over time by Poisson regression analysis.¹⁵ We used midyear populations as person-years and we specified five year age groups (categorical) and calendar year (continuous) as independent variables and mortality as dependent variable. We analyzed the trends from vital statistics over 1969-1978, 1979-1986 and 1986-1993 and trends from the hospital register from 1980-1986 and 1986-1993. We choose 1987 as the cut-off point because of the apparent rupture in trend in that year.

Results

Figure 1 and table 2 show changes in mortality from cardiovascular diseases and in discharge rates from hospital for stroke, acute coronary heart events, and congestive heart failure.

For women, all cardiovascular related death rates declined from 1970 till 1986, suggesting a change in common risk factors of most cardiovascular diseases. During 1987-1993, however, changes in cardiovascular mortality levelled off, from -2.3 % per year in 1969-86 to -0.7 % per year in 1987-93. Mortality from coronary heart disease decreased steeply, but other vascular related death rates started to increase in 1987. Mortality from stroke levelled off after a long period of decline.

For men, changes in cardiovascular related mortality during 1969-1978 were limited to a decline in mortality from stroke. During 1979-1986, the rate of decline in cardiovascular related mortality increased, driven by declining death rates from coronary heart and stroke; rates for other cardiovascular causes of death changed little. In 1987-1993 mortality from coronary heart disease declined steeply, but death rates from stroke declined less steeply and death rates from other cardiovascular causes seemed to increase.

Table 1 - ICD codes used in search for changes in cardiovascular morbidity and mortality in the Netherlands during 1969-93. ICD-8 and ICD-9 (international classification of diseases, eighth and ninth revisions) used for analysis of death registered in Statistics Netherlands,¹³ and ICD-9-CM (international classification of disease, ninth revision, clinical modification) used for analysis of diagnosis at discharge listed in hospital register.

Disease	ICD 8 (1969-78)	ICD 9 (1979-93)	ICD 9-CM (1980-93)
cardiovascular disease	390-458, 782, 795-796	390-459, 798-799	
coronary heart disease	410-414	410-414	
stroke	430-438	430-438	430-438
acute coronary events			410-411
congestive heart failure			428-429

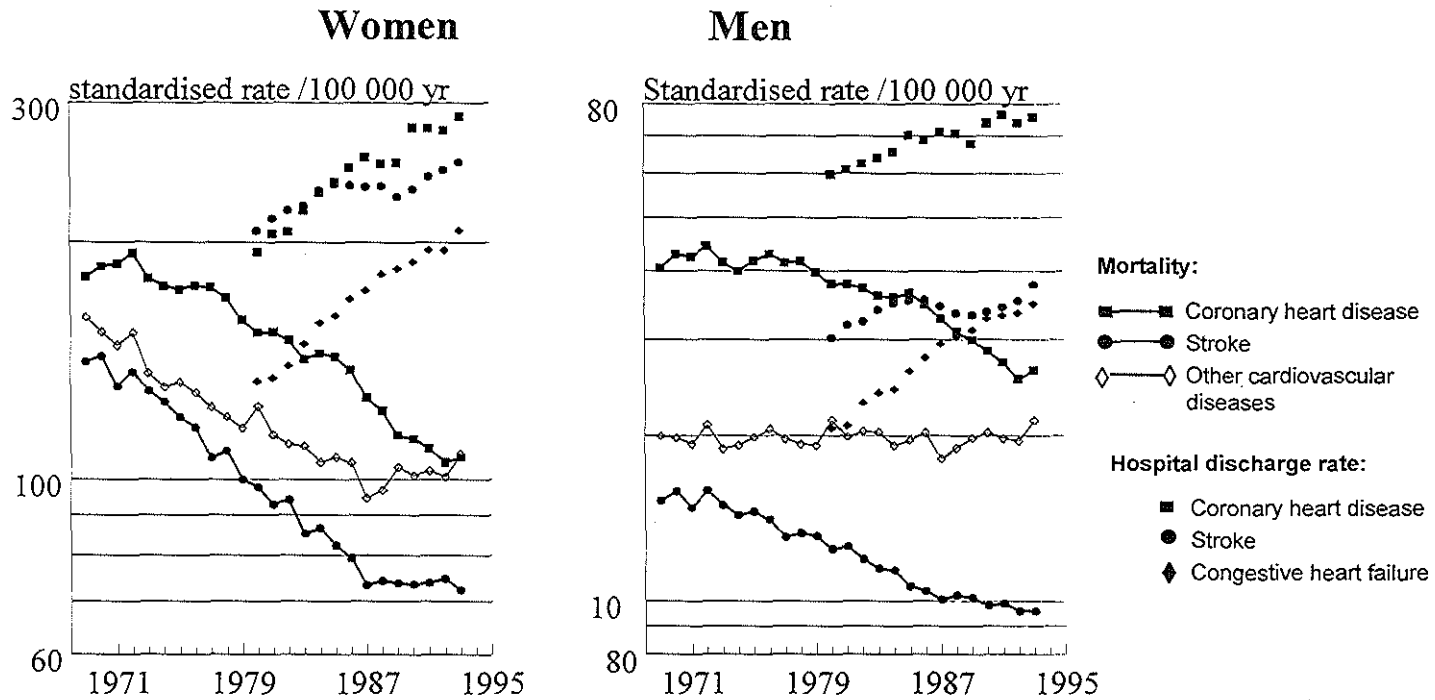


Figure 1

Age standardised (Dutch population of 1980-93 as direct standard) annual mortalities and hospital discharge rates for cardiovascular diseases in the Netherlands.

Table 2

Age standardised rates (per 100 000 population) and annual percentage changes in mortality and hospital discharge rates for the population of the Netherlands during 1969-93. (Data on national mortality from Statistics Netherlands¹³ and data on hospital mortality and morbidity from hospital register)

	Period 1969-78		Period 1979-86 †		Period 1987-93	
	rates (SE)	percentage change (95% CI)	rates (SE)	percentage change (95% CI)	rates (SE)	percentage change (95% CI)
Women						
<i>National mortality from:</i>						
Cardiovascular disease	443 (1.1)	-2.3 (-2.5 to -2.2)	349 (0.8)	-2.3 (-2.5 to -2.1)	289 (0.7)	-0.7 (-1.0 to -0.4)
Stroke	125 (0.3)	-3.3 (-3.6 to -3.0)	90 (0.2)	-3.3 (-3.7 to -2.8)	74 (0.2)	-0.1 (-0.7 to +0.4)
Coronary heart disease	180 (0.4)	-1.0 (-1.3 to -0.8)	148 (0.4)	-1.9 (-2.2 to -1.6)	114 (0.3)	-3.1 (-3.5 to -2.6)
Other	138 (0.3)	-3.2 (-3.5 to -2.9)	111 (0.3)	-2.0 (-2.4 to -1.6)	101 (0.3)	+1.5 (+1.0 to +2.0)
<i>Mortality in hospital from:</i>						
Cardiovascular disease			124 (0.3)	-0.7 (-1.1 to -0.2)	109 (0.3)	-1.6 (-2.0 to -1.1)
Stroke			42 (0.1)	-1.5 (-2.3 to -0.7)	36 (0.1)	+0.4 (-0.4 to +1.3)
Coronary heart disease			38 (0.1)	+0.2 (-0.6 to +1.0)	31 (0.1)	-5.5 (-6.3 to -4.6)
Other			44 (0.1)	-0.7 (-1.5 to +0.1)	42 (0.1)	-0.4 (-1.2 to +0.3)
<i>Hospital discharge rate</i>						
Stroke			213 (0.8)	+0.7 (+0.3 to +1.0)	204 (0.7)	-0.0 (-0.4 to +0.3)
Acute coronary event			211 (0.8)	+2.7 (+2.3 to +3.1)	235 (0.8)	+1.1 (+0.8 to +1.5)
Congestive heart failure			141 (0.6)	+2.5 (+2.0 to +2.9)	160 (0.6)	+1.3 (+0.9 to +1.7)

Table 2 - continued

	Period 1969-78		Period 1979-86 †		Period 1987-93	
	rates (SE)	percentage change (95% CI)	rates (SE)	percentage change (95% CI)	rates (SE)	percentage change (95% CI)
Men						
<i>National mortality from:</i>						
Cardiovascular disease	762 (2.4)	-0.4 (-0.6 to -0.3)	687 (2.4)	-1.5 (-1.7 to -1.3)	583 (2.1)	-1.7 (-1.9 to -1.4)
Stroke	146 (0.5)	-1.9 (-2.2 to -1.5)	117 (0.4)	-3.2 (-3.7 to -2.8)	99 (0.4)	-1.1 (-1.7 to -0.5)
Coronary heart disease	419 (1.3)	-0.1 (-0.3 to +0.1)	369 (1.3)	-1.7 (-1.9 to -1.4)	287 (1.0)	-4.2 (-4.6 to -3.9)
Other	198 (0.6)	0.0 (-0.3 to +0.3)	201 (0.7)	-0.2 (-0.5 to +0.2)	197 (0.7)	+1.9 (+1.5 to +2.3)
<i>Mortality in hospital from:</i>						
Cardiovascular disease			238 (0.9)	-0.7 (-1.1 to -0.3)	214 (0.8)	-1.9 (-2.3 to -1.5)
Stroke			59 (0.2)	-2.2 (-3.0 to -1.4)	53 (0.2)	-0.6 (-1.4 to +0.2)
Coronary heart disease			84 (0.3)	-1.3 (-2.0 to -0.7)	65 (0.2)	-5.8 (-6.5 to -5.1)
Other			95 (0.4)	+0.7 (+0.1 to +1.4)	96 (0.4)	0.1 (-0.5 to +0.7)
<i>Hospital discharge rate</i>						
Stroke			322 (1.1)	+1.3 (+0.9 to +1.6)	302 (1.0)	-0.0 (-0.4 to +0.3)
Acute coronary event			621 (1.5)	+1.4 (+1.1 to +1.6)	635 (1.4)	-0.4 (-0.6 to -0.2)
Congestive heart failure			229 (0.1)	+3.7 (+3.2 to +4.1)	281 (1.0)	+1.1 (+0.8 to +1.5)

† 1980-6 for the hospital register data

These upward changes are similar to, but less clear than, those seen among women.

The hospital register showed an increase in discharge rates for acute coronary heart disease, congestive heart failure and stroke in 1980-86. In 1987-93, the discharge rates stabilised for stroke and coronary heart disease, though rates for coronary heart disease still increased among women. The age adjusted case fatality ratio for an acute coronary event decreased from 13.6% (95% confidence interval 13.3% to 13.9%) for men and 18.4% (17.8% to 19.0%) for women in 1980-81 to 7.9% (7.7% to 8.1%) for men and 9.9% (9.5% to 10.2%) for women in 1992-93. The case fatality ratios for strokes remained constant at about 17.7 % for both sexes since 1987. For congestive heart failure, both sexes showed an age dependent increase in discharges by calendar date. In more recent years, more patients were hospitalised at higher ages.

Discussion

Our study provides circumstantial evidence that the sharp drop in mortality from coronary heart disease between 1985 and 1993, the levelling off of mortality from stroke, and the increase in mortality from congestive heart failure are causally linked by the same process: the increased survival of patients with coronary heart disease. Cardiovascular diseases share many of the same risk factors, and having one disease increases the risk for others: a history of ischemic heart disease increases the risk for other heart diseases, notably heart failure and dysrhythmias, and these increase the risks for cardiogenic stroke.¹⁶ As the prognosis for coronary heart disease improves the increasing numbers of surviving patients will increase the pools of people at high risk of other heart diseases and stroke. These will result in increasing death rates from stroke and other cardiovascular diseases.

Validity of study

The changes we observed might have been caused by changing diagnostic habits, policies for referral or rules of classification. The validity of the Dutch register of causes of death is reasonable.¹⁷ Any changes in coding between cardiovascular and non-cardiovascular causes of death are likely to have been small and unable to bias seriously trends over time. The possibility of misclassification between different causes of cardiovascular related death is high, but the observed patterns in both the hospital register and the mortality statistics were consistent (see table 2).

General Practice registers have shown that nearly all patients suspected of having an acute myocardial infarction are hospitalised in the Netherlands.¹⁸ Death rates both outside hospital and in hospital showed substantial reductions, which makes it unlikely that deaths outside hospital were exchanged for deaths in hospital. Moreover, the secular trend of

improving prognosis has been documented before, in the Netherlands as elsewhere, and has been linked to improved management.^{4 8-10 19}

Part of the increased rates of discharge of patients with stroke and the decrease in case fatality in the early 1980s was caused by the introduction of computed tomography, which ascertained more benign lesions.^{6 7 20} Since 1987, the incidence of and mortality from stroke have remained constant, suggesting a steady state in survival. The sharp decrease in mortality from coronary heart disease and the concomitant levelling off of mortality from stroke after a long period of decline has also been observed in the prospective, population based Minnesota heart survey.^{7 9 20} The age dependent increase in mortality from congestive heart failure rates has also been documented before.²¹⁻²⁴

Secular risk factor changes might explain the observed changes, but these have been modest (at best) in the period under study in the Netherlands.²⁶

Conclusion

Improvements in treating coronary heart disease seem undeniable, but some of the gains made are lost again to deaths from stroke and congestive heart failure. This has important consequences for public health, as increasing numbers of surviving but disabled patients with chronic cardiovascular disease are boosting demand for health care.

References

1. Uemura K, Pisa Z. Trends in cardiovascular disease mortality in industrialised countries. *Wld Health Statist Quart* 1988;41:155-68.
2. Goldman L, Cook E. The decline in Ischemic Heart Disease mortality rates. *Ann Intern Med* 1984;101:825-36.
3. Sigfusson N, Sigvaldason H, Steingrimsdottir L, Gudmundsdottir II, Stefansdottir I, Thorsteinsson T, et al. Decline in ischaemic heart disease in Iceland and change in risk factor levels. *Br Med J* 1991;302:1371-5.
4. Sytkowski P, Kannel W, D'Agostinho R. Changes in risk factors and the decline in mortality from cardiovascular disease. *N Engl J Med* 1990;322:1635-41.
5. Bonita R. Epidemiology of stroke. *Lancet* 1992;339:342-4.
6. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke trends in an aging population. *Stroke* 1993;24:931-939.
7. McGovern PG, Burke GL, Sprafka JM, Xue S, Folsom AR, Blackburn H. Trends in mortality, morbidity, and risk factor levels for stroke from 1960 through 1990. The Minnesota Heart Survey. *JAMA* 1992;268:753-9.
8. Stevenson R, Ranjadayalan K, Wilkinson P, Roberts R, Timmis AD. Short and long term prognosis of acute myocardial infarction since introduction of thrombolysis. *Br Med J* 1993;307:349-53.

9. McGovern PG, Pankow JS, Shahar E, Doliszny KM, Folsom AR, Blackburn H, et al. Recent trends in acute coronary heart disease--mortality, morbidity, medical care, and risk factors. The Minnesota Heart Survey Investigators. *N Engl J Med* 1996;334:884-90.
10. Meeter K, Honkoop P, Verhage A, Boersma H, Fioretti P, Deckers J. Treatment of myocardial infarction during the hospitalisation phase and a short period afterward: now and 10 years ago [in Dutch]. *Ned Tijdschr Geneeskd* 1993;137:1922-6.
11. Kannel WB, Belanger AJ. Epidemiology of heart failure. *Am Heart J* 1991;121:951-7.
12. Hachinski V. Preventable senility: a call for action against the vascular dementias. *Lancet* 1992;340:645-8.
13. Statistics Netherlands. *Deaths by causes of deaths, age and gender*. Voorburg (Netherlands): Central Bureau of Statistics, 1994.
14. Kannel W, Schatzkin A. Sudden death: Lessons from subsets in population studies. *J Am Coll Cardiol*. 1985;5:141B-149B.
15. Breslow NE, Day NE. Rates and rate standardization. *The design and analysis of cohort studies*. Lyon: International Agency for Research on Cancer, 1987:48-79.
16. Cerebral Embolism Task Force. Cardiogenic brain embolism. The second report of the Cerebral Embolism Task Force. *Arch Neurol* 1989;46:727-43.
17. Mackenbach JP, van Duyne WMJ, Kelson MC. Certification and coding of two underlying causes of death in the Netherlands and other countries of the European Community. *J Epid Comm Health* 1987;41:156-60.
18. Fracheboud J. *Coronary Center Care of Home Care? A descriptive study of home care of patients with an acute myocardial infarction in the Netherlands*. (in Dutch). Utrecht: NIVEL, 1987.
19. Gheorghiade M, Ruzumna P, Borzak S, Havstad S, Ali A, Goldstein S. Decline in the rate of hospital mortality from acute myocardial infarction: impact of changing management strategies. *Am Heart J* 1996;131:250-6.
20. Broderick JP, Phillips SJ, Whisnant JP, O'Fallon WM, Bergstralh EJ. Incidence rates of stroke in the eighties: the end of the decline in stroke? *Stroke*. 1989;20:577-82.
21. Reitsma JB, Mosterd A, Koster RW, van Capelle FJ, Grobbee DE, Tijssen JG. Increase in the number of admissions due to heart failure in Dutch hospitals in the period 1980-1992. [in Dutch]. *Ned Tijdschr Geneeskd* 1994;138:866-71.
22. Bonneux L, Barendregt J, Meeter K, Bonsel G, van der Maas P. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. The future rise of heart failure. *Am J Publ Health* 1994;84:20-29.
23. Ghali JK, Cooper R, Ford E. Trends in hospitalization rates for heart failure in the United States, 1973-1986. Evidence for increasing population prevalence. *Arch Intern Med* 1990;150:769-73.
24. Anonymous. Mortality from congestive heart failure - United States, 1980-1990. *Morbidity and Mortality Weekly Report* 1994;43:77-81.
25. Ruwaard D, Kramers PGN, Van den Bergh Jeths A, Achterberg PW, editors. *Public health status of Dutch Population over the period 1950-2010*. The Hague: Sdu, 1994.

Diverging trends in colorectal cancer morbidity and mortality in the Netherlands

Earlier diagnosis comes at a price

Abstract

In developed countries, time trends in incidence of colorectal cancer differ markedly from trends in mortality. This study sought to explain simultaneously changes in both colorectal cancer incidence and mortality. Data on first admissions, interventions and outcome from the national hospital registry over the period 1978-1989 and data on mortality from Statistics Netherlands over the same period were analysed by age-period models and subsequently entered in a Markov-chain model, simulating disease history from first admission to death. Over the period 1978-1989 age adjusted numbers of first admissions and interventions increased with respectively 37 % and 32 %, while mortality declined with 8 %. For every 100 patients admitted in 1987-89, 13 more will survive compared to 1978-80. Of these, 3 will be saved by improving results of primary treatment but the other 10 will survive their diagnosis the subsequent 10 years. Although progress in treatment has been made, therapeutic improvement can account only for the smaller part of the divergence between morbidity and mortality. Increased diagnostic activity, raising incidence and lowering mortality simultaneously, is the most likely cause of the unexplained divergence.

Introduction

Cancer registers in many developed countries detect increasing incidences of cancers unrelated to smoking, and commonly attribute these increases to changing environmental hazards.^{1,2} However, in the same countries, trends in cancer mortality differ strikingly from trends in incidence; as a rule mortality has been going down while incidence has been going up.^{3,4} This divergence between incidence and mortality can be explained by improved therapy and/or decreasing lethality.^{1,2} Advances have been made in specific therapies, such as the dramatic improvements of treatment in juvenile cancers and the more modest gains realized by adjuvant chemotherapy in advanced breast and colorectal cancer.^{5,6} Unspectacular but effective specific changes, however, have contributed probably even more to lowered cancer lethality, because they are applied to the majority of solid tumours: better preparation of the surgical patients, safer procedures and anaesthesia, improved control of infections, more effective reanimation, etc.

An alternative hypothesis explaining this divergence is increased case detection. By lowering diagnostic thresholds lesions with less invasive potential are added, increasing incidence and improving prognosis at the same time.⁷ If this is the case, it implies that part of the observed morbidity increase is iatrogenic and maybe preventable. To shed light on the likeliness of either explanation we have looked at incidence, mortality and survival from colorectal cancer in the Netherlands.

Colorectal cancer is the second most frequent cancer among men and women in the Netherlands, showing an incidence increase relative to mortality (figure 1). In the Netherlands, mass screening for colorectal cancer is not recommended. The Dutch policy makers feel that the unavoidable increase in morbidity and costs, induced by the many false positives (faecal occult blood testing) and/or by more demanding diagnostic procedures (sigmoido- or coloscopy) are not justified by the still uncertain decrease in mortality.^{8,9} But individual physicians may feel otherwise, and are free to act accordingly: long before mass screening for breast cancer was introduced in the Netherlands, the regional cancer registry showed increasing numbers of small tumours, witnessing earlier diagnosis.¹⁰ Contrary to breast cancer we do not possess accurate data about colorectal cancer stages at primary diagnosis in the Netherlands. Such information would be much harder to interpret anyway, because staging in colorectal cancer is more dependent of modern diagnostic imaging. This may give rise to stage migration: previously missed invasion of deeper tissues may be diagnosed by more modern diagnostic imaging, such as magnetic resonance imaging (MRI), causing an artefactual migration from milder to more severe disease stages.¹¹

This paper presents the trends in first admissions for colorectal cancer as primary diagnosis, major interventions during these first admissions to the hospital and mortality over the period 1978-1989 in the Netherlands.

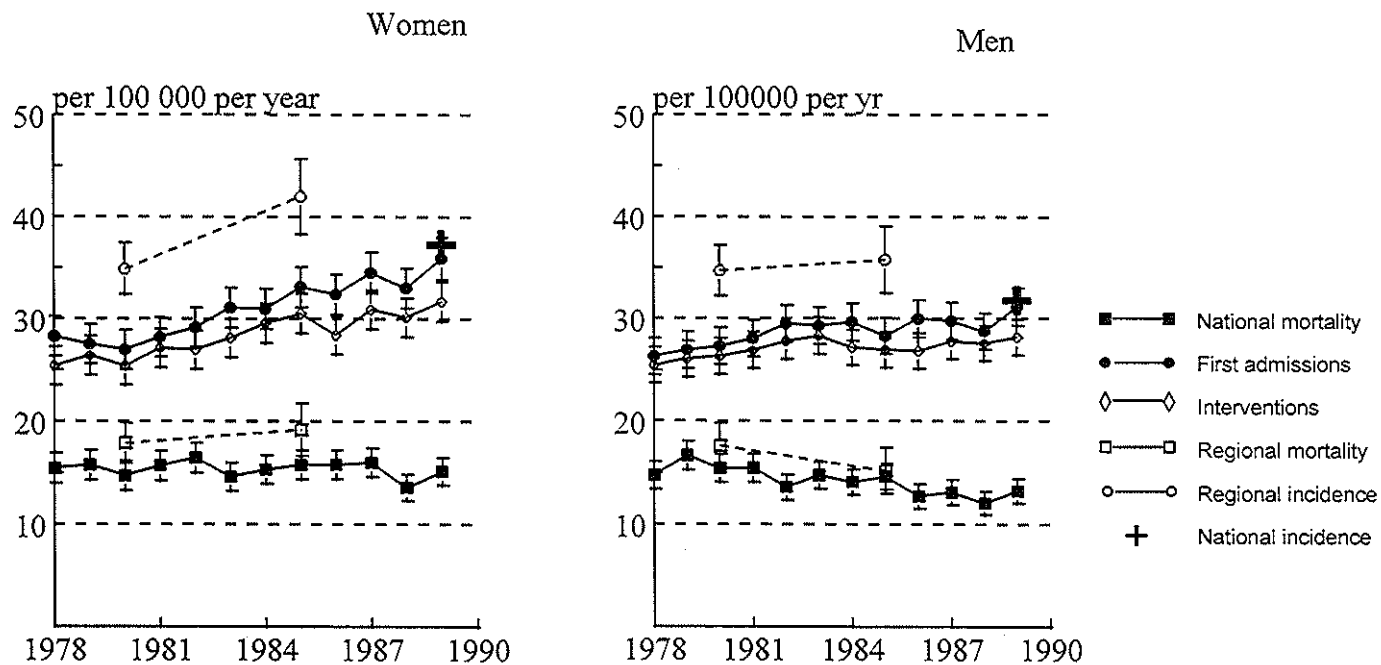


Figure 1

Full lines and symbols show the age standardized rates (30-64 year) of first admissions, major interventions and deaths for the Netherlands. Dotted lines and open symbols show the cancer incidence and mortality from the South Eastern Netherlands (1978-1987) and the national cancer incidence (1989).¹⁴ The error bars show 95% confidence limits (see text).

Table 1

Age standardized incidence and mortality per 100 000 per year and their standard errors (in brackets), by sex, period and source. The cancer register figures from the period 1978-1987 are from the Southeastern region only; those from 1989 are the first available national data (see text).

Men (30-64 y)

	<i>Cancer Register</i>			<i>Hospital register</i>		
	incidence	mortality	ratio	1st admission	mortality	ratio
1978-82	34.9 (1.3)*	17.9 (1.0)*	1.95 (0.14)*	28.2 (0.3)	14.9 (0.2)	1.90 (0.03)
1983-87	42.0 (1.9)*	19.2 (1.3)*	2.19 (0.18)*	33.9 (0.4)	15.3 (0.3)	2.21 (0.05)
1989	37.3 (1.0)	15.1 (0.6)	2.47 (0.12)	35.9 (1.0)	15.1 (0.6)	2.38 (0.12)

Women (30-64 y)

	<i>Cancer Register</i>			<i>Hospital register</i>		
	incidence	mortality	ratio	1st admission	mortality	ratio
1978-82	34.7 (1.3)*	17.6 (1.1)*	1.98 (0.15)*	27.8 (0.3)	14.4 (0.2)	1.93 (0.03)
1983-87	35.8 (1.7)*	15.2 (1.1)*	2.35 (0.21)*	30.2 (0.4)	14.2 (0.3)	2.14 (0.05)
1989	31.8 (0.9)	13.2 (0.6)	2.41 (0.12)	31.2 (0.8)	13.2 (0.9)	2.36 (0.17)

* From the Southeastern region; all other data are national.

The change in prognosis needed to explain the divergence between incidence and mortality is quantified, and we conclude that it is unlikely that improved therapy can explain the improving prognosis.

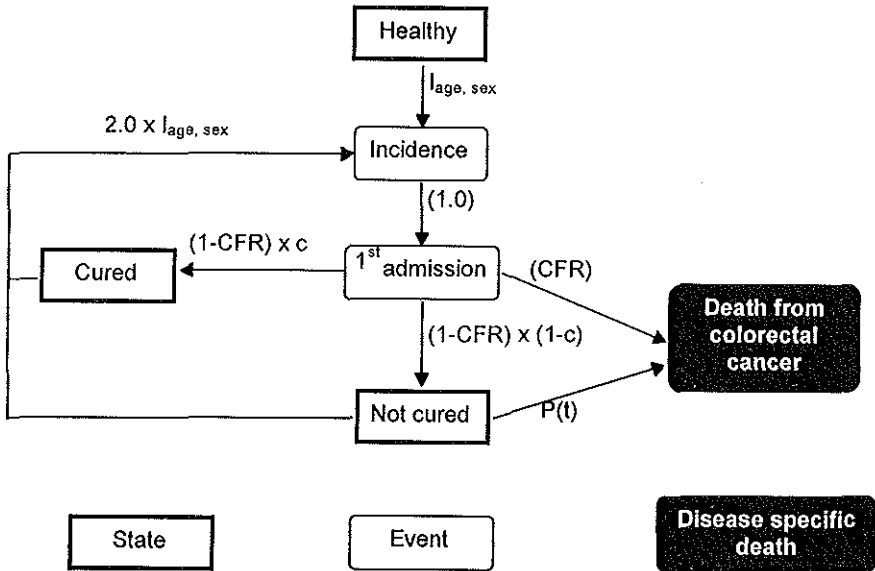
Patients and methods

Numbers of deaths for colon and rectum cancer (International Classification of Diseases, 9th revision, nrs. 153 and 154) by calendar year (1978-1989), 5 year age group (30-64) and sex have been obtained from Statistics Netherlands.¹² Person years at risk are approximated by the midyear population of an age and sex group in a calendar year. Summary estimates over age are calculated by using direct standardisation, using the European Standard Population as weights; standard errors are calculated by

$$\left(\sum_a \frac{w_a^2 O_a}{N_a^2} \right)^{1/2}$$

where O are the numbers, N the midyear population, w the weights and a the 5 year age groups between 30 and 64.¹³ Colon and rectum cancer are taken together in one group to avoid possible changes in codification.³ Nationwide data on colorectal cancer incidence are available since 1989.¹⁴ The regional cancer register in the Southeastern Netherlands (SE-N) publishes cancer incidences since 1975, but the population covered is relatively small (1 million) and in the period of interest (1978-1989) the colorectal cancer mortality in SE-N was significantly higher than in the rest of the Netherlands (see figure 1 and table 1).^{15,16}

Therefore, we used primarily national hospital register data as a proxy for incidence. In addition to administrative data the hospital register records the diagnosis at discharge (primary and secondary), the result at discharge (alive or dead), all major interventions, and whether it is a first admission for the considered cancer or not. We considered first admissions with colorectal cancer as primary diagnosis (International classification of diseases, 9th revision, nrs. 153-154) and major interventions during first admission as partial or total colectomy, rectum amputation or enterostomy.¹⁷ Comparing the (national) incidence for 1989 with the first admissions of the hospital register, we decided to limit the analysis till age 64. The hospital register became increasingly incomplete in the elder age groups, but for the younger age groups the incidence corresponded closely with the first admissions (see figure 1 and table 1). It has been shown before that more than 97% of all patients under 65 with colorectal cancer are treated in the Netherlands.¹⁸ Therefore, the nationwide hospital register of the Netherlands is an acceptable proxy of colorectal cancer incidence, if limited to the young and middle aged patients (30-64 year).



(Death from all other causes occurs in all states at any time)

Figure 2 - The Markov model of colorectal cancer.

Incidence is determined by age and sex. Persons may die in hospital (CFR), or may be cured (c). If not, they face a time dependent probability of cancer death $P(t)$ (see appendix). All patients run a risk of dying from other causes and a higher risk of a second primary tumour.

Period trends in intervention, first admission and mortality rates have been estimated by loglinear regression analysis.¹⁹ The observed rates are related to age-group, sex and calendar year as follows:

$$E_{a,s} = N_{a,s} e^{(\alpha_{a,s} + \beta_{a,s} \cdot x)}$$

where E is the expected number (of deaths, first admissions or interventions) and N the midyear population, a denotes 5 year age groups from 30 to 64 year, s sex and x is the calendar year. β is the slope of the regression line of the logarithm of the rates of every age and sex group versus calendar year, and represents the trend of age and sex specific rates over time. Summary estimates over age are calculated by specifying only sex as explanatory β variable.

The disease history is modelled by a Markov type state transition model (see figure 2). The model assumes three groups of patients after definite diagnosis: a fraction which dies during first admission (further called CFR, 'case fatality rate'), a fraction which leaves hospital 'cured' and a fraction which will die of the disease at some later point in time, provided they do

not die from other causes (further called 'not cured'). The probability of dying from colorectal cancer for the not cured is lognormally distributed over time characterised by a geometric mean (identical to the median survival time) and variance (see appendix).²⁰ Consequently the model accommodates changes in 4 components of prognosis: the probability of surviving primary treatment, the probability of being cured and - if not cured - the median survival time variance before dying from colorectal cancer. The case fatality rate (CFR), defined as the fraction that dies in hospital within 2 months of primary diagnosis, is known from the hospital register. The fraction which is cured approximates the fraction surviving 10 year after hospital discharge (corrected for death from other causes): the risk of dying more than 10 year after diagnosis of colorectal cancer equals the risk of death of the reference population.²¹ All patients, the cured and the not cured, run twice the risk of the reference population for a second primary colorectal cancer.²² All persons, healthy, cured and not cured, run a risk of dying from all other causes, determined by Dutch life tables corrected for colorectal cancer death.¹²

The survival distribution is first estimated by an iterative nonlinear least squares regression, weighted for the numbers of death, based on survival figures from the Norwegian and SE-N cancer registry.^{23,24} Then, by using incidence and survival the model determines expected mortality: the model starts from observed incidence and calculates expected numbers of death, given a stated survival and cure rate. Combinations of cure rates and survival periods will lead to age specific estimates of mortality, which may or may not be different from observed estimates (see appendix). Numerous pairs of cure rates and survival periods have been tested. The variance between calculated and observed numbers of deaths is tested by assuming a Poisson distribution of the probability of death (see appendix).²⁵ If the calculated numbers differ significantly from the observed ($p < 0.05$), that specific pair of cure rate and survival period is rejected as unlikely.

Results

For both sexes, rates of first admissions increase over the period 1978-1989, while mortality remains stable or declines (see figure 1). The cancer incidence and mortality of the South-Eastern Netherlands show the same trend, although the incidence mortality ratios are somewhat higher (see table 1). The cancer incidence of 1989 corresponded closely with the first admission rates of the hospital register of the same year (see figure 1).

Table 2 shows the annual changes by age and sex, estimated by the log-linear age-period models. Over this relatively short period, first admission rates increased with 39.9 % (M) and 20.4 % (W), rates of major interventions after primary diagnosis increased with 32.6 % (M) and 14.0 % (F), but mortality nearly remained stable for men (+ 2.2 %) and declined substantially among women (- 15.3 %). Hence, clinical incidence increased

with more than 35 % relative to mortality, and incidence of major interventions increased with 30 %. The standard errors (SE) in table 1 show that this increase of clinical incidence compared with mortality is highly significant. Major intervention rates during primary admission increased 5 % less steeply than first admission rates; this difference is statistically not significant, but probably indicates a shift from major surgery towards more non invasive colonoscopic treatment for early lesions.

The case fatality rate in these age groups declined quite strongly: from 5.6 % (SE 0.5) for men and 5.9 % (SE 0.5) for women in 1978-1980 to respectively 2.5 % (SE 0.3) and 2.4 % (SE 0.3). However, as shown in the same table 1, this decline in lethality can only explain the smaller part of the observed difference between morbidity and mortality trend: late mortality, excluding deaths during first admission, increased with 9.2 % among men and decreased still among women (- 9.5 %).

Table 2

Annual change of admissions, major interventions and deaths by age and sex over the period 1978-1989; 'late deaths' refers to all patients dying after surviving primary diagnosis and treatment, and corrects for changes in operative lethality.

Men; annual change in % (standard errors between brackets.)

	1st admissions	interventions	all deaths	late deaths
30-34	0.0 (1.9)	-2.2 (2.2)	-5.5 (3.1)	-4.7 (3.2)
35-39	0.2 (1.5)	-0.4 (1.7)	-3.1 (2.3)	-2.4 (2.4)
40-44	5.9 (1.5)	4.2 (1.2)	0.5 (1.7)	0.7 (1.8)
45-49	3.1 (0.9)	2.4 (0.9)	-0.8 (1.3)	-0.4 (1.4)
50-54	3.3 (0.7)	2.4 (0.7)	1.2 (1.0)	1.9 (1.0)
55-59	3.4 (0.6)	3.0 (0.6)	0.0 (0.8)	-0.5 (1.1)
60-64	2.8 (0.5)	2.6 (0.5)	0.7 (0.6)	1.3 (0.7)
30-64	3.1 (0.3)	2.6 (0.3)	0.2 (0.4)	0.8 (0.4)

Women; annual change in % (standard errors between brackets.)

	1st admissions	interventions	all deaths	late deaths
30-34	2.6 (2.1)	3.2 (2.4)	-0.3 (3.7)	-0.6 (3.8)
35-39	-0.2 (1.6)	1.4 (4.2)	-2.2 (2.5)	-1.8 (2.6)
40-44	0.1 (1.2)	-0.1 (1.2)	-2.6 (1.8)	-2.4 (1.8)
45-49	2.2 (0.9)	1.0 (1.0)	-1.9 (1.4)	-0.9 (1.4)
50-54	2.0 (0.7)	1.1 (0.7)	-1.3 (1.0)	-0.5 (1.1)
55-59	1.1 (0.6)	1.0 (0.6)	-1.5 (0.8)	-1.1 (0.8)
60-64	2.2 (0.5)	1.6 (0.5)	-1.4 (0.7)	-0.7 (0.7)
30-64	1.7 (0.3)	1.2 (0.3)	-1.5 (0.4)	-0.9 (0.4)

Table 3 shows relative survival rates which fit closely the observed incidence and mortality rates of two three year periods at the beginning and at the end of the study period. Next to these simulated relative survival rates are the figures from Scandinavian and Dutch cancer registries originally used for estimation of the survival distribution. The simulated (fitted) survival rates are slightly lower than the observed. This can be expected if the hospital register misses a few very early lesions with good prognosis, curable by non-invasive procedures (see discussion).

The relative survival expected for 1987-89 predicts that cure rates (equal to the 10 year relative survival rate after hospital discharge) have increased with 11.8 % points, given a constant medium survival time: from 41.8 % (range: 38.8 - 44.5) in 1978-80 to 53.6 % (range: 51.2 - 55.1) in 1987-89. In other words: for every 100 patients (< 65 y) admitted in 1978-80 6 died in hospital and 55 in the subsequent years; 9 years later less than 3 died in hospital and 45 will die in the subsequent years. 13 extra patients survived a colorectal cancer diagnosis, 3 thanks to lowered hospital mortality.

Figure 3 shows areas of all pairs of cure rates and median survival time which fit the observed incidence and mortality within 95 % confidence limits (see appendix). As median survival increases, the corresponding cure rate has to decrease to fit the given mortality. Indeed, two processes can explain any change of mortality rates, given incidence: death from colorectal cancer can be cancelled (hence, persons are cured; arrow a in figure 3) or postponed (hence, median survival is increased; arrow b in figure 3). Obviously, both processes can also take place at the same time (arrow c in figure 3). For example: the incidence and mortality data of 1978-1980 can be explained by a median survival of 1.5 year and a cure rate of 0.44 or by a median survival of 3 year and a cure rate of 0.37. The incidence and mortality figures of 1987-1989 can be explained by a median survival of

Table 3

Published relative survival rates from Scandinavian countries(23, 26, 34) and the Southeastern Netherlands (SE-N)(24) are compared with the simulated relative survival, which fits best observed first admission and mortality rates from the periods mentioned.

Time after diagnosis	Observed					Simulated	
	Finland 1953-74	Norway 1972-75	Sweden 1970-74	Sweden 1975-79	SE-N 1975-85	Netherlands 1978-80	Netherlands 1987-89
2 m	-	-	88,6	90,6	-	94,2	97,5
1 y	54,2	67,1	70,5	71,7	-	68,0	74,3
3 y	-	47,8	-	-	-	48,2	58,8
5 y	31,1	42,7	45	46,8	46,5	42,6	54,6
10 y	30,2	-	41,9	-	42,2	39,4	52,3

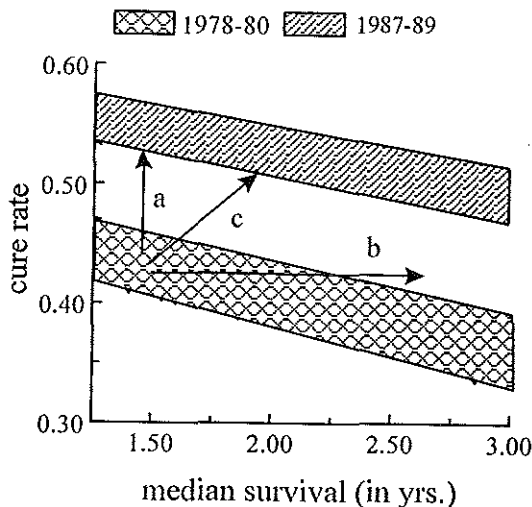


Figure 3 - Two way sensitivity analysis.

The two areas contain all pairs of cure rate and median survival of the not cured which will fit the given incidence and mortality rates of 1978-80 and 1987-89 within 95% confidence limits. The arrows indicate three hypothetical explanations for the difference between the two periods: increase of numbers which are cured (a), increase of survival of the not cured (b), or a combination of both (c).

1.5 year and a cure rate of 0.54 or a median survival of 3 year and a cure rate of 0.50. Figure 3 shows also that the model is much less sensitive to changes in assumptions about the median survival time than cure rate: to fit the same combinations of incidence and mortality, cure rates have to decrease with an average of 4.2 % points (a relative decrease of 10 %) when median survival increases with 1 year (a relative increase of more than 50 %). This is a consequence of the fact that more than 80 % of those who eventually will die of colorectal cancer will do so within 3 years of their diagnosis. Consequently, prolonged survival in the absence of cure can not explain much of the widening gap between incidence and mortality: cure rates must have improved considerably. The most plausible explanation for the increasing cure rates is increasing case detection. Earlier diagnosis probably improves effectivity of treatment⁹, and certainly increases the numbers of more benign lesions.

Discussion

The first question to be addressed concerns the validity of the data. We found increasing clinical incidence and decreasing mortality. For the considered age groups colorectal cancer mortality trends are generally valid.³

Artefactual trends may be caused by higher diagnostic efficacy in more recent periods: fewer fatal cancers are missed and more patients with widespread cancer have the site of primary origin of their cancer determined. But these changes in diagnostic practice will increase, not lower disease specific mortality.

Incidence data are collected from the hospital register, which is not constructed for epidemiological purposes. But less than 3 % of young and middle aged patients with a primary diagnosis of colorectal cancer will not be treated, thus hospitalisation rates for incident colorectal cancer are nearly complete.¹⁸ If anything, patients with early lesions curable by non-invasive colonoscopic procedures tend to be treated more often in outpatient clinics, which will cause an artefactual trend of admission decrease instead of increase. Only substantial changes in codification practices of both admissions for colorectal cancer and of major interventions might have biased trend estimates; we can not exclude a priori such changes, but they seem unlikely for such a short period and for a disease such as colorectal cancer. And finally, the Dutch hospital register shows the same trends as those observed in the regional cancer register in South Eastern Netherlands (table 1 and figure 1): similar analyses will yield similar results. Regional differences do exist, but nevertheless are small: the Netherlands are small, with a homogeneous population.

Our analysis might be weakened by the cross-sectional nature of the incidence and mortality data, biasing our assessment of (longitudinal) changes in prognosis. But the effects of therapy are period-, not cohort-dependent and most deaths from colorectal cancer occur within 3 years after primary diagnosis (table 3). The most recent mortality figures (1990-1992) suggest a sharpening decrease, particularly among men.¹² This would be inconsistent with the hypothesis that death is only postponed, not cancelled: such postponed deaths would cause a 'catch up' increase of mortality.

Increasing incidence and decreasing mortality of colorectal cancer can only be explained by a substantial improvement in prognosis. We have quantified the expected increase in survival needed: cure rates, defined as 10 year survival after hospital discharge, has to increase with 12 % points between 1978-80 and 1987-89; from 42 % to 54 %. The simulated survival distribution for the beginning of the period (1978-1980) was similar to the observed figures of cancer registries. The mortality soon after primary diagnosis decreased indeed, as has been observed elsewhere.^{26,27} This can be attributed at least partly to increases in therapeutic efficacy: safer intervention procedures and better post-operative care.²⁷ However, as shown in table 2, this change can only explain the smaller part of the observed divergence between morbidity and mortality. Recent advances have been made in the adjuvant treatment of advanced colorectal cancer, but such treatments were rarely applied before 1990 in the Netherlands.^{5,6,28} We didn't find any other indication of advances in treatment which might have benefitted more than a small subgroup (such as patients with solitary hepatic metastases), except for the short term results of surgery. Conse-

quently, the statement that the divergence between incidence and mortality trends has been caused by improved therapy, decreasing lethality, remains with very little support.

If we exclude less likely alternative hypotheses, such as decreasing malignancy of colorectal cancer in humans, the most probable explanation of the improving prognosis remains increased case detection. Earlier diagnosis deals with both incidence and mortality simultaneously:

1. Earlier diagnosis may account for decreasing lethality. While this remains a matter for debate, there is now at least some evidence that earlier diagnosis improves long term prognosis.⁹
2. Earlier diagnosis will increase incidence by shifting diagnosis towards an earlier age. In screening theory this is called lead time.^{7,29} Obviously, lead time alone would not cause an incidence increase: a tumour which is diagnosed at the earlier age ($a - t$) will not be diagnosed any more at age (a). But, in period ($a - t$) patients run a risk to die of other causes and colorectal cancer incidence increases sharply with age. The steeper the incidence increases with age, the more lead time will increase observed incidence rates, by moving diagnosis to younger ages. For colorectal cancer, shifting the whole age specific incidence curve of 1978-80 with one year towards a younger age causes an increase of the age standardized incidence with 12.5 %.

But advancing diagnosis one year in the natural disease progression will cause a shift of incidence by age which is always more than one year. This is caused by 'length time bias': the slower tumours grow, the likelier they are to be picked up by earlier diagnosis.^{7,29} How much incidence will increase, given earlier diagnosis, depends again on the relation of incidence with age and of the time distribution of the 'silent' period that tumours would have passed unnoticed previously, but are detected now. This distribution is unknown, but will reflect the variability of disease progression. A high variability implies many slowly growing tumours and a high potential to boost observed incidence. Such high variability seems likely; unsuspected macroscopic colorectal cancer during necropsy varies between 1 and 1.7 % in occidental countries, representing nearly 20 % of all incident colorectal cancers.^{30,31} Without early detection, many of those 'slow growers' will remain unnoticed because the person will have died before, from other causes.

We conclude that the increase in incidence of colorectal cancer and the concomitant decrease in mortality can not be caused by therapeutic improvements only. The most probable explanation of this divergence is increased cancer detection. This has important epidemiological and health policy implications:

- Increasing case-detection, that shifts cancer diagnosis to an earlier age both of the patient and the tumour, increases incidence (at least in age dependent cancers) and decreases mortality simultaneously, biasing both as indicator of underlying cancer hazards. Time series of stage at pri-

- mary diagnosis might confirm this; in the USA, where incidence and mortality diverge similarly, more early lesions are detected, while rates of distant disease remain stable.³² However, to evade stage migration bias, stages should be ascertained independent of modern imaging techniques.
- No recommended screening policy, or even proof of benefit, was needed to increase case detection, and induce a parallel increase in major interventions. It is worrying that we don't know how much of the induced morbidity increase is truly rewarded by a mortality decrease.

Appendix

The survival was modelled by using a discrete approximation of the lognormal distribution, with median survival μ , variance σ^2 and t_i the i -th month after diagnosis. The probability of surviving colorectal cancer till time t_i $P(S)_i$ is then given by

$$P(S_a)_{t_i} = (1 - CFR_a) \left(c_a + (1 - c_a) \sum_{\tau=t_0}^{t_i} \left(\frac{1}{\tau \sqrt{2\pi\sigma}} \exp \frac{-(\log \tau - \mu)^2}{2\sigma^2} \right) \right)$$

The subscript a refers to two age groups (under 55 and 55-64); these age groups have been introduced because younger persons had a lower CFR and a higher cure rate, resulting in a better survival. CFR is the fraction of persons, first admitted for colorectal cancer, dying in hospital within 2 months, and c (for 'cured') is the fraction of long term (10 y) survivors. Models incorporating sex as determinant for survival were not significantly better, and were ignored. c , μ and σ are estimated from observed survival figures, and then varied by the model (see further).

The presented cancer model is a subsector of a comprehensive public health model, modelling several diseases, and is implemented as a continuous time Markov chain, describing discrete sub-populations, cycling in one year steps. The continuous time specification allows multiple transitions in one time step.³³ To this aim, the parsimonious lognormal distribution, defined by two parameters, is translated into a sequence of exponential waiting time distributions, described by four parameters, simulating the lognormal survival distribution in the Markov model.

The subpopulation which survives mortality related to primary diagnosis and therapy but which is not 'cured', enters a first stage. They will leave that stage with transition probability p and median duration $-\ln(0.5)/P$. After leaving this stage, parameter y distributes survivors over two subpopulations in two separate terminal stages, with transition probabilities q and r leading to death from colorectal cancer. For the purpose of this paper, these parameters have no direct practical meaning, except for simu-

lating the lognormal survival distribution for a population of incurable colorectal cancer patients.

The mean (and median) survival of incurable patients is changed by the model through P . The variables y , q and r are kept constant: changing these parameters has the same effect of prolonging the disease process. Random values between 0 and 1 are generated for the cure rate c and for the transition probability P , and the calculated numbers of colorectal cancer deaths are then compared to the observed numbers by the scaled deviance (log likelihood ratio statistic). Let $O_{a,s}$ be the observed number and $E_{a,s}$ the expected calculated by the model, with a referring to all 5 year age-groups between 30 and 64 and s to sex, then the scaled deviance Δ^{25} is calculated by

$$\Delta = 2 \left[\sum_{a,s} O_{a,s} * \ln \left(\frac{O_{a,s}}{E_{a,s}} \right) - (O_{a,s} - E_{a,s}) \right]$$

If the calculated numbers differ significantly ($P < 0.05$), the pair of c and P values is rejected as unlikely.

References

1. Adami HO, Bergstrom R, Sparen P, Baron J. Increasing cancer risks in younger birth cohorts in Sweden. *Lancet* 1993;341:773-77.
2. Davis D, Dinse G, Hoel D. Decreasing cardiovascular disease and increasing cancer among whites in the United States from 1973 through 1987. *JAMA* 1994;271:431-37.
3. Doll R, Peto R. The causes of Cancer. Sources of bias in estimating trends in cancer mortality, incidence and curability (appendix C). *J Natl Canc Inst* 1981;66:1270-81.
4. Rutqvist LE, Mattson B, Signomkla T. Cancer mortality trends in Sweden 1960-1986. *Acta Oncol* 1989;28:771-75.
5. Krook JE, Moertel CG, Gunderson LL, et al. Effective surgical adjuvant therapy for high risk rectal carcinoma. *N Engl J Med* 1991;324:709-715.
6. Slevin ML, Gray R. Adjuvant therapy for cancer of the colon. *Br Med J* 1991;302:1100-01.
7. Black WC, Welch HG. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med* 1993;328:1237-43.
8. Ahlquist DA, Wieand HS, Moertel CG, et al. Accuracy of fecal occult blood screening for colorectal neoplasia. A prospective study using Hemoccult and HemoQuant tests. *JAMA* 1993;269:1262-7.
9. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. *N Engl J Med* 1993;328:1365-71.

10. Coebergh JWW, Crommelin MA, Kluck HM, van Beek M, van der Horst F. Breast cancer in Southeastern Netherlands: trends in incidence and earlier diagnosis in an unscreened female population (*in Dutch*). *Ned Tijdschr Geneesk* 1990;134:760-65.
11. Feinstein AR, Sosin DM, Wells CK. The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med* 1985;312:1604-8.
12. Netherlands Central Bureau of Statistics. Death by causes of death, age and gender 1978-1989: Netherlands Central Bureau of Statistics, published annually.
13. Breslow NE, Day NE. Rates and rate standardization. *in*: II. The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987:48-79.
14. Netherlands Cancer Registry. Incidence of cancer in the Netherlands 1989. First report.: Netherlands Cancer Registry, 1992.
15. Eindhoven Cancer Registry. Cancer incidence and survival 1975-1987 in Southeastern Netherlands: Comprehensive Cancer Center South, 1991.
16. Netherlands Central Bureau of Statistics. Atlas of cancer mortality in the Netherlands 1979-1990: Netherlands Central Bureau of Statistics, 1992.
17. Stichting Informatiecentrum voor de Gezondheidszorg. Classification of interventions (*in Dutch*): SIG, 1977.
18. Coebergh JWW. Epidemiological aspects of cancer in the elderly (*in Dutch*). *The Practitioner* 1990;10:753-57.
19. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Stat Med* 1987;6:449-67.
20. Rutqvist LE. On the utility of the lognormal model for analysis of breast cancer survival in Sweden 1961-73. *Br J Cancer* 1985;52:875-83.
21. Hakulinen T, Pukkala E, Hakama M, Lehtonen M, Saxen E, Teppo L. Survival of cancer patients in Finland in 1953-1974. *Annals of clinical research* 1981;3 suppl 31:32-37.
22. Harvey EB. Second Cancer Following Cancer of the Colon and Rectum in Connecticut, 1935-82. *Natl Cancer Inst Monogr* 1985;68:99-112.
23. Cancer Registry of Norway. Survival of cancer patients. Cases diagnosed in Norway 1968-1975: Cancer Registry of Norway, 1980.
24. Coebergh JWW, Crommelin MA, van der Heijden LH, Hop WCJ, Verhagen-Teulings MT. Survival of cancer patients in southeastern Netherlands in the period 1975-85 (*in Dutch*). *Ned Tijdschr Geneesk* 1991;135:935-943.
25. Breslow NE, Day NE. Fitting models for grouped data. *in*: II. The design and analysis of cohort studies. Lyon: International Agency for Research on Cancer, 1987:120-176.
26. Enblad P, Adami HO, Bergstrom R, Glimelius B, Krusemo U, Pahlman L. Improved survival of patients with cancers of the colon and rectum? *J Natl Cancer Inst* 1988;80:586-91.
27. Gordon NLM, Dawson AA, Bennet B, Innes G, Eremin O, Jones PF. Outcome in colorectal adenocarcinoma: two seven-year studies of a population. *Br Med J* 1993;307:707-10.

28. Wagener DJT, Wils AJM. Progress in adjuvant treatment of colon and rectum cancer (*in Dutch*). Ned Tijdschr Geneesk 1990;134:1689-1691.
29. Cole PH, Morrison AS. Basic issues in population screening for cancer. J Natl Canc Inst 1980;64:1263-72.
30. Berge T, Ekelund G, Mellner C, Pihl B, Wenckert A. Carcinoma of the colon and rectum in a defined population. An epidemiological, clinical and postmortem investigation of colorectal cancer and coexisting benign polyps en Malmo, Sweden. Acta Chir Scand 1973;438 (suppl):1-86.
31. Delendi M, Gardiman D, Riboli E, Sasco AJ. Latent colorectal cancer found at necropsy. Lancet 1989;i:1331-2.
32. Chow WH, Devesa SS, Blot WJ. Colon cancer incidence: recent trends in the United States. Cancer Causes Control 1991;2:419-25.
33. Chiang C. An introduction to stochastic processes. New York: Wiley, 1968.
34. Finnish Foundation for Cancer Research. Cancer in Finland 1954-2008. Helsinki: Cancer Society of Finland, 1989 (vol 42).

Senescence as cause of hip fractures

Abstract

Aim: To estimate theoretically the numbers of hip fractures among Dutch women caused by senescence in contrast to decreasing post-menopausal bone loss.

Design: A Markov chain model in a multi-state life table.

Method: Cross-sectional data of prevalences of bone mineral density, measured in an epidemiological survey, are related to numbers of hospital admissions for hip fractures among women of age 55 and older, using a (published) relative risk for hip fracture of 2,6 (95 % confidence limits 1,9-3,6) per decrease of 1 SD of bone mineral density. We vary the prevalences of bone mineral density in the multi state life table and estimate expected changes in the number of fractures.

Results: If bone loss ceases after the age of 55, this would decrease the number of fractures with 39.4% (95 % confidence limits 30.4%-45.2%). If bone loss is delayed by 7.5 years, the number of hip fractures would decrease by 14.8 % (10.0% - 19.9% after 5 to 10 year delay). If compliance decreases by age, (3 % per year), then the number of hip fractures would decrease by 5,6 % (3.4% - 11 % after 1%-5% decrease of compliance).

Conclusion: Bone loss is but one of the many age related causes of hip fractures. Population based interventions targeting only bone mineral density, while ignoring comorbidity, have only a limited potential for preventing hip fractures.

Introduction

In the Netherlands, as in many other countries of the west, the incidence of hip fractures in both men and women over the age of 65 has increased, exceeding the number expected based on the aging of the population alone.¹ This is assumed to be associated with the decreasing amount of physical activity in the population, as a result of which bone density is lowering.² But bone density is not the only determinant of hip fractures: over 90% of all hip fractures result from a fall.³ Important determinants for the development of hip fractures are, next to bone mineral density, sex, age, bone quality and various factors which can increase the tendency to fall, which we have collectively assigned the term 'comorbidity' (figure 1a).^{4,5}

The incidence of hip fractures over the age of 65 is twice as high in women as in men.¹ The maximum bone density (the peak bone mass reached at age 20-30) is lower in women, and next to the age-dependent bone loss common for both sexes, an accelerated post-menopausal (estrogen-dependent) loss occurs among women.²

The tendency to fall increases in old age due to comorbidity factors such as equilibrium disorders, deteriorating vision, decreased mobility and problems such as cardiovascular diseases and dementia.³⁻⁵ The chance of a fracture also increases due to diminishing protective reflexes.⁶ A longer life, particularly of persons with an elevated risk of falling, may explain part of the increasing number of hip fractures as caused by comorbidity. It was demonstrated earlier how decreasing mortality yields an increasing number of chronically ill.^{7,8}

The object of this study was to estimate, with the help of a simulation model, the number of hip fractures in Dutch women over the age of 55 caused by decreasing bone density. We estimated subsequently the number of hip fractures which could be prevented by preventing or slowing down the rate of bone loss.

Methods

Model

Figure 1a displays a simple conceptual model: the probability of a hip fracture is determined by the risk of falling and degree of bone mineral density. The degree of bone mineral density is determined by the level of peak bone mass and rate of bone mass loss. Bone mass loss is determined by age (this loss is accelerated in postmenopausal women) and comorbidity. The risk of a fall is also affected by these factors. Other determinants of peak bone mass and bone loss are discounted. Figure 1 b provides a highly simplified representation of the model used: the risk of a hip fracture is determined by age and by age-related decrease in bone density. The existence of disorders which affect bone quality more than bone quantity are ignored and increases in bone mineral density are not allowed.

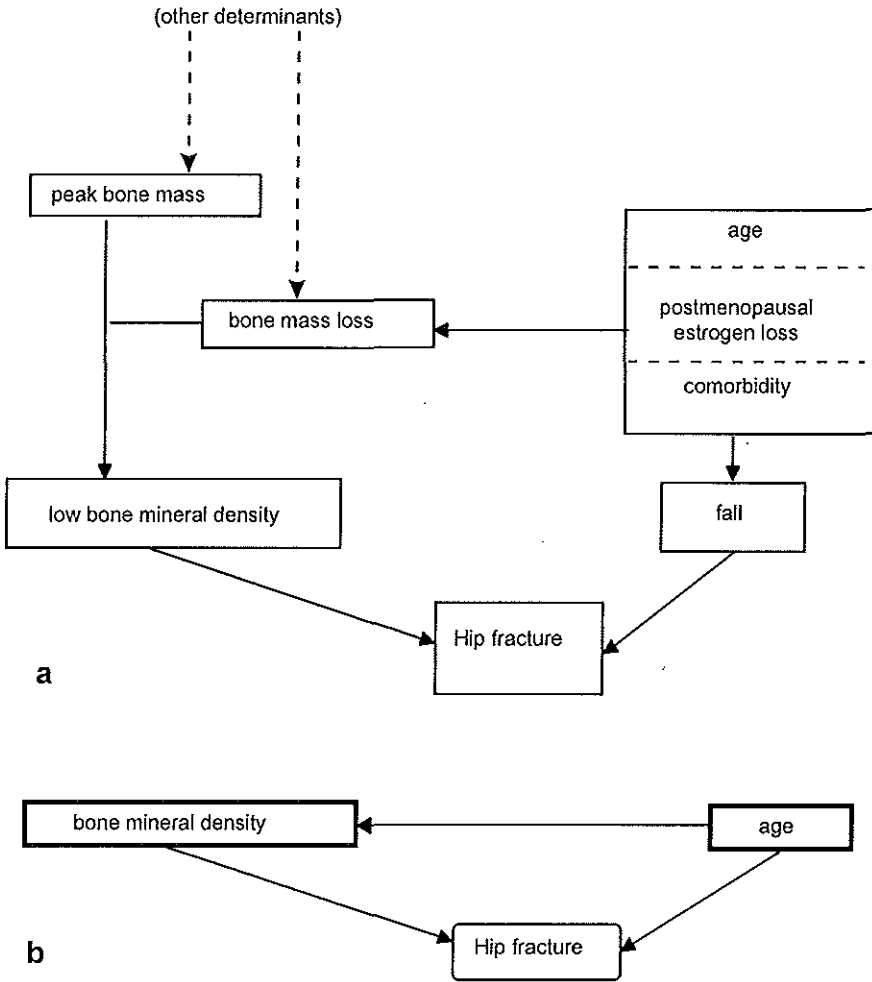


Figure 1 - Conceptual (a) and simplified (b) model for the development of a hip fracture

Technically speaking, the model used was a Markov chain. A so-called discrete Markov chain divides a population into a discrete number of homogenous states; transition probabilities determine the chance of a population changing from one state to another during a space of time (in our model one year). Markov chains have no memory: the future state is determined solely by the current state, not by any past state. The present, highly simplistic model is unable, for example, to relate future bone loss to the bone loss in the past. For this to be possible, the population would have to be redivided into homogenous groups, e.g. groups of persons in which the rate of bone loss is high, and those in which this is slow.

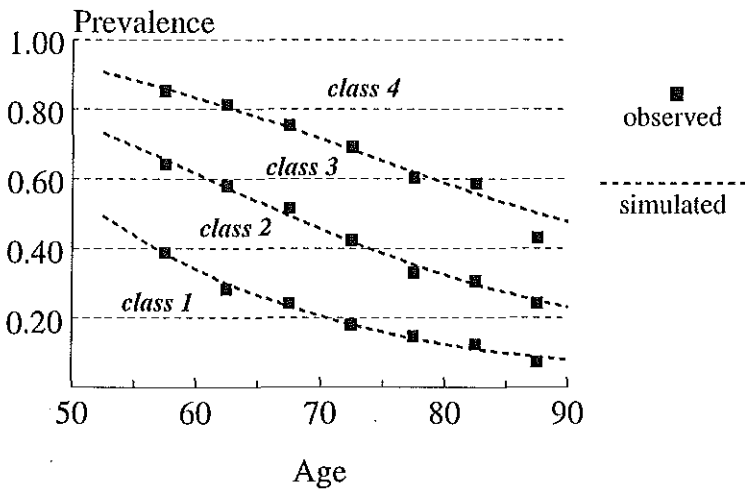


Figure 2 - Observed and simulated prevalences of bone mineral density class by age (BMD class stated between the curves). The prevalence is the area between two curves.

The model used described four classes of prevalence of bone density, ranging from class 1 (highest), to class 4 (lowest), in the female population between 50 and 89 years of age, and related these to the risks of fracture and mortality in a life table. A cohort entered the model at the age of 50 and subsequently faced an annual transition probability of moving from a higher to a lower bone density class. The model optimized the transition probabilities till the prevalence of bone mineral density found in the Rotterdam Study was produced.⁹

This was done by minimizing the maximal likelihood, calculated by the following equation:

$$\sum \left(2 * (O_a * \left(\ln \frac{O_a}{E_a} \right)) - (O_a - E_a) \right)$$

in which a stands for a particular 5-year age group, O is the number of women observed in a defined bone mineral density class (or the number of hip fractures, depending on what we were calculating) and E for the number expected by the model. The sum Σ was rendered as small as possible by using a numeric method of calculation through an iterative changing of the transition probabilities; the model consequently corresponded maximally with the observed reality.

Based on the prevalence rates of the different classes by age (figure 2), the relative risks for hip fractures according to bone mineral density class, and the general age-specific fracture probability by age it then became possible to determine the age-specific fracture probability according to bone density class. The model took into account the risk of dying from other

causes of death (calculated on the basis of the Dutch survival table) and from the consequences of a hip fracture.¹⁰

Data

The bone density data were derived from the Rotterdam study.⁹ The study population consisted of 7983 men and women of age 55 and up. We used only cross-sectional data of the 3375 women in whom the femoral neck bone density had been determined with the help of the dual energy X-ray (DEXA) method.⁹ The bone density distribution of this population was divided into 4 equal quartiles, class 1 through class 4; the bone density of class 1 was > 0.90 g/cm², of class 2 $< 0.90 \geq 0.81$ g/cm², of class 3 < 0.81 and ≥ 0.72 g/cm² and in class 4 < 0.72 g/cm². The prevalence rates for each bone density class were then determined according to 5-year age group (see figure 2).

We obtained the numbers of hip fractures in 1993 in the Netherlands from the Dutch hospital register and calculated fracture rates by using the Dutch population in the middle of that year as person-years follow up.¹¹

We adopted the relative risk of a hip fracture according to bone density class from an article by Cummings et al.; in their population, the risk of a hip fracture increased by a factor of 2.6% (95% confidence limits: 1.9-3.6) per decrease by 1 standard deviation of bone density.¹² This relative risk was converted to bone density class specific risks based on the means of the 4 bone density classes and the standard deviation of the overall distribution of the Rotterdam study group (table 1). The lower and upper limits of the 95% confidence level were used for the sensitivity analyses.

Scenarios

Various scenarios could now be calculated by varying the transition probabilities between the four bone density classes; the age-specific relative figures were subsequently converted into absolute figures for the Dutch female population of 1989.

Table 1 - Relative hip fracture risk for different classes of bone mineral density (BMD)

BMD class	mean BMD (in g/cm ²)	relative decrease (in SD)*	relative risk for fracture (95% CI †)
1	0.98	0‡	1‡
2	0.85	1.1	2.7 (2.0 - 3.9)
3	0.77	1.7	5.1 (3.0 - 8.8)
4	0.65	2.6	11.8 (5.2 - 27.2)

95% CI = 95% confidence interval

* 1 SD = 0.13 g/cm²

† Calculated using published data.¹²

‡ BMD class 1 is the reference class

Results

Experiments were performed to test three lines of reasoning:

Halting of osteoporosis

The first experiment is comparable to the elimination of a cause of death, a much-used technique in life tables.¹³ In an experiment of this kind, a specific cause of death is eliminated and the theoretical cohort will subsequently die from other causes of death based on (independent) substitute mortality risks; this technique illustrates the huge significance of substitute mortality. In our experiment, however, not a cause of death but a cause of disease (namely decreasing bone density) was eliminated and the importance of the other age-related factors (in particular increasing comorbidity) was demonstrated. The decrease in bone density was artificially halted by reducing all transition probabilities between the bone density classes to zero: women consequently remained in the same bone density class to which they had belonged at age 55 for the rest of their lives. The fraction of prevented fractures thus depended on the bone density-dependent risk for fractures, the sole parameter in the model not estimated on the basis of Dutch data. For this reason, this parameter was varied in a sensitivity analysis.

The result demonstrated a reduction in the number of hip fractures by 39.4% (outer limits: 30.4 - 45.2) (table 2 and figure 3). The upper and lower limits were based on the 95% confidence intervals for the relative risks in the study by Cummings et al.¹² In other words: by completely halting all bone loss from age 55 up, 40% of all hip fractures could be prevented.

Slowing the osteoporotic process

In the second experiment, the development of osteoporosis was not halted but slowed, from age 55 up. The model calculated the transition probabilities in such a way that bone density prevalences were achieved an average

Table 2 - Estimated effect of the number of hip fractures in three scenarios according to bone mineral density, versus the figures for 1993

	<i>mean estimation</i>	<i>lower and upper limits</i>
reference (1993)	9299 (100%)	
elimination bone loss	5640 (60.6%)	5093 (54.8%) - 6468 (69.6%)*
slowdown (7.5 years) of bone loss	7927 (85.2%)	7448 (80.1%) - 8369 (90.0%)*
decreased patient compliance [†]	8779 (94.4%)	8448 (90.8%) - 8974 (96.5%)*

* 95% uncertainty limits calculated by sensitivity analysis

† Assuming 80% of the women start at the age of 50 with a product which slows bone loss by 7.5 years and 3% discontinue therapy every year of life.

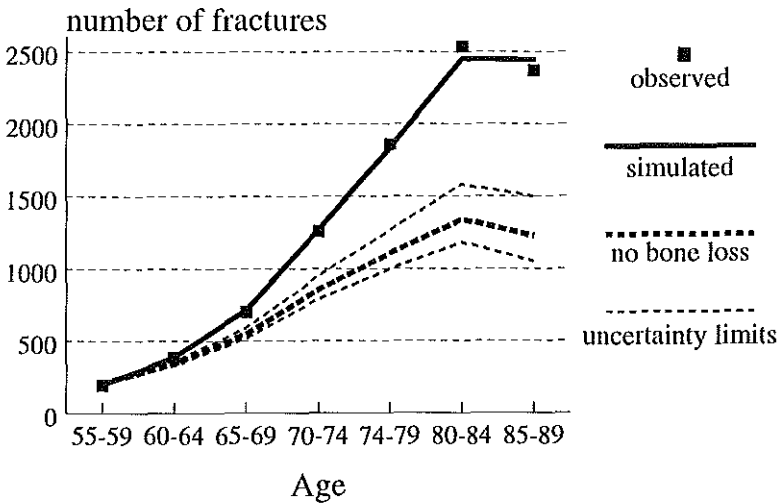


Figure 3 - Expected number of hip fractures if no further bone loss were to occur: number of hip fractures in 5-year groups in Dutch women (observed as hospital admissions in 1993) and the number of hip fractures expected with the Markov model demonstrating the fit of the model. The dashed curves represent the expected number of hip fractures if no bone loss occurred after age 50.

of 7.5 years (5 to 10 years in the extreme scenarios) later; this degree of delay in bone loss is found in women taking estrogen.¹⁴ In other words: the experiment showed a 'treated' female population of 70 years of age (assuming a patient compliance of 100%) with a bone density corresponding to an 'untreated' population of 62.5 years of age (65 to 60 in the extreme scenarios). This yielded a reduction in number of hip fractures of 14.8% (10.0%-19.9%). The gain was limited by the inexorable aging process (see table 2 and figure 4).

Decrease in patient compliance

The third experiment took into account a mixed population with incomplete and changing patient compliance. An initial 80% of the women was assumed to be taking medication at the age of 50 which slowed bone loss by 7.5 years, as was the case in the previous experiment, but of which 3% quit per year (1-5% in the extreme scenarios). This implied that by age 55, 69% (62%-76%) of the women were still taking the medication, a percentage which fell to 44% (29%-65%) by age 70. Those who discontinued the treatment experienced an accelerated rate of bone loss for three years until reaching the bone density they would have had without the protective drug.^{15,16} In this scenario, the number of fractures decreased by 5.6% (4.3-11.0). The reason for this was that elderly persons exposed to the highest risks for fractures had stopped their treatment without deriving any more benefit from this (see table 2 and figure 5).

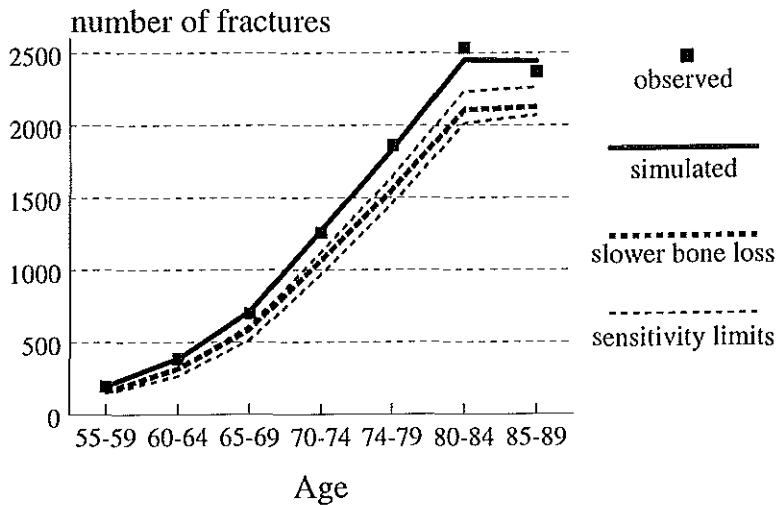


Figure 4 - Expected number of hip fractures if bone loss were to be slowed; numbers of hip fractures in 5-year groups in Dutch women (observed as hospital admissions in 1993) and the number of hip fractures expected with the Markov model. The dashed curves represent the expected number of hip fractures if bone loss were slowed by an average 7.5 (sensitivity limits 5-10) years.

Discussion

The development of hip fractures is largely associated with falling and with diminished bone strength.^{6,17} A hip fracture is nearly always the result of a fall, while the probability of a fracture is influenced inter alia by bone density. Falling is also caused by comorbidity³; calculations generated by the model we used suggested that at population level, primary prevention from middle age on solely aimed at bone density will not be very effective. The morbidity which increases with age (frailty) continues to remain a cause of falls, and if the woman's fall is an unhappy one, she will break her hip no matter what her bone density is. Age is an important predictive factor in respect of hip fractures, regardless of bone density.¹⁸

A Dutch study also demonstrated the low specificity and sensitivity of simple bone density measurements in the finger in relation to the development of non-vertebral fractures.¹⁹ Our study confirmed all this by integrating aggregated figures in a simple model: the fracture probabilities (observed in the Netherlands) increase far more (exponentially) with age than the reduction of bone loss (found in the ERGO study).

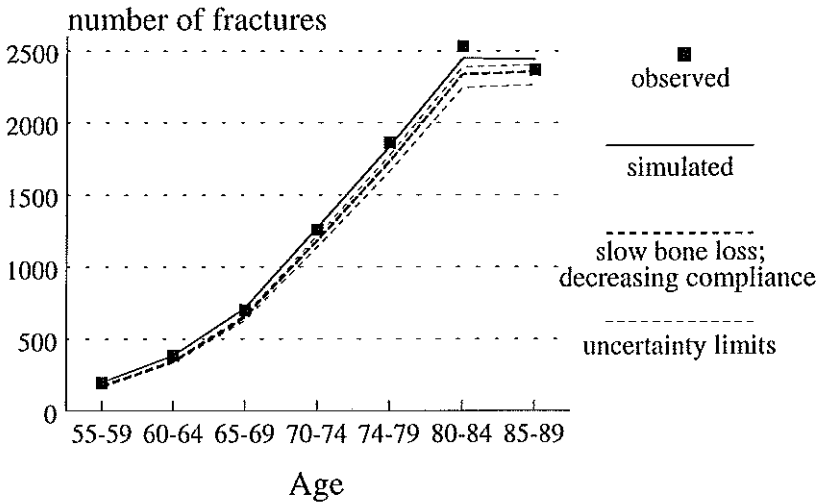


Figure 5 - Expected number of hip fractures with a decline in patient compliance in respect of osteoporosis inhibitors. Numbers of hip fractures in 5-year groups in Dutch women (observed as hospital admissions in 1993) and the number of hip fractures expected with the Markov model. The curves represent the expected number of hip fractures if patient compliance decreased as follows: 80% of the women start to take a product at age 50 which slows bone loss by 7.5 years and 3% (1 till 5 in sensitivity analysis) discontinue their therapy each life year.

Assumptions

The results of theoretical models are determined by assumptions. We assumed that the number of hip fractures registered by the hospitals were an adequate representation of the epidemiological state in the Netherlands. In view of the gravity of the problem and the need for medical intervention this would appear likely. We further assumed that the data on bone density from the ERGO file were representative for the Dutch population. Women in a nursing home were not included; the effect of the underestimated deterioration in bone density with age is compensated to a certain extent by the overestimated chance of fracture for the general population. Moreover, only a small percent of the female population is a resident of a nursing home. Finally, we (still) have not been able to obtain a good estimation for the risk of fracture according to bone density for the Netherlands. The ranges used in the sensitivity analysis were therefore deliberately kept wide: the risk of fractures in bone density class 4 was 5-27 times as high as in class 1. It is hardly likely that the true values in the Netherlands are outside this range.

Although the assumptions applied are consequently robust, caution must be exercised in interpreting the results of such simple models. The model itself is comparable to a survival table²⁰, which generates a synthetic cohort based on cross-sectional probabilities. The results must be inter-

puted like the life expectancy in survival tables, which would only be valid if all other probabilities remain constant in the future: changes in time are ignored.

Moreover, only the two determinative predictive factors of bone density and age existed in the highly simplified world of this theoretical model. A direct extrapolation to the real world would therefore be a far too simplistic representation. For example, if estrogen replacement not only improves bone density but also cardiovascular health, the effect of such replacement therapy will be far more pronounced than simulated here. Observational studies have revealed that postmenopausal women on hormone replacement therapy suffer 25-50% fewer hip fractures, a much greater reduction than predicted by our model²¹. Even if this may to some extent be discounted, it remains reasonable to assume that this cardiovascular improvement will affect the fracture figures.

The model, in this simplified form, is not intended to show the effects of treatment of osteoporosis (for example in women class 3 or 4) or to evaluate the efficiency of screening. It is most certainly not intended for judgments about clinical decisions for individual patients. The sole aim of this study was to indicate the quantitative significance of comorbidity in the aging process which causes hip fractures, for which this simple model is ideally suited. It demonstrates how, if a single determinant of hip fractures is eliminated (bone density loss), 'all other causes' remain which increase rapidly with age. The model demonstrates in quantitative terms that simply reducing the problem of hip fractures to a problem of bone densities is not the right approach.

References

1. Boereboom FT, Groot RR de, Raymakers JA, Duursma SA. The incidence of hip fractures in the Netherlands. *Neth J Med* 1991; 38:51-8.
2. Law, MR, Wald NJ, Meade TW. Strategies for prevention of osteoporosis and hip fractures. *BMJ* 1991; 303:403-9.
3. Grisso JA, Kelsey JL, Strom BL, Chiu GY, Maislin G, O'Brien LA et al. Risk factors for falls as a cause of hip fracture in women. *N Engl J Med* 1991; 324:1326-31.
4. Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, et al. Risk factors for hip fracture in white women. *N Engl J Med* 1995; 332: 767-73.
5. Cooper C, Barker DJ. Risk factors for hip fracture. *N Engl J Med* 1995; 332 814-5.
6. Lauritzen JB, McNair PA, Lund B. Risk factors for hip fractures. A review. *Dan med Bull* 1993; 40: 479-85.

7. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, Maas PJ van der. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure: the future rise of heart failure. *Am J Publ Health* 1994; 84:20-8.
8. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke trends in an ageing population. *Stroke* 1993; 24:931-9.
9. Burger H, Daele PLA van, Hofman A, Schütte HE, Birkenhäger JC, Pols HAP. Werveldeformaties en botdichtheid bij ouderen; het ERGO-onderzoek. *Ned Tijdschr Geneesk* 1995; 139: 2092-6.
10. Boereboom FT, Raymakers JA, Duursma SA. Mortality and causes of death after hip fractures in the Netherlands. *Neth J Med* 1992; 41:4-10.
11. SIG Zorginformatie. Bestanden van de Landelijke Medische Registratie over de Ziekenhuisopnames in 1993. Utrecht: SIG, 1994.
12. Cummings SR, Black DM, Nevitt MC, Browner W, Cauley J, Ensrud K, et al. Bone density at various sites for prediction of hip fractures. *Lancet* 1993; 341:72-5.
13. Namboodiri K, Suchindran C. Life table techniques and their application. Orlando: Academic press, 1987:92-109.
14. Felton DT, Zhang Y, Hannan MT, Kiel DP, Wilson PW, Anderson JJ. The effect of postmenopausal estrogen therapy on bone density in elderly women. *N Engl J med* 1993; 329:1141-6.
15. Christiansen C, Christiansen MS, Transbol I. Bone mass in postmenopausal women after withdrawal of oestrogen/gestagen replacement therapy. *Lancet* 1981; i:459-61.
16. Lindsay R, Hart DM, Maclean A, Clark AC, Kraszewski A, Garwood J. Bone response to termination of oestrogen treatment. *Lancet* 1978; i:1325-7.
17. Rubin CD. Southwestern internal medicine conference: prevention of hip fractures in the elderly. *Am J Med Sci* 1995; 310:77-85.
18. Melton 3d LJ, Kan SH, Frye MA, Wahner HW, O'Fallon WM, Rigs BL. Epidemiology of vertebral fractures in women. *Am J Epidemiol* 1989; 129:1000-11.
19. Hemert AM van, Vandenbroucke JP, Birkenhäger JC, Valkenburg HA. Prediction of osteoporotic fractures in the general population by a fracture risk score. A 9-year follow-up among middle-aged women. *Am J Epidemiol* 1990; 132: 123-35.
20. Manton KG, Stallard E. Chronic disease modelling. London: Griffins, 1988: 207-10.
21. Grady D, Rubin SM, Petitti DB, Fox CS, Black D, Ettinger B et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 1992; 117:1016-37.

The burden of morbidity and mortality from dementia

Abstract

The purpose of this study was to estimate severity-specific mortality and to quantify the global health burden of dementia by assessing the time spent disabled with dementia and the life years lost due to dementia. We used mortality data from the Rotterdam study, a population-based prospective study in the 55+ to calculate overall and severity-specific excess mortality for the demented. Lost life years were calculated by decomposing the (mixed) Dutch life table of 1990-92 in two populations, the demented and the healthy, using prevalence and excess (all cause) mortality. Healthy life loss was calculated by a modified Sullivan technique, weighting for disease severity. Our results indicate that mortality is increased in the demented, in all age, sex and severity groups. Mortality rate ratios are 2.1 (men) and 2.3 (women) with a range from 1.7 to 3.4 (men) and 2.0 to 3.1 (women), depending on severity. 55-year old men lose 1.2 life years due to morbidity and mortality and 0.7 life years due to mortality resulting from dementia. Women lose 3.1 and 1.9 life years respectively. This population-based study provides evidence that mortality is increased in the demented at all stages including minimal dementia. The quantified health impact on the general population is in the same order as that of lung cancer or stroke.

Introduction

Dementia is gradually being recognized amongst health policy decision makers as a major source of disability in the general population. Strategies aiming at prevention, early detection and treatment of the various types of dementia with a primary focus on Alzheimer's disease are increasingly be-

ing discussed and advocated (eg Mayeux, 1996; Whitehouse and Reidenbach, 1997).

Dementia is a common health problem; its prevalence rises exponentially after the age of 60. At least ten percent of the population beyond the age of 80 are affected in Western countries (Katzmann, 1976; Jorm, Korten and Henderson, 1987; Hofmann et al., 1991;). Dementia is associated with an increase in dependent life expectancy and therefore presents an important burden for both the individual and society (Kay, 1991; Katzmann and Kawas, 1994). The socioeconomic consequences will presumably become more critical as life expectancy increases and the post war baby boom cohort reaches old age.

There is sufficient empirical evidence which documents decreased survival in the demented (Roth, 1955; Martin et al, 1987; Evans et al, 1991; Skoog et al, 1993). The relationship between disease severity and mortality has, however, not clearly been established (Mayeux, 1996), probably because of lack of information on severity from epidemiological studies or of sufficient sample sizes. The Rotterdam study, a large population based longitudinal study of the 55+ (Hofman et al., 1991), permits estimation of age- and stage-specific excess mortality among the demented.

Meaningful indicators expressing the public health impact of a disease are the number of life years lost due to the disease and the number of life years spent, disabled, with the disease. One way of expressing the global burden of a disease is to use a composite measure of both. This type of health status index was first introduced by Sullivan (Sullivan, 1971). To allow for comparisons between lost life years due to death and due to disease-related changes in severity, the concept of disability-adjusted life years has been developed (Murray and Lopez, 1994). Disease stages are assigned weights between 0 and 1, where 0 denotes healthy and 1 the severest disability stage or death. The composite figure shows the loss of life due to excess mortality and disability.

Methods

Study population

The Rotterdam study is a community-based, prospective study among the population aged 55 years and over to estimate the prevalence, incidence and determinants of degenerative diseases in order to identify preventable causes of morbidity and mortality. The study population comprises the total elder population from a district of the city of Rotterdam. It consists of 3105 men and 4878 women (total 7983), aged 55 and older. They either lived at home or in one of six old peoples' homes which provide a wide range of care covering all stages of disability. In this population, 494 subjects (6.2%) were identified with dementia.

The approach to case-finding was a three-step procedure (Ott et al., 1995). The MiniMental State Examination (MMSE; Folstein M., Folstein

S. and McHugh, 1975) and Geriatric Mental State Schedule (GMS-A, organic level; Copeland, Dewey and Griffith-Jones, 1986) were used to screen the study population.

Screen-positive persons underwent further testing, based on the Cambridge Examination for Mental Disorders of the Elderly (Roth et al., 1988). The third step consisted of neuroimaging and clinical examinations by a neurologist and a neuropsychologist. This procedure was applied to 94% of the study participants. For the remaining 6% the relevant medical information was obtained from records of general practitioners and the regional institute for outpatient mental health care. Diagnosis was established on the basis of all available test results and other information by an expert panel using DSM-III-R criteria (American Psychiatric Association, 1987). Subjects were classified as having minimal, mild, moderate, or severe disease. The classification was based on the Clinical Dementia Rating (CDR; Hughes et al., 1982) scale and, if a CDR score was not available, on MMSE scores. MMSE cut-off points of 30-24 (minimal), 23-17 (mild), 16-11 (moderate) and 10-0 (severe) were selected on the basis of correlation with CDR scores.

Mortality

The number of deaths over a mean follow-up period of 2.34 years were available to establish all cause mortality risks. Mortality rates were calculated using person-years, by five-year age groups and sex. Mortality rate ratios were calculated for each age group and a combined estimate was obtained by the Mantel-Haentzel method for stratified analyses. A Poisson regression model was fitted to assess the association between disease severity and mortality, using the non-demented study population as reference group. The model had the general form $\log(\text{rate}) = \beta_0 + \beta_1 \text{sev} + \beta_2 \text{age}$. Severity was used as a categorical variable (5 stages, including healthy) and age as a continuous variable. Factoring of the age groups did not improve the model.

Life years spent with dementia and life years lost due to dementia

Life tables based on 1990-1992 vital statistic data from the Netherlands Central Bureau of Statistics were generated. We split the mixed life table population in two homogeneous subpopulations, the healthy and the demented, using the age- and sex-specific prevalence rates observed in the study and the age, sex and severity-specific mortality risks. The difference between the life expectancy of the total and the dementia-free population shows the effect of dementia on life expectancy.

To calculate a disability adjusted life expectancy life years lived by the synthetic life table cohort (the L_x column from a life table) are modified so that $HL_x = L_x (1 - P_{dx} D_{dx})$ (Murray, 1994; Barendregt, Bonneux and van der Maas, 1995). Where HL_x is the number of healthy years lived at age x , L_x is the number of years lived at x from the life table, P_{dx} is the prevalence of the various stages of dementia and D_{dx} is the disability weight assigned to a

state of dementia. The following weights were used: 0.6 for mild, 0.81 for moderate and 0.92 for severe (Murray and Lopez, 1996). For minimal dementia, not assigned in the Global Burden of Disease study, we used an intermediate weight of 0.3. In terms of health policy, a weight of 0.6 implies that a policy maker would not be able to choose between an intervention lengthening the life of 1000 healthy individuals for one year or an intervention lengthening the life of 2500 mildly demented individuals for one year (Murray and Lopez, 1994). 10 years gained among mildly demented patients is equivalent to 4 year gained among healthy individuals.

Results

The demented population was much older than the non-demented. There were more women than men in both the demented and the non-demented study population. Mild and moderate were the most prevalent disease stages (table 1).

Table 1 - Baseline characteristics of the study population

	Non-demented		Demented	
Total number	7489		494	
Sex				
Male	2986	(39.9%)	119	(24.1%)
Female	4503	(60.1%)	375	(75.9%)
Age group				
55 - 64	2706	(36.1%)	10	(2.0%)
65 - 69	1408	(18.8%)	12	(2.4%)
70 - 74	1260	(16.8%)	27	(5.5%)
75 - 79	973	(13.0%)	60	(12.2%)
80 - 84	622	(8.3%)	125	(25.3%)
85 - 89	359	(4.8%)	155	(31.4%)
90+	161	(2.2%)	105	(21.3%)
Mean age (range)	69.7	(55-106)	84.5	(58-101)
Severity of dementia				
Minimal			68	(13.8%)
Mild			204	(41.3%)
Moderate			148	(30.0%)
Severe			74	(15.0%)

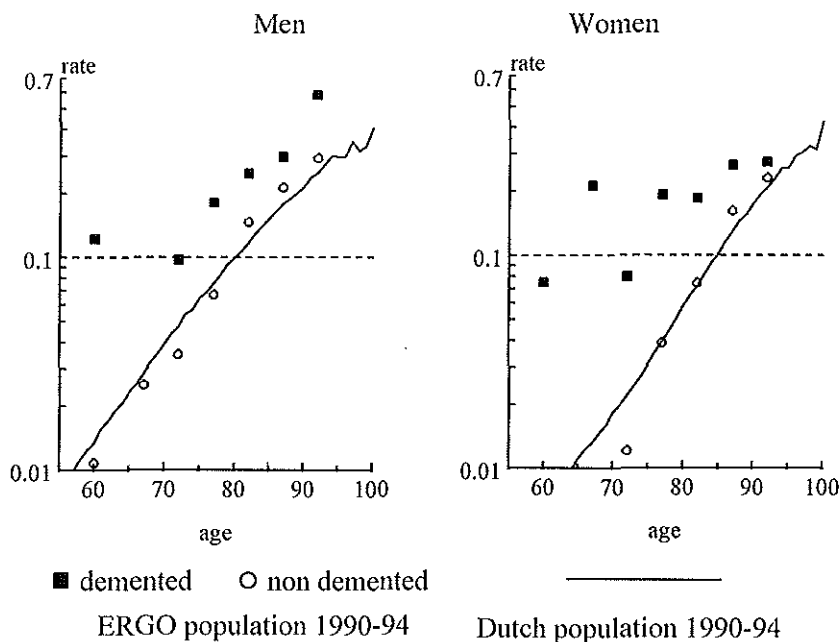


Figure - Mortality in men and women in the Netherlands (—) and in the non demented (■) and demented (○) study population

Overall and severity-specific mortality

Mortality in the study population compared well with corresponding 1990-92 mortality data from the total population of the Netherlands (figure). After a mean follow-up period of 2.34 years (standard deviation 1.28; range 0.05, 6.08), 7.9 % of the total study population had died. In contrast, 43.4 % of the demented population had died. After adjusting for age, the overall mortality rate ratio of the male and female demented population was 2.1 (95% confidence interval 1.5 to 2.9) in men and 2.3 (1.9 to 2.9) in women (table 2). In both men and women this ratio tended to decrease with age.

The mortality among the demented increased by disease severity. Poisson regression analysis confirmed this association (table 3). Remarkable is the increased mortality risk of minimal dementia: 1.7 (95% confidence interval 0.8 to 3.6) and 2.0 (1.1 to 3.6) times higher in men and women than in the healthy population. This difference was statistically significant for women. Results in men lack power because of the smaller sample. More severe dementia was associated with higher excess mortality; severe dementia increased the mortality risk by 3.4 (95% confidence interval 1.8 to 6.5) in men and 3.1 (2.1 to 4.4) in women.

Table 2

Mortality (deaths/1000/year) in the non-demented and demented populations, mortality rate ratios (MRR) and 95% confidence intervals

	Men				Women			
	Non-demented	Demented	MRR	95% CI	Non-demented	Demented	MRR	95% CI
Age group								
55-64	10.4	122.6	12.0	(0.3 - 74.0)	5.8	79.5	13.3	(0.3 - 84.0)
65-69	25.5	0.0	0.0	(0.0 - 11.0)	5.4	212.2	40.0	(7.0 - 60.6)
70-74	33.9	97.5	2.8	(0.3 - 10.9)	10.6	79.3	7.5	(1.4 - 25.6)
75-79	61.4	181.0	2.9	(1.1 - 6.5)	31.0	192.5	6.2	(3.1 - 11.5)
80-84	132.5	247.2	1.9	(0.9 - 3.4)	52.0	185.2	3.6	(2.3 - 5.6)
85-89	183.1	296.3	1.6	(0.8 - 3.1)	121.3	264.9	2.2	(1.5 - 3.1)
90+	176.2	579.2	3.2	(1.3 - 8.7)	202.1	273.4	1.4	(0.9 - 2.0)
Mantel-Haenszel estimate	combined		2.1	(1.5 - 2.9)			2.3	(1.9 - 2.9)

Table 3

Number (%) of deaths observed (by severity) and associated rate ratios (95% confidence intervals), adjusted for age in demented men and women

	Men				Women			
	Deaths observed (%)	Rate ratio ¹⁾	95%CI	p	Deaths observed (%)	Rate ratio ¹⁾	95%CI	p
Severity								
Minimal	7 (30.4)	1.7	(0.8-3.6)	0.184	12 (26.7)	2.0	(1.1-3.6)	0.017
Mild	18 (39.1)	1.7	(1.0-2.8)	0.035	63 (39.9)	2.4	(1.8-3.2)	0.000
Moderate	18 (48.6)	2.2	(1.3-3.7)	0.002	50 (45.1)	2.2	(1.6-3.1)	0.000
Severe	10 (76.9)	3.4	(1.8-6.5)	0.000	39 (63.9)	3.1	(2.1-4.4)	0.000
Total	53 (44.5)				164 (43.7)			
Age (per year)		1.1	(1.1-1.1)	0.000		1.1	(1.1-1.1)	0.000

¹⁾Based on Poisson regression

The burden of morbidity and mortality from dementia

Elimination of the excess mortality risks for dementia yielded 0.7 and 1.9 years for men and women at age 55 (table 4). In men the number of lost life years increased with age, in women it remained constant. Women lose 1.2 "disability-adjusted" life years through morbidity, men 0.5 life years in the male life table cohort. In total, 55 year old men lose 1.2 years of expected disability-adjusted life years, 5.3% of the life expectancy after elimination of dementia and women lose 3.1 years, 10.7% of the remaining life expectancy, through dementia. However, the dementia-free life expectancy is still much longer in women.

Discussion

We have established the global burden of dementia for a district of Rotterdam and other populations global comparable in terms of prevalence of dementia and mortality figures. Decreased survival has been shown by others for selected populations such as nursing home residents or patients attending outpatient geriatric services or for small or otherwise restricted population samples (Roth, 1955; Kay, 1962; Nielson, Homma and Biorn-Henrickson, 1977; Vitaliano et al., 1981; Barclay et al., 1985; Diesfeld, van Houte and Moerkens, 1986; Martin et al., 1987; Rodney et al., 1987; Knopman et al., 1988; Walsh, Welch and Larson, 1990; Evans et al., 1991; Heeren, van

Table 4

Life expectancy (LE), dementia-free LE, disability-adjusted LE, LE after elimination of dementia, lost life years and disability-adjusted lost life years in men and women at age 55, 65, 75 and 85

Age	Life expectancy	Dementia-free LE	Disability-adjusted LE	LE after elimination of dementia	Lost life years	Disability-adjusted lost life years
55	21.8	21.1	21.3	22.5	0.7	1.2
65	14.2	13.3	13.6	14.9	0.7	1.3
75	8.4	7.3	7.6	9.4	1.0	1.8
85	4.6	3.2	3.6	6.0	1.4	2.4
Women						
55	27.1	25.4	25.9	29.0	1.9	3.1
65	18.6	16.8	17.4	20.5	1.9	3.1
75	11.2	9.2	9.8	12.9	1.7	3.1
85	5.7	3.8	4.3	6.7	1.0	2.4

Hemert and Rooymans, 1992; Sulkava, Vaden and Erkinjuntti, 1992; Skoog et al., 1993) our analysis was population-based and allows more general conclusions. The most striking finding was significant excess mortality in the minimally demented, which is a valid disease category because of the careful diagnostic procedure.

This is in contrast with the finding of Evans (Evans et al, 1991) who used a cognitive instead of a global measure to assess severity. Our results indicate that minimal and mild dementia are not negligible when estimating the burden of disease, as is sometimes suggested (Ritchie et al., 1994).

The mortality risk increased with severity in our analysis. These results are biologically plausible and they are in agreement with most of the few published observations addressing the association of excess mortality and disease severity (Nielson et al., 1977; Vitaliano et al., 1981; Barclay et al., 1985; Martin et al., 1987; Cooper, Bickel and Schäufole, 1996; Heyman et al., 1996). The magnitude of the association in our study was similar to that reported previously. Looking at age effects, excess mortality was extremely high in the youngest in both men and women. A similar observation was made in demented nursing home residents (Van Dijk et al., 1992). The relative risks in the latter population were higher than in our analysis, as might be expected, because they needed care; the age-related gradient, however, was essentially the same. This is consistent with the concept that dementia in the young has a more severe course (Barclay et al., 1985; Diesfeld, van Houre and Moerkens, 1986).

We used changes in life expectancy and disability-adjusted life expectancy as a comprehensive quantitative indicator of disease impact on health. The magnitude of the impact of dementia is comparable with that of other major diseases as a cause of death, such as stroke or lung cancer (Bonneux et al., 1997).

This observation establishes dementia as one of the leading causes of death and disability and is consistent with an assumption made by Katzmann in 1976 (Katzmann, 1976) that dementia may rank as the fourth of fifth most common cause of death in the US.

A limitation of our analysis is the omission of nursing home populations with the possible consequence of underestimating mortality and disability. Furthermore, we did not differentiate between dementias with different aetiologies. This seemed justified, however, since there are only minor differences in mortality between the two major subgroups: Alzheimer's disease and vascular dementia (Van Dijk et al., 1991). We assumed that all excess mortality in the demented results from dementia. This may seem naive. In the elderly, death is rarely monocausal, but the result of a complex process involving several diseases. Death from vascular dementia, in particular, is often caused by the underlying vascular disease. However, Alzheimer's disease is the predominant cause of dementia and among these patients all excess mortality is most likely more attributable to dementia than anything else.

Quantification of the burden of disease, using a global comprehensive estimate, establishes dementia as one of the leading causes of mortality and

morbidity in the population of 55 and over. Mortality from dementia ranges in the same order as from lung cancer or stroke. These findings suggest that health policies should aim to eliminate dementia as cause of death and disability.

Appendix

We calculated mortality for the demented and non-demented by observing that the total mortality rate (M) is the average of the rates for demented (M^1) and non-demented (M^0):

$$M = p M^1 + (1 - p) M^0 \quad (1)$$

with p the prevalence of dementia. Furthermore by definition it is true that:

$$M^1 = M^0 R \quad (2)$$

with R the relative risk. By substituting (2) in (1) and rewriting we obtain:

$$M^0 = \frac{M}{pR + 1 - p} \quad (3)$$

for the risk of the non-demented, and through (2) the risk for the demented.

References

- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders, third edition, revised*. Washington, DC: American Psychiatric Association, 1987.
- Barclay LL, Zemcov A, Blass JP, McDowell FH. Factors Associated with duration of survival in Alzheimer's Disease. *Biol Psychiatric* 1985; 20: 86-93.
- Barclay LL, Zemcov A, Blass JP, Sansone J. Survival in Alzheimer's disease and vascular dementias. *Neurology* 1985; 35: 834-840.
- Barendregt JJ, Bonneux L, Maas PJ van der. Health Expectancy: from a population health indicator to a tool for policy making. Paper presented at the REVES 8 conference, Chicago, Illinois, 1995.
- Bonneux L, Barendregt JJ, Nusselder WJ, Van der Maas PJ. Preventing fatal diseases increases health care cost. *BMJ* 1997 (in press).
- Cooper B, Bickel H, Schäufele M. Early development and progression of dementing illness in the elderly: a general-practice based study. *Psychol Med* 1996;26: 411-419.

- Copeland JRM, Dewey ME, Griffiths-Jones HM. Computerized psychiatric diagnostic system and case nomenclature for elderly subjects: GMS and AGE-CAT. *Psychol Med* 1986; 16: 89-99.
- Crook T, Bartus RT, Ferris SH, Whitehouse P, Cohen GD, Gershon S. Age-associated memory impairment: proposed diagnostic criteria and measures of clinical change: Report of a National Institute of Mental Health workgroup. *Dev Neuropsychol* 1986; 2: 261-276.
- Diesfeld HFA, van Houte LR, Moerkens RM. Duration of survival in senile dementia. *Acta Psychiatr Scand* 1986; 73: 366-371.
- Evans DA, Smith LA, Scherr PA, Albert MS, Funkenstein HH, Herbert LE. Risk of death from Alzheimer's disease in a community population of older persons. *Am J Epid* 1991; 134: 403-412.
- Folstein M, Folstein S, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189-198.
- Heeren TJ, van Hemert AM, Rooymans HGM. A community-based study of survival in dementia. *Acta Psychiatr Scand* 1992; 85: 415-418.
- Heyman A, Peterson B, Fillenbaum G and Pieper C. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part XIV: Demographic and clinical predictors of survival in patients with Alzheimer's disease. *Neurology* 1996; 46:656-660.
- Hofman A, Rocca WA, Brayne C, Breteler MMB, Clarke M, Cooper B, *et al*. The Prevalence of Dementia in Europe: A Collaborative Study of 1980-1990 Findings. *Int J Epid* 1991; 20: 736-748.
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A New Clinical Scale for the Staging of Dementia. *Brit J Psychiat* 1982; 140: 566-572.
- Jorm AF, Korten AE, Henderson AS. The prevalence of dementia: A quantitative integration of the literature. *Acta Psychiatr Scand* 1987; 76: 465-479.
- Katzmann R, Kawas C. The Epidemiology of Dementia and Alzheimer's Disease. *Raven Press, Ltd.* 1994; 105-122.
- Katzmann R. The prevalence and malignancy of Alzheimer's disease: a major killer. *Arch Neurol* 1976; 33: 217-218.
- Kay DWK. Outcome and cause of death in mental disorders of old age: A long-term follow-up of functional and organic psychoses. *Acta Psychiatr Scand* 1962; 38: 249-276.
- Kay DWK. The Epidemiology of Dementia: A Review of Recent Work. *Reviews in Clinical Gerontology* 1991; 1: 55-66.
- Knopman DS, Kitto J, Deinard S, Heiring J. Longitudinal study of death and institutionalization in patients with primary degenerative dementia. *J Am Geriatr Soc* 1988; 36: 108-112.
- Martin DC, Miller JK, Kapoor W, Arena VC, Buller F. A controlled study of survival with Dementia. *Arch Neurol* 1987; 44: 1122-26.
- Mayeux R. Development of a National Prospective Study of Alzheimer's Disease. *Alzheimer's Disease and Associated Disorders* 1996; 10 (1): 38-44.
- Murray CJ, Lopez AD, ed. The Global Burden of Disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. *Boston: Harvard School of Public Health*, 1996: 990+xxxi. (The Global Burden of Disease and Injury Series, vol I).

- Murray CJL, Lopez AD. Quantifying disability: data, methods and results. *World Health Organization* 1994; 72 (3): 481-494.
- Murray CJL. Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bulletin of the World Health Organization* 1994; 72: 429-445.
- Nielson J, Homma A, Biorn-Henrickson T. Follow-up 15 years after a gerontopsychiatric prevalence study-conditions concerning death, cause of death, and life expectancy in relation to psychiatric diagnosis. *J Gerontol* 1977; 32: 554-561.
- Ott A, Breteler MB, Harskamp F, Claus JJ, van der Cammen TJM, Grobbee DE, Hofman A. Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study. *BMJ* 1995; 310: 970-973.
- Ritchie K, Robine JM, Letenneur L, Dartigues JF. Dementia-Free Life Expectancy in France. *Am J Publ Health* 1994; 84: 232-236.
- Rodney CP, Todorov AB, Elston RC, Constantinidis J. The Malignancy of Dementia. *Ann Neurol* 1978; 3: 559-561.
- Roth M, Huppert FA, Tym E, Mountjoy CQ. CAMDEX: The Cambridge examination for mental disorders of the elderly. *Cambridge UK: Cambridge Univ Press* 1988; 120-186.
- Roth M. The natural history of mental disorder in old age. *Br J Psychiatry* 1955; 101: 281-301.
- Skoog I, Nilsson L, Palmertz B., Andreasson L.A., Svanborg A. A Population-Based Study of Dementia in 85-Year-Olds. *New Engl J Med* 1993; 328: 153-158.
- Sulkava R, Vaden J, Erkinjuntti T. Survival in Alzheimer's disease (AD) and multi-infarct dementia (MID) in the 1980s. *Neurology* 1992; 42: 143.
- Sullivan DF. A single index of mortality and morbidity. *Health Serv Mental Health Admin Health Rep.* 1971; 86: 347-354.
- Van Dijk PTM, Dippel DWJ, Habbema JDF. Survival of Patients with Dementia. *J Am Geriatr Soc* 1991; 39: 603-610.
- Van Dijk PTM, van de Sande HJ, Dippel DWJ, Habbema JDF. The nature of excess mortality in nursing home patients with dementia. *J Gerontol: Med Sci* 1992; 47: M28-34.
- Vitaliano PP, Peck A, Johnson DA, Prinz PN, Eisdorfer C. Dementia and Other Competing Risks for Mortality in the Institutionalized Aged. *J Am Geriatr Soc* 1981; 11: 513-519.
- Walsh JS, Welch HG, Larson EB. Survival of Outpatients with Alzheimer-type Dementia. *Ann Intern Med* 1990; 113: 429-434.
- Whitehouse PJ, Reidenbach F. Guidelines for Early Identification of Alzheimer's Disease. *Alzheimer's Disease and Associated Disorders* 1997; 11: 61-62.

PART III

Multi-disease models

Introduction to part III

Introduction

In part II several disease models have been described and applied to analyze compression and expansion of morbidity on the disease specific level. This part and the next describe how the disease specific results can be aggregated to a general description of population health status, with this part concentrating on the problem how to model multiple diseases and causes of death simultaneously. Part IV deals with the question how the resulting multi-dimensional description of population health status can be mapped to a public health indicator.

There are many criteria possible to distinguish between types of modelling, but an important distinction at this point is the one between micro- and macro-simulation. Micro-simulation is a technique where the unit of analysis and description is the individual, with macro-simulation this unit is the population (or a part of it). This distinction is important because it determines how population heterogeneity is handled. With micro-simulation population heterogeneity is essentially not a modelling issue: the population is simply the sum of all individuals described, and each individual may have as many stochastically determined characteristics as is desirable (Habbema, de Vlas, Plaisier, & van Oortmarsen, 1996; van Oortmarsen, 1995). The hardest part is to extract the details without being overwhelmed by the generated information.

Macro-simulation, on the other hand, deals with whole groups of individuals at a time, and therefore these groups must be homogeneous at the level of description. But in reality populations are heterogeneous: people run different risks to contract various diseases and to die from them depending on age, sex, risk factor exposure, and genetic disposition. The standard way to deal with this heterogeneity is to sub-divide the population into groups that are sufficiently homogeneous for the purpose of the research question. The art of modelling is to decide when the subdivision has achieved a sufficient level of homogeneity, while keeping the number of groups manageable.

The disease models in the previous part all distinguished age and sex, and in addition at least two (healthy and diseased), but usually more health states (through distinction of more than one disease stage). This already generates a large number of sub-divisions, but when multiple diseases are modelled simultaneously the number of sub-divisions tends to explode because of disease interactions. The techniques described in this part are designed to allow an adequate description of multiple diseases and causes of death and their interactions, while avoiding an explosion of the number of sub-populations, and are as such particular to macro-simulation. In the General Discussion we will get back to this issue.

There are five disease and causes of death interactions that are of importance in a multi-disease model, i.e. the relations between total mortality and disease specific mortality, substitution and competition of causes of death, risk factors for diseases (which may be diseases themselves), mortality selection, and comorbidity. To model these interactions we distinguish two cases: homogeneity versus heterogeneity of disease risk. In both cases age and sex are standard dimensions of differentiation.

Homogeneity of disease risk

A standard assumption in demography when dealing with multiple causes of death is independence: mortality risk of any one cause does not depend on the risk of any other cause. It can be shown that when the mortality risks are described by hazards (also called 'force of mortality', and mathematically the probability to die from a given cause during the time interval Δt , with $\Delta t \rightarrow 0$) and under the assumption of independence, total mortality hazard is simply the sum of the disease specific hazards (Manton & Stallard, 1988). When in addition it is assumed that the hazards are constant in a given age-interval they are equal to disease specific mortality rates (Manton & Stallard, 1988).

Since disease specific mortality rates and total mortality probability are available from Statistics Netherlands it is possible under these assumptions to partition total mortality (after conversion to a rate) between a number of disease specific mortality rates and an 'all-other-causes' mortality rate simply by subtraction.

The convenience of this approach is that it describes mortality on the level of risks. A change in any of the mortality risk leaves the others, by virtue of the independence assumption, unchanged. When, for example, one cause of death is eliminated, the other causes are unaffected. But numbers of deaths from the other causes will increase: a lower mortality risk in one age group will result in a higher number alive in the next. A higher number of persons alive, subject to the same mortality risk results in a higher number of deaths from all the still existing causes of death.

This mechanism, called 'substitution of causes of death', thus describes the interaction between causes of death on the level of numbers by assuming independence on the level of risks. In chapter 12 we expand the

mechanism to include disease incidence and prevalence. A set of independence assumptions, again on the level of risks, is given, and proof is provided that these are indeed sufficient conditions to describe substitution of diseases and causes of death. In addition it is shown that comorbidity prevalence can be derived from the prevalence of specific diseases by multiplication. This latter property allows to omit the explicit modelling of comorbid states, which hugely diminishes the number of sub-populations to be distinguished.

Heterogeneity of disease risk

As already observed above, however, populations are not homogeneous in disease risk. Some people run higher risks on one or several diseases, for example smokers run higher risks than non-smokers on cardiovascular disease, a large number of cancers, chronic obstructive lung disease, and several more. Other well-known risk factors are hypertension and cholesterol. Also diseases themselves can act as risk factors for other diseases, for example diabetes, which is a risk factor for cardiovascular disease, renal disease, and others.

Heterogeneity of disease risk results in mortality selection: because the exposed to a risk factor have a higher risk to contract a disease and to die from it, they are being selectively removed from the population. This mortality selection lowers the prevalence of the risk factor: when a population consisting of a cohort of smokers and one of non-smokers is followed through time (none of the smokers quits, and none of the non-smokers takes up, and apart from the smoking the cohorts are comparable), then the prevalence of smoking in the population will decline with time, because the smokers have a higher risk of dying than the non-smokers (Manton & Stallard, 1988).

Changes in mortality selection cause a phenomenon called 'competition of diseases and causes of death'. When, for example, a cure for lung cancer would be found, the then surviving smokers would keep their higher risks on cardiovascular and chronic obstructive lung disease. Because the group at high risk in the population increases relative to the low risk group of non-smokers, the average risk increases: the incidence and mortality of cardiovascular and chronic obstructive lung disease in the population will go up after elimination of lung cancer mortality. Chapter 7 in part II describes a likely case of competition of cardiovascular causes of death, which was presumably sparked off by the huge improvement in survival after myocardial infarction that was achieved in the period 1985-1990.

Chapters 13 and 14 describe how multi-disease models can handle heterogeneity of disease risks. In chapter 13 the incidence of non-insuline dependent diabetes is estimated from prevalence and mortality. Since diabetes is a risk factor for a number of diseases the mortality used has to include the risk from those causes as well. Chapter 13 doubles as an ex-

ample of the usefulness of incidence-prevalence-mortality models to provide an incidence estimate in cases as difficult as non-insuline dependent diabetes.

This incidence is used in chapter 14 to model diabetes as a disease and as risk factor for heart disease. Once persons have developed diabetes they get assigned a higher risk on heart disease incidence than the non-diabetics, hence higher prevalence and mortality. To allow for this population heterogeneity the population is divided into non-diabetics and diabetics. This sub-division provides for 'local independence': within each sub-population the independence conditions of chapter 12 hold, only for the population as a whole they do not (Manton & Stallard, 1988). It is shown in chapter 14 that in such a heterogeneous population the prevalence of diabetes, and in particular the comorbidity of diabetes and heart disease, will rise if the improvement in heart disease survival extends to diabetics as well.

Risk factors in a homogeneous model

Modelling risk factor heterogeneity through sub-division of the population into groups that are homogeneous in their risk factor exposure is feasible. When the number of risk factors increases, this solution quickly becomes tiresome: the number of sub-populations needed is 2^n , with n the number of risk factors. In many instances that follow (the life tables of chapters 19, 20, and 24, as well as in the *Prevent Plus* model in the latter chapter) a simpler approach is used: disease incidence in a homogeneous population is a function of risk factor prevalence. Because of the homogeneous population mortality selection is not modelled. The relative merits and disadvantages of this simpler approach, as compared to the heterogeneous modelling presented in this part, will be discussed in chapter 27.

References

- Habbema, J. D. F., Vlas, S. J., de, Plaisier, A. P., & Oortmarssen, G. J., van (1996). The microsimulation approach to epidemiological modeling of helminthic infections, with special reference to schistosomiasis. *Am J Trop Med Hyg*, 55(5), 165-169.
- Manton, K. G., & Stallard, E. (1988). *Chronic Disease Modelling* (Vol. 2). London: Charles Griffin & Company Ltd.
- Oortmarssen, G. J. van (1995). *Evaluation of mass screening for cancer: a model-based approach*, Erasmus Universiteit Rotterdam.

Coping with multiple morbidity in a life table

Abstract

One of the applications of the multi-state life table is in the field of Public Health, with states defining various levels of health or functional ability. Another approach is to model Public Health by looking at the impact of individual diseases, but, unfortunately, then two practical problems arise: there are many diseases, and due to comorbidity people may be in several disease states simultaneously. Both problems tend to make the number of states in the life table impractically large.

In this paper we introduce the proportional multi-state life table. It is especially designed to cope relatively easily with a large number of diseases simultaneously, while allowing for comorbidity. We provide proof of validity and an example implementation for cardiovascular disease.

Introduction

The life table is a simple and much used instrument for assessment of population health status. The standard life table produces life expectancy at different ages, and the life expectancy at birth is probably still the most employed indicator of Public Health.

It is also a limited indicator, since it ignores morbidity. This limitation becomes more severe when the burden of disease shifts from acutely fatal, mostly infectious disease, to chronic degenerative, mostly non-communicable disease. After this so called 'epidemiologic transition' (Omran 1971), that has taken place in the developed and is underway in an increasing number of developing countries, the need for an indicator of population health status that includes morbidity becomes acute.

An important point of interest is the assessment of the health impact of specific diseases, both on morbidity and mortality, and the effect of

changes in incidence and survival, caused by autonomous trends and possible interventions. This could help explain observed trends in population health status, make projections of future developments, and perhaps clarify the issue of compression versus expansion of morbidity, that is being debated for quite some time now (Fries 1980; Olshansky, Rudberg et al. 1991).

Interest arises also from the need for optimal allocation of scarce resources in the health care sector. For example, the 1993 World Development Report from the World Bank made an estimate of the Global Burden of Disease by combining prevalence and mortality data from a large number of diseases with a measure of the seriousness of a disease (the DALY, Disability Adjusted Life Year), and with this indicator compared the cost-effectiveness of various proposed interventions (World Bank 1993; Murray and Lopez 1994).

To estimate the total burden of morbidity one has to look at many diseases simultaneously, because, unlike with mortality, people may suffer from various conditions simultaneously (comorbidity), in particular at older ages. The total impact of diseases on population health status and disability does not simply add up: the relation between diseases, health status and disability is a complex one, and more so with comorbidity (Verbrugge, Lepkowski et al. 1989). At high ages comorbidity becomes substantial, and it has implications for intervention possibilities and effects, and for the amount of disability caused.

The method of choice for the inclusion of morbidity is the multi-state life table (Schoen 1987). It is a well established method, backed by a considerable body of theory. In a multi-state life table the inclusion of multiple diseases while taking comorbidity into account is straightforward: just add states defining the combined prevalence of diseases. This works fine for a limited number of diseases, but soon becomes unwieldy when that number increases. For example, with only one state for each disease the number of states to be defined is $2^n - 1$, with n the number of diseases.

We present the proportional multi-state life table method, that makes the inclusion of multiple diseases better manageable and allows for comorbidity implicitly, without the need to define additional states. We implement the method for heart disease and stroke, and look at the effect of hypothetical but not unrealistic changes in incidence and survival on disease prevalence and comorbidity. Finally we discuss limitations and extensions of the method.

Method

Description

Given the large number of diseases, of which many are rare, in practice one will have to limit the number of diseases in the life table to the ones deemed most important. This means that a life table with specific causes of death

will always have to include a category 'all other mortality'. This mortality rate from all other causes can be obtained by subtracting the sum of the included disease specific mortality rates from the total mortality rate, as is done in cause elimination life tables. In the remainder we will distinguish two disease specific mortality rates, and one 'all other causes' mortality rate. Of course there will likewise be an 'all other causes' morbidity, for the present study, however, this is ignored.

The basic idea of the proportional multi-state life table method is very simple: express age specific disease prevalence, instead of in numbers, in proportions of diseased among all alive at that age. Under certain independence conditions the comorbidity rate of any number of diseases is then just the product of their respective rates. Sufficient independence assumptions for the case of two diseases and an 'all other mortality' cause of death are listed as A1-A3 in the next section, and they can be summarized as:

- the incidence of each disease should be independent from all causes of death, except its own disease specific mortality (A1);
- disease incidences are independent (A2);
- all causes of death are independent (A3).

The proportional multi-state life table is divided into sections: first a general section, that is in fact a standard cause elimination life table, and secondly, one section for each disease with an independent illness-death process, that describes disease incidence, prevalence, and disease specific mortality. Total mortality in the general section is calculated from an 'all other' mortality rate and the disease specific mortality rates from the disease sections. Therefore changes in the disease sections that cause a change in its disease specific mortality feed into the total mortality, and thus these disease specific changes are reflected in the total mortality experience of the life table cohort.

Because the disease specific sections are independent and do calculations in rates they need not take into account the mortality from other causes of death: the general section computes for each age the number of deaths from all causes, and the age and disease specific prevalence in numbers is just the prevalence rate of that disease times the number of years lived in the age interval by the cohort. The same holds for the number of incidents and disease specific deaths.

The age specific comorbidity rate from any number of diseases is then the product of their respective prevalence rates, which just like the disease specific prevalence rates can be multiplied by the number of years lived in the age interval in the general section of the life table to obtain comorbidity in numbers.

Because each disease specific section does not need to be aware of total mortality or disease specific mortalities from other diseases each section is quite independent from all other disease sections and only linked to the general section by its disease specific mortality.

Proof

To prove the method described above to be equivalent to a standard life table we must show that prevalence, comorbidity, and mortality thus calculated are identical to results from a standard life table. The following three propositions state this more formally (all three propositions concern age specific rates):

- I) The prevalence in the disease section is equal to the prevalence calculated in the presence of other diseases and causes of death.
- II) The comorbidity of the two diseases is the product of their respective prevalences.
- III) The disease specific mortality in the disease section is equal to the disease specific mortality calculated in the presence of other causes of death.

The general independence condition for $n-1$ number of diseases and 1 'all other causes of death' category is:

$$Pr\{A_{1,i} \leq a, \dots, A_{i,n-1} \leq a, A_{D,i} > a, \dots, A_{D,n} > a\} = Pr\{A_{D,n} > a\} \cdot \prod_{i=1}^{n-1} Pr\{A_{1,i} \leq a, A_{D,i} > a\} \quad (A)$$

with:

a : index for age

$A_{1,i}$: age at incidence of disease i

$A_{D,i}$: age at death from disease i

$A_{D,n}$: age at death from all other causes

For specific numbers of diseases slightly weaker conditions can be derived from (A) by integration over incidences or causes of death, and without loss of generality we will restrict ourselves to a situation with 2 diseases and an additional 'all other causes' mortality. We will show that in this case the three independence assumptions below are sufficient for proposition I-III to hold.

Incidences and mortalities are independent (except for the same disease):

$$Pr\left\{A_{1i} \leq a \mid \bigcap_{j \in S} \{A_{Dj} > a\}\right\} = Pr\{A_{1i} \leq a \mid A_{Di} > a\} \quad \forall i \in Z$$

$$Pr\left\{A_{Dj} > a \mid \bigcap_{i \in Z} \{A_{1i} \leq a\}\right\} = \begin{cases} Pr\{A_{Dj} > a \mid A_{1j} \leq a\} & \forall j \in Z \\ Pr\{A_{Dj} > a\} & j \in S, j \notin Z \end{cases} \quad (A1)$$

Incidences are mutually independent:

$$Pr\left\{\bigcap_{i \in Z} \{A_{iI} \leq a\}\right\} = \prod_{i \in Z} Pr\{A_{iI} \leq a\} \quad (\text{A2})$$

Causes of death are mutually independent:

$$Pr\left\{\bigcap_{j \in S} \{A_{Dj} > a\}\right\} = \prod_{j \in S} Pr\{A_{Dj} > a\} \quad (\text{A3})$$

with:

Z: the set of diseases and S: the set of causes of death.

Now let:

$Z = \{L, H\}$.

$S = \{L, H, O\}$

$N_L(a)$: population prevalence of disease L at age a .

$N_H(a)$: population prevalence of disease H at age a .

$N_{HL}(a)$: population prevalence of comorbidity of diseases H and L at age a .

$n_H(a)$: prevalence of disease H at age a in disease section.

$M_H(a)$: population disease specific mortality of disease H at age a .

$m_H(a)$: disease specific mortality of disease H at age a in disease section.

Proposition I

The disease section prevalence and the population disease prevalence can be expressed respectively as:

$$n_H(a) = Pr\{A_{IH} \leq a \mid A_{DH} > a\} \quad (1)$$

$$N_H(a) = Pr\{A_{IH} \leq a \mid A_{DL} > a, A_{DH} > a, A_{DO} > a\} \quad (2)$$

From equation (2) and A1 it follows that:

$$N_H(a) = Pr\{A_{IH} \leq a \mid A_{DH} > a\} = n_H(a) \quad (3)$$

Proposition II

The population prevalence of disease L and of the comorbidity of H and L are:

$$N_L(a) = Pr\{A_{IL} \leq a \mid A_{DL} > a, A_{DH} > a, A_{DO} > a\} \quad (4)$$

$$N_{HL}(a) = Pr\{A_{IH} \leq a, A_{IL} \leq a \mid A_{DL} > a, A_{DH} > a, A_{DO} > a\} \quad (5)$$

By proposition I we know that $N_H(a)=n_H(a)$, and similarly for disease L. Proposition II now becomes: are A1-A3 sufficient conditions for (6) to be true:

$$n_H(a)n_L(a) = N_{HL}(a) \quad (6)$$

Rewriting (5) gives:

$$N_{HL}(a) = \frac{\Pr\{A_{IH} \leq a, A_{IL} \leq a, A_{DH} > a, A_{DL} > a, A_{DO} > a\}}{\Pr\{A_{DH} > a, A_{DL} > a, A_{DO} > a\}} \quad (7)$$

The denominator of (7) can be partitioned by using (A3), to partition the numerator we use, in addition to (A1) and (A2), the property that incidence of a disease by definition has to occur before its mortality. This property allows us to write:

$$\Pr\{A_{IL} \leq a, A_{DL} > a\} = \Pr\{A_{IL} \leq a\} - \Pr\{A_{DL} \leq a\} \quad (8)$$

Equation (7) then becomes:

$$N_{HL}(a) = \frac{\left[\Pr\{A_{IH} \leq a, A_{IL} \leq a, A_{DH} > a\} - \Pr\{A_{IH} \leq a, A_{DL} \leq a, A_{DO} > a\} + \Pr\{A_{DH} \leq a, A_{DL} \leq a, A_{DO} > a\} - \Pr\{A_{DH} \leq a, A_{IL} \leq a, A_{DO} > a\} \right]}{\Pr\{A_{DH} > a\}\Pr\{A_{DL} > a\}\Pr\{A_{DO} > a\}} \quad (9)$$

$$= \frac{\Pr\{A_{DO} > a\}\Pr\{A_{IH} \leq a, A_{DH} > a\}\Pr\{A_{IL} \leq a, A_{DL} > a\}}{\Pr\{A_{DL} > a\}\Pr\{A_{DH} > a\}\Pr\{A_{DO} > a\}} \quad (10)$$

$$= \Pr\{A_{IH} \leq a | A_{DH} > a\}\Pr\{A_{IL} \leq a | A_{DL} > a\} = n_H(a)n_L(a) \quad (11)$$

Proposition III

The population and disease section probabilities are respectively:

$$M_H(a) = \Pr\{A_{DH} \leq a | A_{DL} > a, A_{DO} > a\} \quad (12)$$

$$m_H(a) = \Pr\{A_{DH} \leq a\} \quad (13)$$

From A3 it follows that:

$$M_H(a) = \Pr\{A_{DH} \leq a\} = m_H(a) \quad (14)$$

An example

A simple disease model

We will illustrate the method with two diseases and an 'all other causes' mortality. Each disease is described by an illness-death process with only one diseased state, consisting of just disease specific incidence, prevalence and mortality, with age specific incidence and mortality rates to be given. Recovery is ignored. This disease model can be described by a continuous time Markov process with three states: h : non-diseased; z : diseased; and d : dead. There are two transition hazards: γ from healthy to diseased; and φ from diseased to dead.

The equations below assume the transition hazards to be constant within age intervals, and to minimize the impact of this assumption the equations use 1-year age intervals. We consider the age interval $[a, a+1]$, with persons distributed over the three health states, and define r_a to be the probability for a person in the healthy state h at exact age a to make two transitions within the age interval: from healthy to diseased and from diseased to dead. Let U be the time spent in that age interval in the healthy state, and V the time spent in the interval in the diseased state. Two transitions occur in the 1-year interval when the sum of these times is smaller than 1:

$$\begin{aligned}
 r_a &= P\{U+V \leq 1\} \\
 &= \int_0^1 P\{V \leq 1-u | U = u\} P\{U = u\} du \\
 &= \int_0^1 (1 - e^{-\varphi_a(1-u)}) \gamma_a e^{-\gamma_a u} du \\
 &= \begin{cases} \frac{\varphi_a(1 - e^{-\gamma_a}) - \gamma_a(1 - e^{-\varphi_a})}{\varphi_a - \gamma_a}, & \gamma_a \neq \varphi_a \\ 1 - e^{-\gamma_a} - \gamma_a e^{-\gamma_a}, & \gamma_a = \varphi_a \end{cases} \quad (15)
 \end{aligned}$$

The state variables h_a , z_a , and d_a denote the proportion of the population in the various states at age a . The states at $a+1$ are described by:

$$h_{a+1} = h_a e^{-\gamma_a} \quad (16)$$

$$z_{a+1} = h_a(1 - e^{-\gamma_a} - r_a) + z_a e^{-\varphi_a} \quad (17)$$

$$d_{a+1} = h_a r_a + z_a(1 - e^{-\varphi_a}) + d_a \quad (18)$$

Prevalence at exact age a is:

$$n_a = \frac{z_a}{1-d_a} \quad (19)$$

Using eq. 16-19 we can express prevalence at age $a+1$ as:

$$n_{a+1} = \frac{n_a e^{-\varphi_a} + (1-n_a)(1-e^{-\gamma_a} - r_a)}{1-n_a(1-e^{-\varphi_a}) - (1-n_a)r_a} \quad (20)$$

Mortality probability at a is given by:

$$m_a = \frac{d_{a+1} - d_a}{1-d_a} \quad (21)$$

and incidence density at a by:

$$i_a = 1 - e^{-\gamma_a} \quad (22)$$

We can now write prevalence at $a+1$ as a function of prevalence, incidence and mortality at a :

$$n_{a+1} = \frac{n_a - m_a + (1-n_a)i_a}{1-m_a} \quad (23)$$

Life table setup

In table 1 we show the setup of the general section of the life table. The first column contains the 'all other causes' mortality rate, the second adds

Table 1
The general section of the proportional multi-state life table.

δ_a^o	δ_a	q_a	l_a	d_a	L_a	T_a	e_a
δ_0^o	$\delta_0^o + \sum_d \delta_0^d$	$1 - e^{-\delta_0}$	100000	$l_0 q_0$	$0.5(l_0 + l_1)$	$\sum_{a=0}^{95} L_a$	$\frac{T_0}{l_0}$
δ_1^o	$\delta_1^o + \sum_d \delta_1^d$	$1 - e^{-\delta_1}$	$l_0 - d_0$	$l_1 q_1$	$0.5(l_1 + l_2)$	$\sum_{a=1}^{95} L_a$	$\frac{T_1}{l_1}$

δ^o 'all other causes' mortality rate

δ^d disease specific mortality rate

δ total mortality rate

Other notation follows standard life table convention.

to this the disease specific mortality rates from the disease sections, and the third converts this to the total mortality probability. From there on the general section is equal to a standard life table.

Table 2 shows the setup of a disease section. The first two columns give the input, the incidence and mortality rates respectively, the latter one is used in the general section to determine total mortality probability. The third and fourth column convert these to probabilities, and the fifth calculates the prevalence.

Implementation

We have implemented this method, employing a spreadsheet, for heart disease (ischemic heart disease and congestive heart failure combined) and stroke among Dutch males in 1988. Table 3 shows incidence and mortality rates used in the calculations. Incidence of heart disease is based on various studies and a nationwide hospital register (Bonneux, Barendregt et al. 1994), incidence of stroke on a Dutch regional and several international studies (Niessen, Barendregt et al. 1993), disease specific and total mortality rates are from the national bureau of statistics (Statistics Netherlands 1991; Statistics Netherlands Published annually).

Calculations are done in a full life table (i.e. with 1-year age intervals) because of the crucial assumption, made to be able to subtract and sum mortality rates, that all hazard rates are constant during the age interval (Manton and Stallard 1988). Table 4 shows part of such a full life table with a general section and a heart disease section. In the latter nothing is happening until, at age 23, the incidence hazard γ_a^d becomes greater than 0. When the heart disease mortality rate δ_a^d is big enough it becomes clear that the total mortality hazard δ_a is the sum of mortality hazard from all other causes (δ_a^o) and the disease specific rate. The disease prevalence in numbers (N_a^d) is obtained by multiplication of the prevalence rate n_a^d with

Table 2
A disease specific section of the proportional multi-state life table

γ_a	δ_a	i_a	m_a	n_a
γ_0	δ_0	$1 - e^{-\gamma_0}$	$1 - e^{-\delta_0}$	0
γ_1	δ_1	$1 - e^{-\gamma_1}$	$1 - e^{-\delta_1}$	$\frac{n_0 - m_0 + (1 - n_0)i_0}{1 - m_0}$

- γ incidence rate
 δ disease specific mortality rate
 i incidence probability
 m disease specific mortality probability
 n disease prevalence

the l_a column of the general section: both n_a^d and N_a^d are prevalences at exact age a . Note that at these younger ages the difference between hazards and probabilities does not show up in the rounded figures. Also note that by using cross-sectional data in a cohort model the estimated prevalences will not fully comply with observed prevalences in the presence of past trends.

With this life table we first estimated the baseline age specific case fatalities (the disease specific mortality probability, given disease prevalence) from observed disease specific mortality and estimated prevalence, using simply:

$$p_a = \frac{m_a}{n_a} \quad (24)$$

The resulting base line values are then used to estimate the impact on disease prevalence and on disease specific and total mortality of changes in incidence and case fatalities.

Table 3 - Disease specific incidence and mortality rates for heart disease and stroke, Netherlands, 1989, males.

Incidence is defined as incidence rate among the non-prevalent population, heart disease is ischemic heart disease and congestive heart failure combined.

Age	Heart disease		Stroke	
	incidence	mortality	incidence	mortality
25-29	0.000115	0.000012	0.000071	0.000009
30-34	0.000316	0.000048	0.000127	0.000021
35-39	0.001190	0.000175	0.000225	0.000047
40-44	0.003068	0.000462	0.000388	0.000100
45-49	0.005730	0.000943	0.000670	0.000175
50-54	0.009275	0.001710	0.001241	0.000389
55-59	0.013079	0.003478	0.002319	0.000771
60-64	0.016666	0.006122	0.003598	0.001431
65-69	0.021097	0.010198	0.005607	0.002783
70-74	0.028204	0.016316	0.008775	0.005379
75-79	0.038341	0.024833	0.013772	0.009980
80-84	0.051322	0.035256	0.021621	0.018235
85-89	0.068518	0.046781	0.025794	0.025746
90-94	0.098469	0.057755	0.029295	0.033099
95+	0.147829	0.065907	0.031636	0.036967

Cardiovascular disease in the Netherlands (as in many other western countries) has seen declining age specific mortality, for stroke at least since the early sixties, for heart disease since the early seventies. A closer look at the available evidence reveals that this decline is most likely due to a combination of declining incidence and increasing survival (Niessen, Barendregt et al. 1993; Bonneux, Barendregt et al. 1994). We assess the impact of such combined trends by multiplying age specific incidence of both diseases with a factor of 0.9 (all ages), and multiply case fatalities (i.e. 1-survival probability) likewise with 0.8, always comparing outcomes with the 1988 base year. These changes in incidence and survival are compatible with the presumed changes in the past decade.

Baseline results

In table 5 the resulting disease prevalences for the 1988 base year are shown, both in rates and numbers. The prevalence rates are calculated using the average of n_a^d and n_{a+1}^d , the number are obtained by multiplication of the average rate with the general sections L_a column, and are therefore higher than the prevalences in table 4.

Improved survival

First we increase survival with heart disease and stroke, while leaving all other parameters at their base value. Figure 1 shows outcome expressed as indices (with 1988=100) of prevalences by age of heart disease and stroke and of the comorbidity of heart disease and stroke. Both prevalence rates show the same pattern: increases for all age groups, but much more so in older age groups. This pattern is much more visible for stroke than for heart disease, because the base survival of stroke is worse and therefore an equal relative improvement has a larger impact. The comorbidity rate of heart disease and stroke, being the product of the two disease specific prevalence rates, shows the same pattern even more strongly: for the highest age group it is 50% higher than the base line.

The effect of improved disease specific survival on total mortality causes a further strengthening of this pattern when we look at prevalence numbers, because a larger part of the initial birth cohort lives to high age. Comorbidity of heart disease and stroke is 65% higher for the oldest age group.

Incidence decline

Secondly we multiplied the age specific incidences with 0.9, again with all other parameters at their base value (figure 2). The age specific prevalences at the lowest ages stand at a corresponding 90% of the base rate, but these percentages edge up a bit with age. The lower prevalence rate in any age group results in a larger group at risk for incidence, which counter-acts the lower incidence rate. A similar result for cardiovascular mortality was observed by Rose and Shipley (1990). The effect is stronger for heart disease

Table 4. Proportional multi-state life table (partly) with a general section and a heart disease section.

N_a^d is the product of l_a and n_a^d , all other equations are in tables 1 and 2.

age	General section					Heart disease section					N_a^d
	δ_a^o	δ_a	q_a	l_a	d_a	γ_a^d	δ_a^d	i_a^d	m_a^d	n_a^d	
0	0.0078	0.0078	0.0078	100000	776	0.00000	0.00000	0.00000	0.00000	0.00000	0
1	0.0011	0.0011	0.0011	99225	108	0.00000	0.00000	0.00000	0.00000	0.00000	0
2	0.0005	0.0005	0.0005	99117	51	0.00000	0.00000	0.00000	0.00000	0.00000	0
3	0.0004	0.0004	0.0004	99066	36	0.00000	0.00000	0.00000	0.00000	0.00000	0
4	0.0003	0.0003	0.0003	99030	28	0.00000	0.00000	0.00000	0.00000	0.00000	0
5	0.0003	0.0003	0.0003	99002	25	0.00000	0.00000	0.00000	0.00000	0.00000	0
6	0.0003	0.0003	0.0003	98977	25	0.00000	0.00000	0.00000	0.00000	0.00000	0
7	0.0002	0.0002	0.0002	98952	22	0.00000	0.00000	0.00000	0.00000	0.00000	0
8	0.0002	0.0002	0.0002	98930	19	0.00000	0.00000	0.00000	0.00000	0.00000	0
9	0.0002	0.0002	0.0002	98911	19	0.00000	0.00000	0.00000	0.00000	0.00000	0
10	0.0002	0.0002	0.0002	98892	19	0.00000	0.00000	0.00000	0.00000	0.00000	0
11	0.0002	0.0002	0.0002	98873	19	0.00000	0.00000	0.00000	0.00000	0.00000	0
12	0.0002	0.0002	0.0002	98854	21	0.00000	0.00000	0.00000	0.00000	0.00000	0
13	0.0002	0.0002	0.0002	98833	23	0.00000	0.00000	0.00000	0.00000	0.00000	0
14	0.0003	0.0003	0.0003	98810	26	0.00000	0.00000	0.00000	0.00000	0.00000	0
15	0.0004	0.0004	0.0004	98784	36	0.00000	0.00000	0.00000	0.00000	0.00000	0
16	0.0005	0.0005	0.0005	98748	49	0.00000	0.00000	0.00000	0.00000	0.00000	0
17	0.0006	0.0006	0.0006	98699	54	0.00000	0.00000	0.00000	0.00000	0.00000	0
18	0.0006	0.0006	0.0006	98645	59	0.00000	0.00000	0.00000	0.00000	0.00000	0

19	0.0007	0.0007	0.0007	98585	67	0.00000	0.00000	0.00000	0.00000	0.00000	0
20	0.0007	0.0007	0.0007	98519	70	0.00000	0.00000	0.00000	0.00000	0.00000	0
21	0.0007	0.0007	0.0007	98449	71	0.00000	0.00000	0.00000	0.00000	0.00000	0
22	0.0007	0.0007	0.0007	98378	72	0.00000	0.00000	0.00000	0.00000	0.00000	0
23	0.0007	0.0007	0.0007	98306	73	0.00002	0.00000	0.00002	0.00000	0.00000	0
24	0.0007	0.0007	0.0007	98233	73	0.00005	0.00000	0.00005	0.00000	0.00002	2
25	0.0007	0.0007	0.0007	98160	71	0.00007	0.00001	0.00007	0.00001	0.00006	6
26	0.0007	0.0007	0.0007	98089	71	0.00009	0.00001	0.00009	0.00001	0.00012	12
27	0.0007	0.0008	0.0008	98018	74	0.00012	0.00001	0.00012	0.00001	0.00021	20
28	0.0007	0.0008	0.0008	97943	73	0.00016	0.00002	0.00016	0.00002	0.00031	30
29	0.0007	0.0008	0.0008	97870	74	0.00020	0.00003	0.00020	0.00003	0.00045	44
30	0.0007	0.0008	0.0008	97796	75	0.00024	0.00003	0.00024	0.00003	0.00061	60
31	0.0008	0.0008	0.0008	97721	79	0.00028	0.00004	0.00028	0.00004	0.00082	80
32	0.0008	0.0009	0.0009	97642	86	0.00032	0.00005	0.00032	0.00005	0.00105	103
33	0.0009	0.0009	0.0009	97556	90	0.00049	0.00007	0.00049	0.00007	0.00132	129
34	0.0009	0.0010	0.0010	97465	94	0.00067	0.00010	0.00067	0.00010	0.00174	169
35	0.0009	0.0011	0.0011	97371	102	0.00084	0.00012	0.00084	0.00012	0.00230	224
36	0.0010	0.0011	0.0011	97269	111	0.00102	0.00015	0.00101	0.00015	0.00302	293
37	0.0011	0.0012	0.0012	97158	121	0.00119	0.00018	0.00119	0.00017	0.00388	377
38	0.0011	0.0013	0.0013	97037	126	0.00157	0.00023	0.00156	0.00023	0.00489	475
39	0.0011	0.0014	0.0014	96911	134	0.00194	0.00029	0.00194	0.00029	0.00622	602
40	0.0012	0.0016	0.0016	96777	151	0.00232	0.00035	0.00231	0.00035	0.00786	760

because the base line prevalence of heart disease is higher than the one of stroke.

With both diseases at 90% of their prevalence rates for the young, the comorbidity rate of course stands at 81% for those ages, and is also edging up with age. Through the implied lower disease specific mortalities this pattern is more pronounced in prevalence numbers: at the highest age disease prevalences even exceed the base line value.

Improved survival and lower incidence combined

The combined picture of improved survival and lower incidence is in figure 3. Lower incidence is reflected in 5 to 10% lower prevalence rates at young

Table 6 - Base line values for the prevalence of heart disease, stroke, and their comorbidity.

Age	prevalence rates			prevalence numbers		
	heart	stroke	comob	heart	stroke	comob
0-4	0.0000	0.0000	0.0000	0	0	0
5-9	0.0000	0.0000	0.0000	0	0	0
10-14	0.0000	0.0000	0.0000	0	0	0
15-19	0.0000	0.0000	0.0000	0	0	0
20-24	0.0000	0.0000	0.0000	5	3	0
25-29	0.0003	0.0002	0.0000	140	82	0
30-34	0.0013	0.0006	0.0000	624	287	0
35-39	0.0046	0.0013	0.0000	2241	633	3
40-44	0.0137	0.0025	0.0000	6606	1195	17
45-49	0.0318	0.0045	0.0001	15144	2118	70
50-54	0.0612	0.0078	0.0005	28455	3648	229
55-59	0.1003	0.0138	0.0014	44828	6159	629
60-64	0.1441	0.0228	0.0033	60136	9533	1390
65-69	0.1884	0.0349	0.0066	70041	12957	2459
70-74	0.2327	0.0495	0.0115	71426	15201	3555
75-79	0.2785	0.0660	0.0184	62772	14870	4156
80-84	0.3274	0.0816	0.0267	45286	11291	3705
85-89	0.3838	0.0888	0.0341	25242	5841	2241
90-94	0.4640	0.0801	0.0372	10202	1762	814
95+	0.5403	0.0671	0.0363	2819	350	189
Total				445968	85932	19458

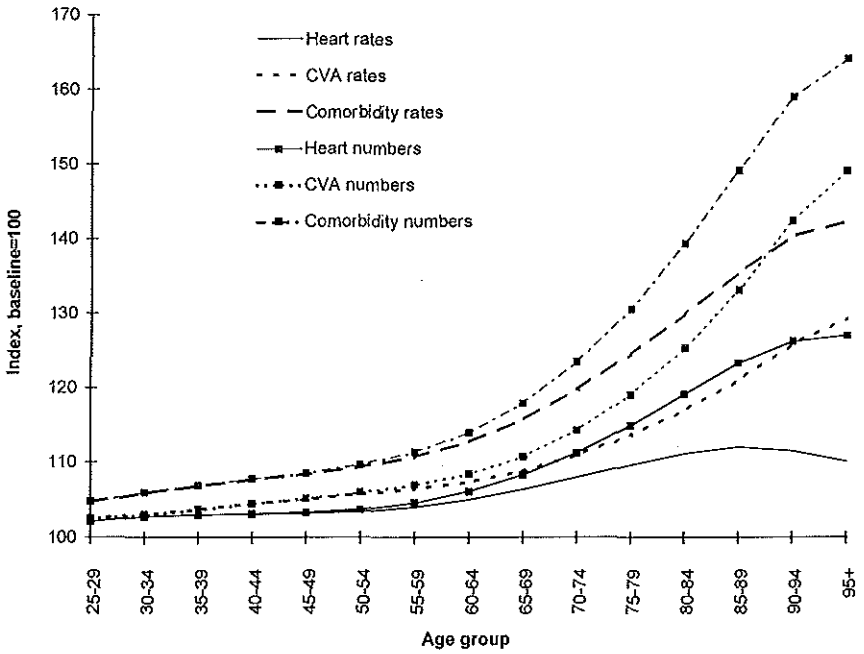


Figure 1 - Prevalences of heart disease and stroke (rates and numbers) by age after improved survival, index with baseline = 100. Legends: heart rates: prevalence rate of heart disease; CVA rates: prevalence rate of stroke; comorbidity rates: comorbidity prevalence rate of heart disease and stroke; heart numbers: prevalence numbers of heart disease; CVA numbers: prevalence numbers of stroke; comob numbers: comorbidity prevalence numbers of heart disease and stroke

ages, and the increased survival causes up to about 15% higher prevalence rates at older ages. The now combined mortality effect pushes up disease prevalence numbers at the highest ages above the improved survival-only case. And comorbidity, while being more than 15% lower for the young, shows a 20% (rates) and 55% (numbers) increase for the oldest age group. The result are tilted index lines, with pivot points somewhere between age 65 and 70.

Summed across ages the lower prevalence at young and higher prevalence at older ages on balance result in somewhat higher prevalence numbers: 3 and 7% for heart disease and stroke respectively. Number of years lived with disease as a percentage of total number of years lived also increases, but a bit less: 2 and 5% respectively.

Discussion

Improved survival and decreased incidence have, as has been demonstrated above, a similar effect on disease prevalence: a relative shift of the burden of disease towards older ages. With improved survival all age groups have

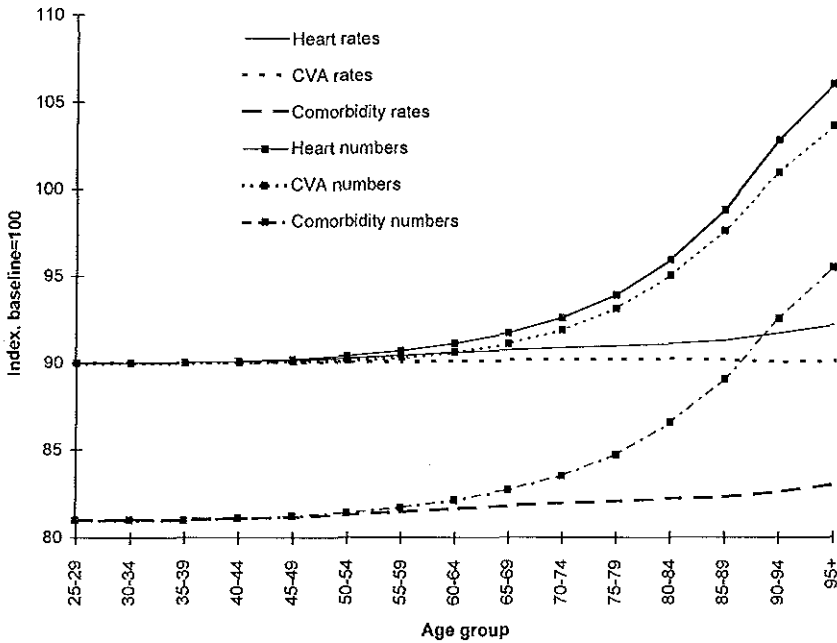


Figure 2 - Prevalences of heart disease and stroke (rates and numbers) by age after incidence decline, index with baseline = 100. Legends: heart rates: prevalence rate of heart disease; CVA rates: prevalence rate of stroke; comorbidity rates: comorbidity prevalence rate of heart disease and stroke; heart numbers: prevalence numbers of heart disease; CVA numbers: prevalence numbers of stroke; comorbidity numbers: comorbidity prevalence numbers of heart disease and stroke

higher prevalences, but the oldest much more so. Declining incidence lowers prevalence for most age groups, but older ages benefit less and the oldest even have a higher prevalence. Combining both changes turns the relative shift into an absolute one: younger age groups experience less disease, older age groups more. In terms of disability this pattern will be even more pronounced because of comorbidity: disability in many cases might be an almost exponential function of the number of comorbid conditions (Verbrugge, Lepkowski et al. 1989).

The decline in cardiovascular disease mortality is most likely due to a combination of declining incidence and increasing survival. Declining incidence may be unique to cardiovascular disease, cancer incidence, for example, has been reported up, albeit based on inconclusive evidence (Adami, Bergstrom et al. 1993; Bonneux, Van Oortmarsen et al. 1993; Bonneux, Barendregt et al. 1995). But increased survival probably extends to many other chronic conditions: medical technology has seen more progress in management of chronic disease than in cure. Barring major breakthroughs this pattern is likely to continue.

The question whether morbidity has been compressing or expanding must be seen against the epidemiological background of improved survival

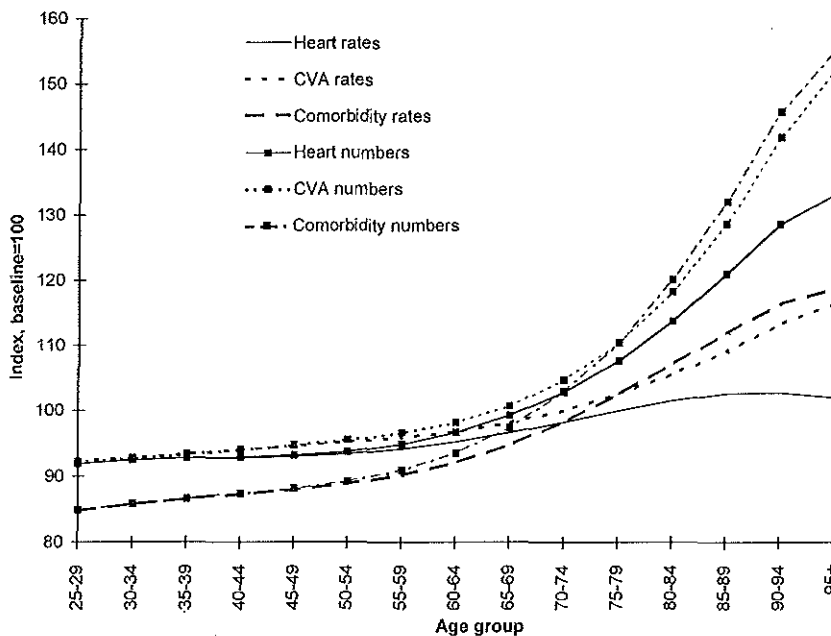


Figure 3 - Prevalences of heart disease and stroke (rates and numbers) by age after improved survival and incidence decline combined, index with baseline = 100. Legends: heart rates: prevalence rate of heart disease; CVA rates: prevalence rate of stroke; comorbidity rates: comorbidity prevalence rate of heart disease and stroke; heart numbers: prevalence numbers of heart disease; CVA numbers: prevalence numbers of stroke; comorbidity numbers: comorbidity prevalence numbers of heart disease and stroke

for most chronic diseases, with simultaneously decreasing incidences for the large cardiovascular causes of morbidity. Although it is difficult to say what the overall effect of these changes has been, it is likely, in particular when expressed in terms of disability, that there has indeed been an absolute shift in the burden of disease from the young towards the old.

If this pattern of a decreasing prevalence at younger and an increasing one at older ages will indeed turn out to be the prevailing development, it would seem that the debate on compression versus expansion of morbidity is being held in too simplistic terms. While morbidity is compressing in the sense that there is less of it at younger age and it is being concentrated at higher ages, it is at the same time expanding in the sense that there is more of it at older age and in total (both absolute and relative). Perhaps we should rethink our terminology.

The shift in age distribution of disease described above has economic consequences as well: it will inflate health care costs because of increased numbers of patients, who are older on average, have more comorbidity, and therefore are more expensive.

Similar exercises have been done by Crimmins, Hayward et al. (1994). They use a standard multi-state life table with states of varying degrees of dependency, and look at the effects of changes in mortality and morbidity

rates. Their results seem compatible with ours, but unfortunately they do not report on changes in the age pattern of dependency, only on expected dependent years and the proportion of total life expectancy in a dependent state.

The proportional multi-state life table method employed in this paper provides a relatively simple method to incorporate several diseases in a life table, and keep track of comorbidity. To be sure, the standard multi-state life table method can handle multiple diseases with comorbidity, but the resulting life table soon becomes awkward when the number of diseases increases. In particular, adding a disease to an existing standard multi-state life table requires a major update of the life table.

With a proportional multi-state life table, in contrast, only the first two columns of the general section need to be updated, and the four columns of a new disease section added. This aspect makes a step-wise expansion of the life table to incorporate an increasing number of diseases an attractive option.

The price for this relative simplicity and flexibility is assumptions A1-A3 (or a similar set, depending on the number of diseases). They require the diseases to be independent, and the life table cohort to be homogeneous. When other than independent diseases are desirable these can only be modeled by defining additional states. For example, when diabetes and cardiovascular disease are modeled it is imperative to let diabetics have a higher risk on cardiovascular disease incidence, and perhaps also a worse survival. This can be achieved by doubling the number of states of the life table, one set for diabetics and one for non-diabetics. The independence conditions are thus relaxed to local independence. A similar setup can be done for risk factors like smoking and hypertension, but obviously there are practical limits to the number of dependent diseases and risk factors that can thus be accommodated.

Our implementation of the life table used 1-year age intervals. As such this not a requirement, eq 15-23 can easily be rewritten for wider intervals, although in particular at higher ages wider intervals will bias results. Minimizing this bias is a good reason for using 1-year intervals, but an additional reason is a wider application of the methodology that avoids a limitation the proportional multi-state life table shares with all life tables: it either describes a single cohort through time, or a population at a single point in time (Shryock and Siegel 1976).

Often in Public Health it is interesting to look at intervention effects on an entire population through time. *Prevent* is a dynamic population model that links risk factor prevalences with disease specific and total mortality, thus allowing estimates of decreased mortality after risk factor intervention (Gunning-Schepers 1989; Gunning-Schepers, Barendregt et al. 1989). This model has now been extended to *Prevent Plus* that incorporates multiple morbidity, using dynamic versions of the disease section of the proportional multi-state life table. *Prevent Plus* has a 1-year time step, and to avoid unwanted blending of disease prevalences between adjacent

age groups the age interval and time step of the dynamic disease models must be equal.

The proportional multi-state life table method also forms the methodological basis of NIMPH, the Netherlands Integrated Model of Public Health. NIMPH combines the dynamic population and risk factor approach of *Prevent Plus* with dependent diseases (like diabetes and cardiovascular disease) and multi-stage disease models (to allow for duration dependence and differences in severity). This family of models, proportional multi-state life table, *Prevent*, *Prevent Plus*, and NIMPH, with various levels of complexity and data requirements, is used to explore the research field of Public Health, with research question and available data determining which model is most appropriate.

References

- Adami, H. O., Bergstrom, R., Soren, P., & Baron, J. (1993). Increasing cancer risk in younger birth cohorts in Sweden. *Lancet*, *341*, 773-777.
- Bonneux, L., Barendregt, J. J., Looman, C. W. N., & van der Maas, P. J., (1995). Diverging trends in colorectal cancer morbidity and mortality: earlier diagnosis comes at a price. *Eur J Cancer*, *31A*(10), 1665-71.
- Bonneux, L., Barendregt, J. J., Meeter, K., Bonsel, G. J., & van der Maas, P. J. (1994). Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J of Public Health*, *84*, 20-28.
- Bonneux, L., Van Oortmarsen, G. J., & Barendregt, J. J. (1993). Increase in cancer incidence in younger birth cohorts (letter). *Lancet*, *341*, 1409.
- Crimmins, E. M., Hayward, M. D., & Saito, Y. (1994). Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography*, *31*(1), 159-75.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *New England Journal of Medicine*, *303*, 130-135.
- Gunning-Schepers, L. J. (1989). The health benefits of prevention, a simulation approach. *Health Policy*, *12*, 1-256.
- Gunning-Schepers, L. J., Barendregt, J. J., & Van der Maas, P. J. (1989). Population interventions reassessed. *Lancet*(I), 479-481.
- Manton, K. G., & Stallard, E. (1988). *Chronic Disease Modelling* (Vol. 2). London: Charles Griffin & Company Ltd.
- Murray, C. J. L., & Lopez, A. D. (Ed.). (1994). *Global comparative assessments in the health sector*. Geneva: World Health Organization.
- Niessen, L. W., Barendregt, J. J., Bonneux, L., & Koudstaal, P. J. (1993). Stroke Trends in an Aging Population. *Stroke*, *24*, 931-939.
- Olshansky, S. J., Rudberg, M. A., Carnes, B. A., Cassel, C. K., & Brody, J. A. (1991). Trading off longer life for worsening health: the expansion of morbidity hypothesis. *Journal of Aging and Health*, *3*(2), 194-216.
- Omran, A. R. (1971). The epidemiological transition: a theory of the epidemiology of population change. *Milbank Memorial Fund quarterly*, *49*, 509-538.

- Rose, G., & Shipley, M. (1990). Effects of coronary risk reduction on the pattern of mortality [see comments]. *Lancet*, 335(8684), 275-7.
- Schoen, R. (1987). *Modelling Multigroup Populations*. New York and London: Plenum Press.
- Shryock, H. S., & Siegel, J. S. (1976). *The methods and materials of demography, condensed edition*. Orlando: Academic Press, Inc.
- Statistics Netherlands (1991). Life tables for men by age, 1986-1990. *Maandstatistiek bevolking*(12), 76.
- Statistics Netherlands (Published annually). Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1. Voorburg: CBS.
- Verbrugge, L. M., Lepkowski, J. M., & Imanaka, Y. (1989). Comorbidity and its impact on disability. *Milbank Q*, 67(3-4), 450-84.
- World Bank (1993). *Investing in Health*. New York: Oxford Univ Press.

Consistent estimates of NIDDM incidence, prevalence and mortality

Abstract

Objective: To estimate NIDDM incidence in the Netherlands in the absence of a sufficiently large empirical study.

Research design and methods: An incidence/prevalence/mortality (IPM) model of NIDDM forms the basis. Prevalence data were obtained from a study which pooled existing prevalence estimates. Diabetes-related mortality was estimated using relative risks on all cause mortality, the NIDDM prevalence estimate, and Dutch all cause mortality. Diabetes-related mortality and NIDDM prevalence allow to estimate incidence using the IPM model.

Results: Annual NIDDM incidence estimates by age and sex, that are consistent with the prevalence estimate and the diabetes-related mortality. Estimates range from 8.1 per 10,000 (7.7..8.8) for males aged 40-44, and 7.0 (6.8..8.0) for females, to 79.7 per 10,000 (69.5..90.9) for males aged 75-79, and 85.8 (80.6..91.0) for females. This compares well with the few observations available.

Conclusions: When empirical estimates of incidence are largely lacking the methodology described offers a useful alternative. The internal consistency with the prevalence and mortality estimates allows to use the set for the assessment of potential intervention effects.

Introduction

Describing the epidemiology of NIDDM is surprisingly difficult. An important reason for this is the difficulty to separate cases from non-cases. Unlike IDDM the onset of NIDDM is a gradual process, where at some point in time a clinically defined threshold is reached and the patient becomes a case. Initially the patient experiences few, if any, symptoms, and therefore a large part of all NIDDM patients is undiagnosed: some estimates are as high as 50% (1).

In such a situation (difficult case definition and a high prevalence of occult disease) the observed incidence and prevalence become highly dependent on factors other than the true occurrence of disease. Factors include patient and doctor awareness, level of case finding, campaigns by patient organizations or governments, and changes in insurance policies. This results in highly variable estimates of incidence and prevalence. Studies of prevalence and incidence in the Netherlands show a variation in estimates by a factor of about 3 (2-6).

Looking at disease specific mortality, for most diseases the best recorded epidemiological aspect, is in the case of diabetes of not much help either. Diabetes is a risk factor for a number of diseases, such as cardiovascular disease. When diabetes is an underlying cause for cardiovascular death it is supposed to be mentioned as such on the death certificate. This is not always the case since it is difficult for physicians to establish the role of diabetes in the death process, especially at older ages when co-morbidity is more frequent. Coding of the death certificates also promotes difficulties in interpretation. It is suspected that the cause of death registration underreports the number of deaths due to diabetes to a varying degree over time, with large swings between subsequent ICD versions. The usefulness of the cause of death registration for the study of diabetes epidemiology is therefore limited, in particular for the elderly (7, 8).

The uncertainty about diabetes occurrence hampers the assessment of the burden of disease, and of the potential gains in population health status that could be made by a better control of risk factors for diabetes, and of diabetes itself. Assessing such gains requires a consistent set of incidence, prevalence, and mortality rates, which can be used as input in an epidemiological model. The impact of changes in incidence, for example by weight control, on diabetes prevalence and diabetes related mortality can be estimated with such a model.

The purpose of the present paper is to obtain a consistent set of diabetes incidence, prevalence and mortality. In a previous study we presented an estimate of diabetes prevalence in the Netherlands, based on a pooled analysis of existing studies (2). By combining the increased mortality risk of diabetes with the pooled prevalence estimate in an incidence-prevalence-mortality model, an age and sex specific estimate of diabetes incidence is obtained. The estimates are compared with the few observations available.

Data and methods

The model

Incidence-prevalence-mortality (IPM) models of disease processes are useful tools in epidemiological research. Recent applications concern supporting estimation of incidence, prevalence and survival in the face of unreliable or incomplete data (9), an analysis of trends in colorectal cancer incidence and mortality (10), and an analysis of trends in stroke incidence and survival (11). The main advantage of IPM models is that incidence, prevalence and mortality figures are linked through the causal chain of a disease process, and this chain limits the possible combinations of incidence, prevalence and mortality rates. Limits are imposed because any prevalent case must have become incident at some younger age, and any person dead with a disease must have become incident previously and have been prevalent, however shortly. Jointly estimated incidence, prevalence and mortality rates, using a causal model, are therefore internally consistent (9).

The most convenient way to model a disease process is to assume independence from all other causes of death. Under that assumption the mortality rate of the disease can be obtained from the disease specific mortality as reported by the national statistics bureau, and an 'all other causes' mortality rate can easily be calculated (12).

An independence assumption, though convenient, is not appropriate in the case of diabetes mellitus: in addition to being coded as a cause of death itself (ICD-9 250), diabetes also acts as a risk factor for a range of causes of death, such as cardiovascular disease.

The high mortality of diabetics from other diseases results in an excess total mortality: several studies have reported relative risks on all-cause mortality ranging from about 2-3 (men) and 3-4 (women) around age 40 to a little over 1 in ages around 80 (both sexes) (13-17). This excess mortality gives rise to mortality selection: because diabetics are selectively being removed from the population, the prevalence of diabetes is lower than it would have been if diabetics ran no excess mortality risk. However, part of the diabetes excess mortality is not expressed in ICD 250 mortality rates. Under these circumstances a diabetes IPM model that assumes independence of diabetes and all other mortality risk would, for a given incidence and ICD 250 mortality, result in too high prevalence estimates.

Instead of using the ICD 250 mortality we therefore constructed a diabetes-related mortality (m^d), independent from the cause-of-death statistics. This diabetes-related mortality equals all excess mortality caused by diabetes, however coded in death statistics. Total mortality for diabetics and non-diabetics is estimated from the average population total mortality, the relative risk on mortality for diabetics, and diabetes prevalence using:

$$m = p m^d + (1 - p) m^0 \quad (1)$$

and:

$$m^l = m^0 R \quad (2)$$

Substituting eq (2) in (1) we can write the mortality of the non-diabetics as:

$$m^0 = \frac{m}{pR + 1 - p} \quad (3)$$

The mortality due to diabetes from all causes in the population m^d then becomes:

$$m^d = pm^0(R - 1) = \frac{p(R - 1)m}{pR + 1 - p} \quad (4)$$

Where:

m : average total mortality rate;

m^0 : total mortality rate of non-diabetics, m^l : total mortality rate of diabetics,

m^d : diabetes related mortality rate of population;

p : prevalence of diabetes;

R : relative risk on total mortality, given exposure to diabetes.

Age and sex indexes are suppressed.

Given this mortality rate in the population and the prevalence of diabetes we can now express incidence in the age interval $a..a+1$ (i_a) as a function of prevalence at a and $a+1$, and mortality at a :

$$i_a = \frac{p_{a+1}(1 - m_a^d) - p_a + m_a^d}{1 - p_a} \quad (5)$$

Equation 5 is based on the description of the disease process as a continuous time markov process, and is derived elsewhere (18). A moving average is applied to the resulting age-specific incidences to smooth the discontinuities that are caused by the age group subdivision of the relative risks used (see table 1).

Table 1 - Relative risk (95% confidence interval) on all cause mortality for diabetics from the Verona study (16).

Age	Men	Women
0-44	4.69 (1.71-10.2)	3.03 (0.03-16.86)
45-54	2.33 (1.38-3.69)	3.43 (1.43-6.77)
55-64	2.13 (1.76-2.56)	2.33 (1.63-3.22)
65-74	1.50 (1.30-1.72)	2.27 (1.92-2.66)
75+	1.13 (1.00-1.28)	1.32 (1.20-1.44)

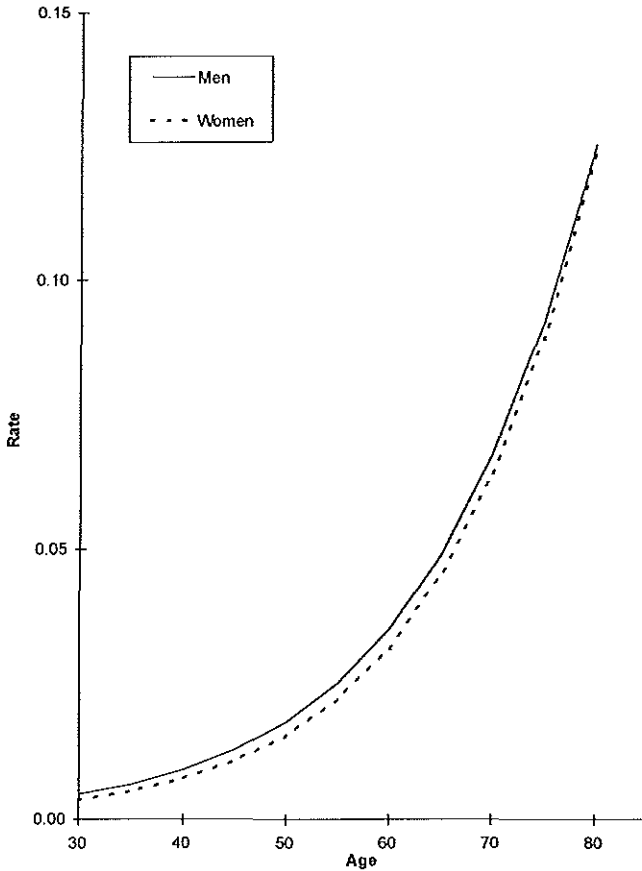


Figure 1 - Prevalence of Diabetes by age and sex, pooled estimates of 15 studies (2).

Data

The pooled estimate of prevalence was based on 15 studies, performed since 1970, which estimated the clinical prevalence in the Netherlands (2). The studies largely fall into two groups: one based on registrations by general practitioners, and one based on self report in a surveys. For each group we fitted an exponential curve on age-specific prevalence using logistic regression. The fit based on the survey group of studies yielded a somewhat higher pooled prevalence than the one on the GP studies, due to one large GP study with rather low reported prevalence. The fitted curve of the survey studies forms the basis of the current analysis. Because of data limitations the pooled prevalence estimate is restricted to the age range of 30-80, and therefore the prevalence estimate concerns mainly NIDDM patients. Figure 1 shows the prevalence rates by age of men and women respectively. More details on the

pooled estimate have been published elsewhere (2). The relative risks for the calculations in this analysis are based on from the Verona study (table 1) (16). This decision was based on an earlier study, in which we have compared several studies reporting relative risks on all cause mortality for diabetes (19).

We are aware of four studies in the Netherlands that have reported incidence of diabetes mellitus (Table 2), these will be compared with our estimation of diabetes incidence.

Results

Figures 2 (men) and 3 (women) present an incidence estimate by age and a lower and upper bound, based on the estimates of relative risks and their 95% confidence intervals from the Verona study. The incidence estimates are very similar for men and women, increasing over the whole age range, and reaching a level of about 1% per year at age 80 (see also table 3). The lower bound relative risks result in an incidence estimate below the central estimate, the higher bound in one above: with lower mortality selection a lower incidence will be needed to reproduce the prevalence and vice versa. The difference between the lower and upper estimate is bigger for men than for women. This is due to a relatively wider confidence interval of the relative risk (see also Table 1) and a higher absolute mortality risk.

In the figures 2 and 3 also the observations from four Dutch incidence studies (table 2) are presented. The data points are plotted in the middle of the (wide) age intervals, for the open ended highest age groups we used the average age of these groups as calculated with a life table. The four incidence studies give rather similar results for younger ages, but show an increasingly wide divergence with age, with the Continuous Morbidity Registration (CMR)-Nijmegen estimate tending to be the highest, and the National Study the lowest. Comparing our results with the four studies, our estimates seem to be more or less in the middle of the observations for all age groups, except for the (usually open-ended) oldest where only the CMR observation is within the estimated range, while the other studies report (sometimes much) lower rates.

Table 2 - Characteristics of the studies of incidence of diabetes mellitus in the Netherlands

Name of the study	Period	Sample size	Age	Case-ascertainment
Sentinel Stations (3)	1980-1983	162,626	0-65+	report gp
Sentinel Stations (4)	1990-1991	160,000	0-80+	report gp
CMR [*] -Nijmegen (5)	1971-1993	12,000	0-75+	report gp
National Study (6)	1987-1988	335,000	0-75+	report gp in 3 months period

* CMR: Continuous Morbidity Registration
Report GP is a report by a general practitioner

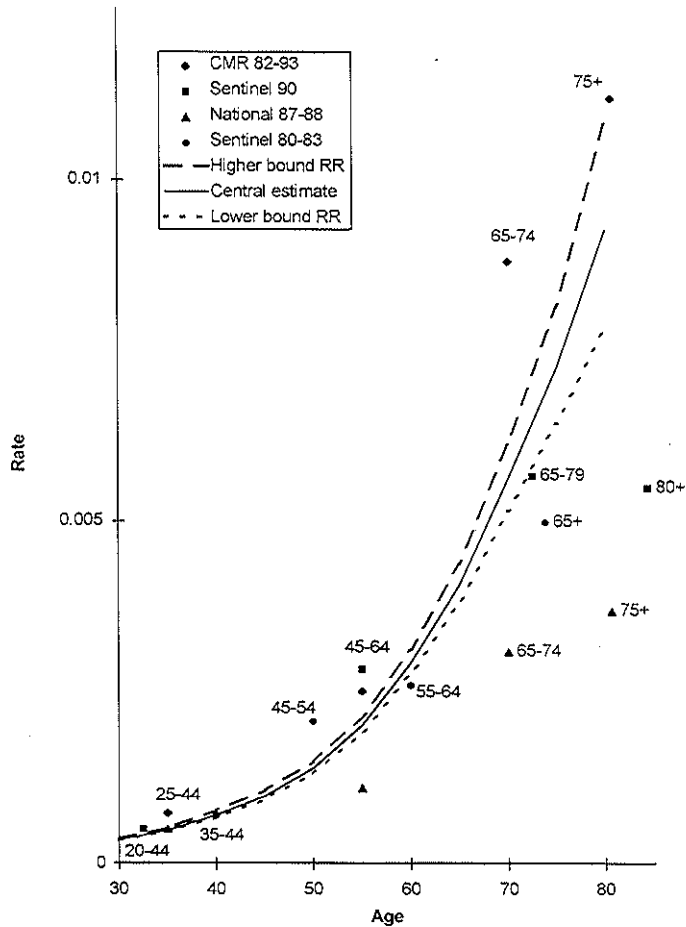


Figure 2 - Incidence of NIDDM in the Netherlands by age, Men, estimates based on IPM model, compared with age-specific observations from 4 empirical studies.

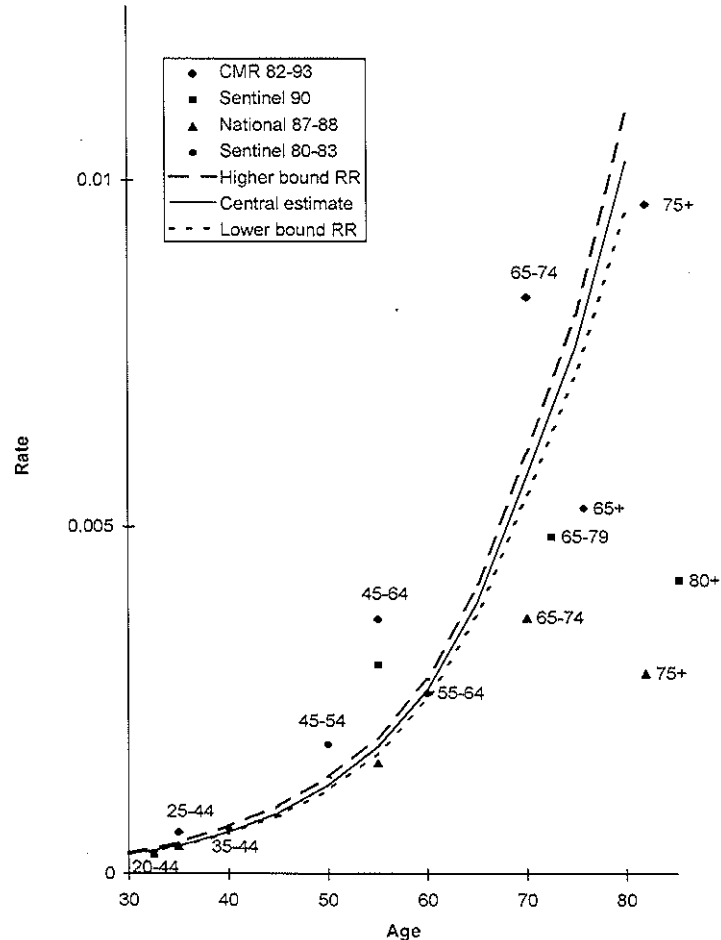


Figure 3 - Incidence of NIDDM in the Netherlands by age, Women, estimates based on IPM model, compared with age-specific observations from 4 empirical studies.

Discussion

To estimate incidence and prevalence of NIDDM sufficiently large empirical studies with adequate follow-up are preferred. To conduct such a study is difficult and expensive though, and consequently most studies are too small to yield reliable estimates. This is true in particular for incidence estimates, of which there are only a few, and they are plagued by small numbers when age specific rates are required. As a result the estimates vary widely.

As an alternative to direct observation, this paper describes an indirect method to estimate the incidence of NIDDM. It is based on the prevalence estimate from a previous study, a pooled analysis of 15 empirical studies. Added to that is information about the excess mortality of diabetics as compared to non-diabetics (16), and theoretical knowledge about the relation between incidence, prevalence and mortality.

Since no sex and age-specific estimates of relative risks for the Netherlands are available, excess mortality is estimated using relative risks on all cause mortality from the Verona study. This of course assumes that these relative risks are applicable to the Netherlands. We feel that this assumption is justified, since, (although performed in different counties, in different periods and with different methods), most studies report rather similar estimates of relative risks, taken into account the differences in age categories (13-17). For a number of reasons the Verona study was chosen: it is the most recent study, they reported age and sex specific relative risks, using the non-diabetic population as the reference.

With the relative risks from the Verona study, a diabetes-related mortality rate is constructed. Because the diabetes-related mortality rate includes all excess mortality for diabetics, it is independent from 'all other' mortality. Therefore it can be used, unlike the specific mortality restricted to diabetes mellitus (ICD 250), in an incidence-prevalence-mortality (IPM) model that assumes independence from the non-diabetes-related mortality.

The diabetes-related mortality and the prevalence rate are subsequently used as input for such an IPM model to yield incidence estimates. The estimates comply well with the available empirical incidence estimates from four Dutch incidence studies, but have upper and lower bounds that are considerably more narrow than the observed range of values. Three of the four empirical studies reported for the open ended (oldest) age groups a (much) lower incidence as compared with our estimates. This may be due in part to the calculation of the average age for the oldest age groups with a life table. Institutionalized people, who, on average, will be older than non-institutionalized, were not included in the empirical studies, but are in the life table. The calculated average age will therefore tend to be higher than the actual one in the studies, causing the data points to be plotted too far to the right.

A second reason why these three studies report lower rates than the CMR-Nijmegen study might be study design. CMR-Nijmegen is a small study, based on 4 general practices, with an active policy of completeness. The three other studies are much larger, and more dependent on the willing

participation of GPs. This will produce a propensity for too low estimates, which, as always, will tend to be stronger for the oldest patients.

An assumption implicit in using cross-sectional data is that there are no time trends in incidence and survival. With strong recent time trends in incidence or survival currently observed cross-sectional prevalence will not be at their equilibrium values. For example, if incidence has recently increased, it will take several years for prevalence to reach the new higher equilibrium value. Since the incidence estimate is based on prevalence, it will reflect conditions prevailing in the past, so our estimates may diverge from current incidence.

The IPM model used in this study ignores remission: an extra parameter would be required, with uncertain value. One study evaluated the diabetic status of patients after 8 years of follow-up, and reported a remission of 14% (20). However, there are reasons to suspect misclassification at baseline to be responsible for at least some of these cases. We assumed remission to be mostly temporary and too small to affect the incidence estimates substantially.

A weak point is the lack of an estimate for ages over 80. It would be desirable to extend the age range of the analysis beyond this age: NIDDM is a disease of the elderly, and on current mortality rates over half of the women and more than a third of the men survive to at least that age. Since the estimates will be used to assess the burden of disease and potentials for prevention for NIDDM, we feel that such large parts of the population ought not to be excluded. However, without a reliable prevalence estimate the procedure to estimate incidence breaks down. The message here is that better data, in particular on the oldest old, are badly needed.

The resulting incidence estimates are internal consistent with the prevalence and mortality estimates, which is not necessarily a property of empirical estimates, even when done in the same population. For example, the huge increases in prevalence (+40%) simultaneous with large decreases in incidence (-20%), as reported for Manitoba between 1986 and 1991, are difficult to explain (21, 22).

In a recent study Garancini *et al* estimated NIDDM incidence from data on prevalence and disease duration using data from the Cremona study (23). In table 3 their results are compared with the results of the present study. Their results are rather different: for ages between 40 and 70 their estimates are much higher, and they

Table 3: NIDDM incidence estimates per 10,000, from the Cremona study (23) and present study, by age group and sex

Age	Cremona		present study	
	Men	Women	Men	Women
30-39	5	3	5	4
40-49	15	15	10	9
50-59	43	30	21	19
60-69	70	78	42	41
70-79	58	67	74	79
80-89	31	50	93*	103*

*At age 80

reach a peak of 70 (males) and 78 (females) per ten thousand for the 60-69 age group (our estimate is about 40 per 10,000), and subsequently slump down to about 30 (males) and 50 (females) per 10,000 for the 80-89 group.

This result strikes us as rather implausible: the declining incidence suggests a pool of susceptibles that is being drained. But NIDDM is a senescent disease, glucose tolerance keeps declining with age for most people, and not just for a limited group (24, 25). While it is possible to nevertheless obtain such a peaked incidence curve of known diabetes, for example with a strongly age dependent intensity of case finding, the usefulness as a description of diabetes epidemiology seems limited.

The main result from our study are estimates of NIDDM incidence, that compare well with the few empirical estimates available, and, importantly, are consistent with a corresponding set of prevalence and excess mortality rates. This latter property allows to use the set as the basis for further work, that will concentrate on the assessment of the burden of disease of diabetes, and on the potentials for prevention.

References

1. King H, Rewers M. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. *Diabetes Care* 1993;16:157-177.
2. Baan C, Bonneux L, Ruwaard D, Feskens E. Prevalences of diabetes in the Netherlands in the Netherlands. A quantitative review. *European Journal of Public Health* (in press)
3. Steering Committee on Future Health Scenarios. Chronic diseases in the year 2005. Dordrecht-Boston-London: Kluwer Academic Publishers, 1991. (vol Volume 1: Scenarios on Diabetes Mellitus 1990-2005).
4. Ruwaard D. Diabetes Mellitus, from epidemiology to health policy [Thesis]. Erasmus University, 1996.
5. Lisdonk van de EH, Bosch WJHM, Huygen FJA, Lagro-Jansen ALM, ed. *Ziekten in de huisartspraktijk*. Utrecht: Wetenschappelijke uitgeverij Bunge, 1990:342. .
6. Velden van der J, Bakker de DH, Claessens AAMC, Schellevis FG. Een nationale studie naar ziekten en verrichtingen in de huisartspraktijk. Basisrapport: Morbiditeit in de huisartsenpraktijk. Utrecht: Nederlands Instituut Voor Onderzoek van de Eerstelijnsgezondheidszorg, 1991. .
7. Mackenbach JP, Snels IAK, Friden-Kill LM. Diabetes mellitus als doodsoorzaak. *Ned Tijdschr Geneesk* 1991;135:1492-1496.
8. Jouglé E, Papoz L, Balkau B, Maguin P, Hatton F. Death certificate coding practices related to diabetes in European countries - the EURODIAB subarea C study. *Int J Epidemiol* 1992;21:343-351.
9. Murray CJ, Lopez AD. Quantifying disability: data, methods and results. *Bull World Health Organ* 1994;72(3):481-94.

10. Bonneux L, Barendregt JJ, Looman CWN, van der Maas PJ. Diverging trends in colorectal cancer morbidity and mortality: earlier diagnosis comes at a price. *Eur J Cancer* 1995;31A(10):1665-71.
11. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke Trends in an Aging Population. *Stroke* 1993;24:931-939.
12. Manton KG, Stallard E. *Chronic Disease Modelling*. London: Charles Griffin & Company Ltd, 1988: 279+viii. (Baily NTJ, ed. *Mathematics in Medicine*; vol 2).
13. Panzram G, Zabel-Langhennig R. Prognosis of diabetes mellitus in a geographically defined population. *Diabetologia* 1981;20:587-591.
14. Kleinman JC, Donahue RP, Harris MI, Finucane FF, Madans JH, Brock DB. Mortality among diabetics in a national sample. *Am J Epidemiol* 1988;128:389-401.
15. Wong JSK, Pearson DWM, Murchison LE, Williams MJ, Narayan V. Mortality in diabetes mellitus: experience of a geographically defined population. *Diab Med* 1991;8:135-139.
16. Muggeo M, Verlato G, Bresson F, et al. The Verona diabetes study: a population-based survey on known diabetes prevalence and 5-year all-cause mortality. *Diabetologia* 1995;38:318-325.
17. Fuller JH, Elford J, Goldblatt P, Adelstein AM. Diabetes mortality: new light on an underestimated public health problem. *Diabetologia* 1983;24:336-341.
18. Barendregt JJ, van Oortmarsen GJ, van Hout BA, van den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Mathematical Population Studies* (in press).
19. Baan C, Nusselder W, Barendregt J, Ruwaard D, Bonneux L, Feskens E. The burden of mortality of diabetes mellitus in the Netherlands. (Submitted).
20. Stern MP, Valdez RA, Haffner SM, Mitchell BD, Hazuda HP. Stability over time of modern diagnostic criteria for type II diabetes. *Diabetes Care* 1993;16:978-983.
21. Blanchard J, Ludwig S, Wadja A, Dean H, Anderson K, Kendall O, et al. Incidence and prevalence of diabetes in Manitoba, 1986-1991. *Diabetes Care* 1996;19:807-811.
22. Leibson C, Milton L, Palumbo PJ. Temporal trends in Diabetes Incidence and Prevalence [Letter]. *Diabetes Care* 1997;20(3):460-461.
23. Garancini M, Gobbi C, Errera A, Sergi A, Gallus G. Age-specific incidence and duration of known diabetes. The Cremona study. *Diabetes Care* 1996;19:1279-1282.
24. West K. *Epidemiology of diabetes and its vascular lesions*. New York: Elsevier, 1978.
25. King H, Dowd JE. Primary prevention of type 2 (non-insulin-dependent) diabetes mellitus. *Diabetologia* 1990;33:3-8.

Changing cardiovascular epidemiology and its impact on diabetes prevalence: a quantification

Abstract

Background: *Several studies have reported increases in diabetes prevalence. Various causes for such an increase may exist, but one might be the recent marked improvement in heart disease survival. Since diabetes is a risk factor for heart disease better heart disease survival will tend to cause higher diabetes prevalence. We quantitatively examine the impact of this effect.*

Methods: *Life table techniques are used to estimate diabetes specific heart disease epidemiology. The changes in heart disease epidemiology between 1980-83 and 1990-93 are used to estimate the impact on diabetes prevalence of the decline in mortality selection. Sensitivity of the result to alternative assumptions on the excess risk for diabetics is determined.*

Results: *Diabetes prevalence shows an age related increase with a maximum of about 10% (numbers) and 2% (rates) for men at high ages. The impact on diabetes prevalence among women is much smaller. The comorbidity of heart disease and diabetes shows age related increases that reach 50 to 70% at higher ages (both sexes).*

Conclusion: *The changes in heart disease epidemiology are unlikely to be a major cause of increasing diabetes prevalence, but they are likely to cause a sizable increase in the comorbidity of heart disease and diabetes.*

Introduction

Recently studies from various countries have reported increases in diabetes prevalence.(1-3) While the evidence is not always wholly convincing, the matter certainly deserves attention, due to the considerable impact of diabetes on public health. Diabetes mellitus is an important risk factor for a number of diseases, among them cardiovascular disease (4), and an increase of prevalence of diabetes may therefore adversely affect cardiovascular incidence, prevalence, and mortality.

Understanding the causes of the increase in diabetes prevalence is vital to the development of effective interventions. Increased incidence is an obvious potential cause, and many countries have indeed seen increases in IDDM incidence.(5, 6) For NIDDM the evidence is less clear (1, 2, 7), if only because incidence of NIDDM is less well-defined. Indeed, on some estimates about half of NIDDM patients goes about undiagnosed (8), and this circumstance in itself might be responsible for the reported prevalence increase: with such a large pool of undiagnosed patients a minor enhancement in case finding practice will have a major impact on reported prevalence.

Yet another potential explanation for increased diabetes prevalence is related to its role as a risk factor for cardiovascular disease. Because diabetics run a higher risk on cardiovascular disease they are subjected to cardiovascular mortality selection: the increased risk of cardiovascular death tends to lower diabetes prevalence, since diabetics are selectively being removed from the population. For small causes of death the effect on diabetes prevalence will hardly be discernable, but since cardiovascular disease is such a large one it will.

The existence of mortality selection opens a potential for increases in diabetes prevalence: when the mortality selection eases, diabetes prevalence will go up. Since the early 1970's mortality from cardiovascular disease has been going down tremendously in the Netherlands, as in many other countries. Between 1970 and 1990 age-standardized mortality from ischemic heart disease and stroke halved, and this decline can be attributed to a combination of lower incidence and improved survival.(9-11)

In this article we estimate the effect of changes in heart disease epidemiology on diabetes prevalence. We use data on heart disease incidence and survival and their trends, and life table techniques to combine them with diabetes incidence and prevalence. We calculate the increase in diabetes prevalence and in the comorbidity of diabetes and heart disease as a consequence of the trends in heart disease incidence and survival.

Methods

To model the diabetes and heart disease processes we use an extension of the standard life table, the multi-state life table, that distinguishes not just between 'alive' and 'dead', but allows for additional states like 'alive, healthy',

'alive, with heart disease', etc (12). The model includes two disease processes: heart disease and diabetes, which are modelled as continuous time markov processes.(13) To allow for the heterogeneity of a population consisting of diabetics and non-diabetics the population is divided over two life tables: one for diabetics and one for non-diabetics.

Each life table distinguishes a non-heart disease state, and a heart disease state. In each state the life table cohort is subjected to 'all other mortality' probabilities. In the heart disease state an additional disease specific mortality probability is in force, which we assume to be higher for diabetics. People not in the heart disease state run a risk on incidence probabilities for heart disease, which in the diabetics life table are also higher. Therefore diabetics have a higher incidence of heart disease and worse survival than non-diabetics.

Diabetes incidence is used to link the two life tables: the life table cohort of 100,000 starts in the non-diabetics life table with the lower heart disease risks, at each age incident diabetes cases are calculated using the incidence rates. These cases are transferred to the diabetics life table, where they are subjected to the higher heart disease incidence rates.

Incidence and survival of heart disease is predominantly based on a nationwide hospital register, various other studies are used in addition.(11) Diabetes prevalence is based on a pooled analysis of 15 Dutch studies.(14) Excess mortality for diabetics was calculated using relative risks on total mortality from the Verona study.(15) Diabetes incidence was backcalculated from the excess mortality and prevalence, using the same continuous time markov process technique.(16)

We estimate diabetes specific heart disease epidemiology based on the average incidence, prevalence, and survival for the 1980-'93 period.(11) We assumed a rate ratio for heart disease mortality for diabetics of 3 at age 30, declining linearly to 1 at age 95.(4, 17) In addition we assumed a relative 10-year mortality risk of about 1.4 for diabetic heart disease patients, based on several studies reporting mortality risk.(18-20) Together these assumptions form the reference case, which allows to backcalculate diabetic specific heart disease incidence. To gauge sensitivity we calculated alternatives with the excess risks put to zero, and with double excess risks.

To estimate the influence of changes in heart disease on diabetes epidemiology the calculations were repeated with identical diabetes incidence and excess risk, but with heart disease epidemiology averaged over the 1980-83 and 1990-93 periods.(11)

Results

Figures 1 and 2 show diabetes specific prevalence of heart disease for the reference estimate 1980-1993 for men and women respectively. The differences between men and women in heart disease epidemiology are clear: men have much higher prevalences (note that the Y-axes have different scales), and female prevalence is shifted more to higher ages. Heart disease prevalence for diabetics is higher, except at the highest ages: incidence of heart disease at

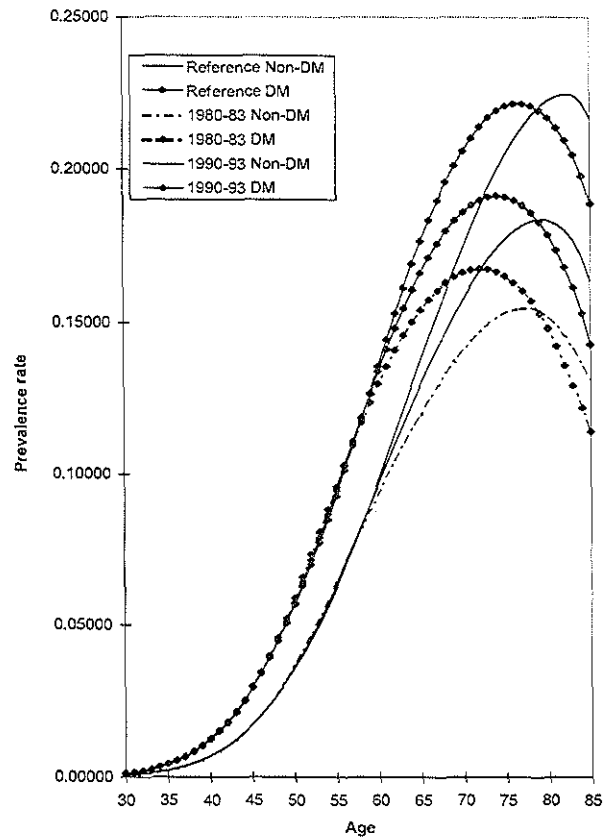


Figure 1 - Diabetes specific prevalence of heart disease, males, for the reference 1980-93 period, and for the 1980-83 and 1990-93 periods.

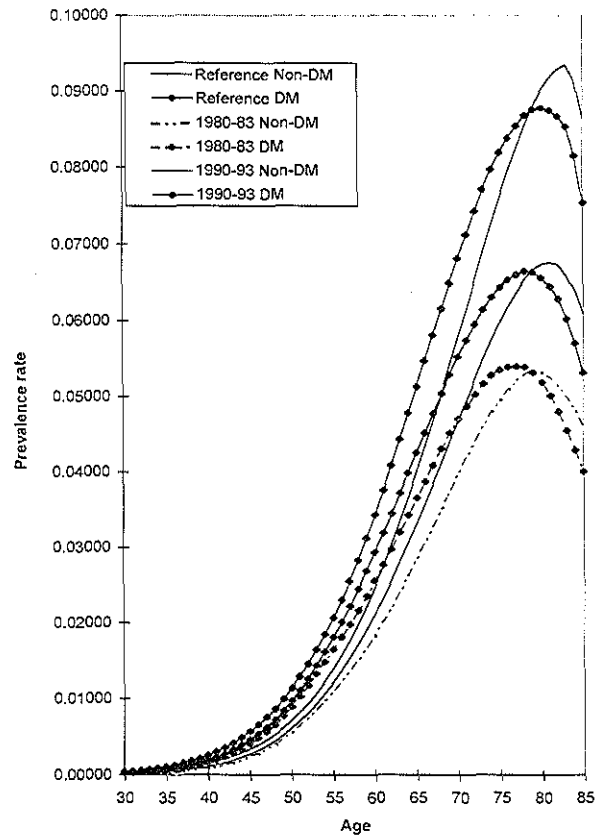


Figure 2 - Diabetes specific prevalence of heart disease, females, for the reference 1980-93 period, and for the 1980-83 and 1990-93 periods.

those ages is still higher for diabetics, but survival with heart disease declines with age and does so faster for diabetics because of the relative risk on mortality risk.

The higher heart disease mortality risk of diabetics causes mortality selection: figure 3 shows the effect on diabetes prevalence of changes in the relative risks. Putting the relative risks to 1 causes diabetes prevalence to increase, doubling the excess risks causes it to decline. Two things can be observed: the effect increase with age, but it is generally small (results for women are not shown, but even smaller).

The age effect is because mortality selection is stronger where mortality is higher, so in the younger age groups effects are smaller than in the higher. A higher relative risk on heart disease increases the mortality selection, and therefore lowers diabetes prevalence rates at higher ages. The estimate with a relative risk of 1 gives prevalence in the absence of mortality selection, and the difference between this estimate and the best estimate is the potential increase of diabetes prevalence when mortality selection diminishes.

Figures 1 and 2 also show heart disease prevalence based on the periods 1980-83 and 1990-93. The difference between the periods is a considerable increase in prevalence for the higher age groups, mostly caused by improved survival. In addition there was a small decrease in incidence for the younger groups among men, but the reverse seems to have happened among women. (11)

Figures 4 and 5 show the effect on diabetes prevalence and the comorbidity of diabetes and heart disease, expressed as indices with 1980-83=100. The effect on diabetes prevalence rate is again small: it reaches a maximum of about 2% for men at high ages, for women it is even less. More pronounced is the increase of diabetes prevalence in numbers: improved survival with heart disease causes an age dependent rise, which reaches over 10% at age 80 (men, with women again the effect is considerably less).

The most striking result in figures 4 and 5 is the increase in the comorbidity of heart disease and diabetes: in both rates and numbers an age-dependent increase that reaches a maximum of 50 to 70% for the eldest. A notable difference between men and women is that the latter also see a considerable increase for ages under 65, while for the former the increase in that age group is much more limited. This is caused by the difference in heart disease epidemiology changes between the sexes in the age group: a decline in incidence for the men, but an increase for the women.

Discussion

Although the evidence is not unequivocal there are several reasons to suspect that diabetes prevalence is increasing. One reason is increased incidence, observed for type I. For type II a possible increase in incidence might be caused by changes in risk factors, like obesity and physical activity.(1) The situation is complicated by the presumed large pool of undiagnosed type II diabetics,

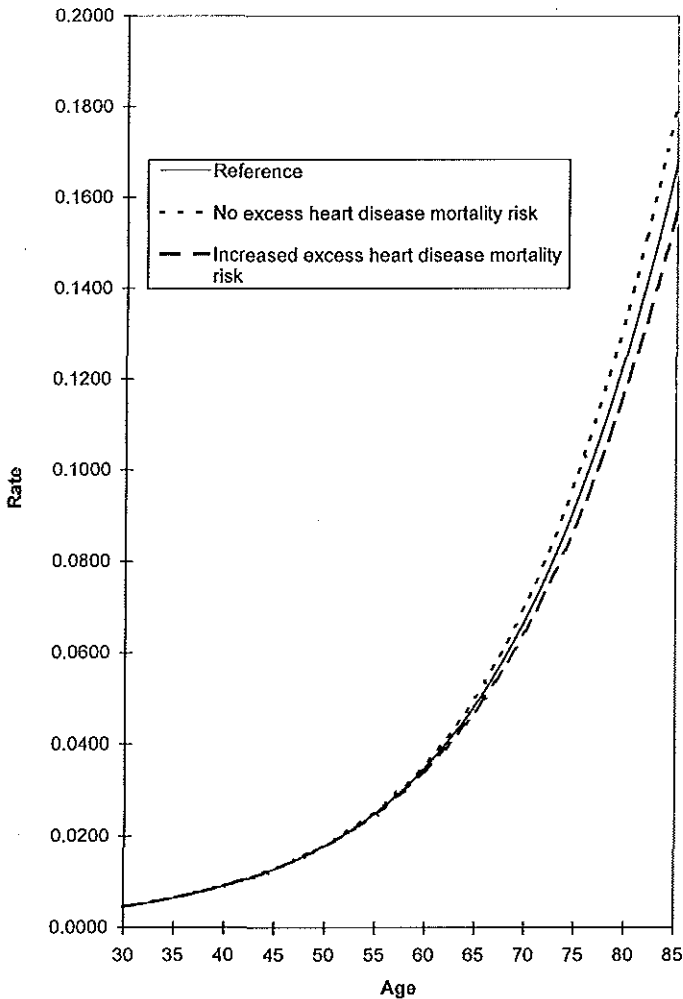


Figure 3 - Diabetes prevalence under various heart disease mortality risks, males.

that makes stepped-up case finding result in a large increase in clinical prevalence, even in the absence of any real epidemiologic change.

To this already confusing picture yet another explanation for increased diabetes prevalence is added: diminished cardiovascular mortality selection. We have shown that if the changes in heart disease epidemiology as observed in the last decade, mostly an improvement of survival, apply equally to diabetics this results in an increase in diabetes prevalence in the life table population, both in rates and in numbers.

The fact that not just prevalence numbers increase, but prevalence rates as well, shows that it concerns more than substitution of disease. Larger

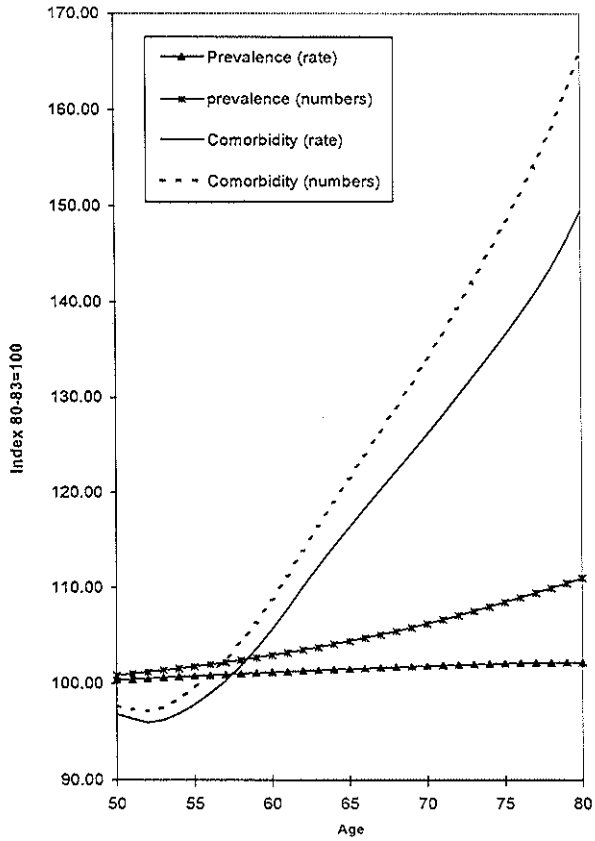


Figure 5 - Prevalence of diabetes and of comorbidity of heart disease and diabetes. Males by age, indices of 1990-93 estimates with 1980-83 =100.

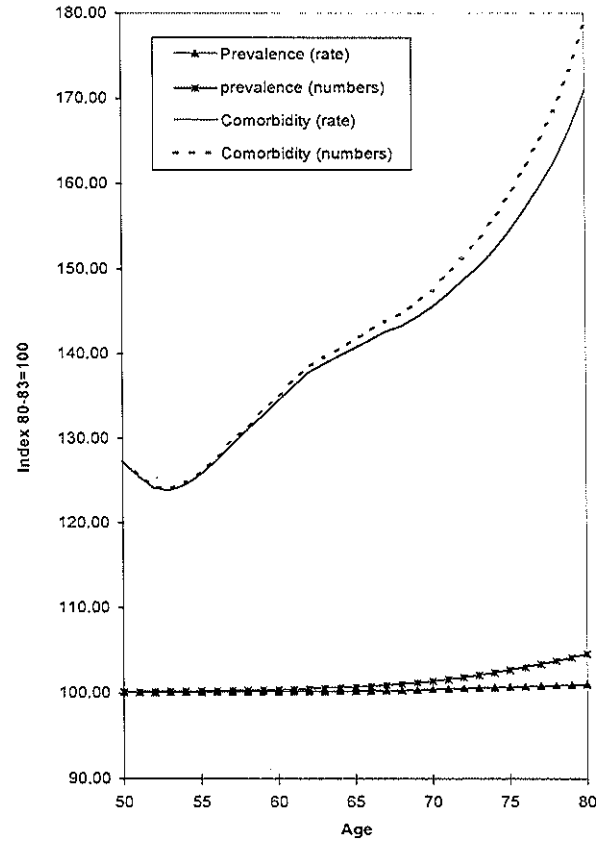


Figure 4 - Prevalence of diabetes and of comorbidity of heart disease and diabetes. Females by age, indices of 1990-93 estimates with 1980-83 =100.

numbers of people alive, as is the case in the higher age groups after the heart disease trends, means larger numbers of diabetics, even when the diabetes prevalence rate is unchanged. But since diabetics have higher heart disease risks than non-diabetics, they are more strongly affected by the lower mortality, mortality selection eases, and the prevalence numbers of diabetics in the higher age groups increase faster than non-diabetics. Hence the higher diabetes prevalence rates in the population.

The increase has a distinct age pattern: none for the younger age groups, and growing with age. This age pattern is inherent to diminished heart disease mortality selection: it only operates in the age groups where heart disease mortality is substantial.

A similar but enhanced age pattern can be seen in the comorbidity of diabetes and heart disease: for the men the comorbidity prevalence in the younger age groups is even lower (due to lower heart disease incidence) with substantial increases in the highest age groups. The comorbidity increase is almost completely due to the increase in heart disease prevalence rates, the diabetes rate increase only adds marginally.

Our calculations show that the impact of improved heart disease survival on diabetes prevalence is generally rather small. A notable increase in the number of elderly diabetic men can be observed, but since the number of non-diabetic men of the same age also increases the diabetes prevalence rate hardly moves. For women, due to the much smaller importance of heart disease, the impact is even less. This result is rather robust: doubling the excess risk of diabetics or removing it altogether hardly makes a difference (see figure 3).

Moreover, there are two reasons why the impact may have been overestimated. The first is that the introduction of new medical interventions are probably in large part responsible for improved heart disease survival. With new interventions comorbidity initially, and perhaps permanently, acts as a contraindication. Presumably, for example, thrombolysis was applied to diabetic heart patients more reluctantly than to the general population.(21) This affects our assumption that the improvement in survival observed for the general population equally applies to the diabetic population, with the result that measured effects will tend to be smaller.

The second reason for overestimation lies in the nature of life table calculations. Life table analysis produces an estimate that fully reflects the impact of the changed incidence and survival, in reality it takes time for this impact to develop. The increase in heart disease prevalence between 1980-83 and 1990-93 will, due to these dynamic effects, have been smaller, and consequently the same goes for the impact on diabetes prevalence.

On the other hand two reasons to suspect underestimation exists. The heart disease epidemic seems to have peaked in the early seventies, so the decline in mortality has been going on for a longer time.(22) And not just heart disease mortality has been declining: stroke, another disease diabetes is a risk factor for, has also seen declining mortality and presumably incidence for a long time.(10)

In any case the comorbidity increase that our calculations produce seems to be big enough to have been observed empirically. The Minnesota Heart Survey reported increased diabetes prevalence among cardiovascular patients.(20, 23)

On balance we conclude that the changes in heart disease epidemiology are unlikely to be a major cause of increasing diabetes prevalence, but that they are likely to cause a sizable increase in the comorbidity of heart disease and diabetes.

References

1. Leibson CL, O'Brien PCO, Atkinson E, Palumbo PJ, Melton III LJ. Relative contributions of incidence and survival to increasing prevalence of adult-onset diabetes mellitus: a population-based study. *Am J Epidemiol* 1997;146:12-22.
2. King H, Zimmet P. Trends in the prevalence and incidence of diabetes mellitus: non-insulin dependent diabetes mellitus. *World Health Stat Q* 1980;41:190-6.
3. Zimmet PZ, Alberti KGMM. The changing face of macrovascular disease in non-insulin dependent diabetes mellitus: an epidemic in progress. *Lancet* 1997;350:S1.1-S1.4.
4. Nathan DM, Meigs J, Singer DE. The epidemiology of cardiovascular disease in type 2 diabetes mellitus: how sweet it is....or is it? *Lancet* 1997;350:S1.4-S1.9.
5. Green A, Slojie AK, Eshoj O. Trends in the epidemiology of iddm during 1970-2020 in Fyn County, Denmark. *Diabetes Care* 1996;19:801-806.
6. Ruwaard D, Hirasng RA, Reeser HM, et al. Increasing incidence of type I diabetes in the Netherlands. *Diabetes Care* 1994;16:1-3.
7. Ruwaard D, Gijsen R, Bartelds AIM, Hirasng RA, Verkleij H, Kromhout D. Is the incidence of diabetes increasing in all age-groups in The Netherlands. Results of the second study in the Dutch Sentinel Practice Network. *Diabetes Care* 1996;19:214-218.
8. King H, Rewers M. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. *Diabetes Care* 1993;16:157-177.
9. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure: the future rise of heart failure. *Am J Public Health* 1994;84(1):20-8.
10. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke Trends in an Aging Population. *Stroke* 1993;24:931-939.
11. Bonneux L, Barendregt JJ, Van der Maas PJ. The new old epidemic of coronary heart disease. (Submitted).
12. Schoen R. Modelling Multigroup Populations. New York and London: Plenum Press, 1987: 297+xii. (Land KC, ed. The Plenum Series on Demographic Methods and Population Analysis;).
13. Barendregt JJ, van Oortmarssen GJ, van Hout BA, van den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Mathematical Population Studies* (in press).

14. Baan C, Bonneux L, Ruwaard D, Feskens E. Prevalences of diabetes in the Netherlands in the Netherlands. A quantitative review. *EJPH* (in press).
15. Baan C, Nusselder W, Barendregt J, Ruwaard D, Bonneux L, Feskens E. The burden of mortality of diabetes mellitus in the Netherlands. (submitted).
16. Barendregt JJ, Baan CA, Bonneux L. Consistent estimates of diabetes incidence, prevalence and mortality. (submitted).
17. Sniderman A, Michel C, Racine N. Heart disease in patients with diabetes mellitus. *J Clin Epidemiol* 1992;24:1257-1370.
18. Abbud ZA, Shindler DM, Wilson AC, Kostis JB. Effect of diabetes mellitus on short- and long-term mortality rates of patients with acute myocardial infarction: a statewide study. *Am Heart J* 1995;130:51-57.
19. Behar S, Boyko V, Reicher-Reiss H, Goldbourt U. Ten-year survival after acute myocardial infarction: comparison of patients with and without diabetes. *Am Heart J* 1997;133:290-296.
20. Sprafka JM, Burke GL, Folsom AR, McGovern PG, Hahn LP. Trends in prevalence of diabetes mellitus in patients with myocardial infarction and effect of diabetes on survival. The Minnesota Heart Survey. *Diabetes Care* 1991;14:537-543.
21. Hansen HT, Fog L, Kjergaard SC, Christensen PD, Bulow I. Thrombolytic therapy in diabetic patients with acute myocardial infarction. *Diabetes Care* 1996;19(10):1135-1137.
22. Bonneux L, Looman CW, Barendregt JJ, Van der Maas PJ. Regression analysis of recent changes in cardiovascular mortality and morbidity in the Netherlands. *Brit Med J* 1997;314:789-792.
23. Sprafka JM, Virnig BA, Shahar E, McGovern PG. Trends in diabetes prevalence among stroke patients and the effect of diabetes on stroke survival: the Minnesota Heart Survey. *Diabet Med* 1994;11(7):678-84.

PART IV

*Measuring
compression and expansion
of morbidity*

Introduction to part IV

Introduction

Compression and expansion of morbidity are explicitly concerned with the combination of the two fundamental aspects of population health status: morbidity and mortality. Whether either compression or expansion occurs depends on the balance between incidence of morbidity, cure and survival, and the resulting balance is a particular outcome of mortality and of prevalence of morbidity, weighted for severity.

Mortality is (and has since long been) measured routinely in many countries, including the Netherlands. Mortality by age and cause of death is being published by Statistics Netherlands annually. While the cause of death registration is not without its problems, in particular for the elderly, mortality data are generally considered robust, and are available.

Widely used indicators based on mortality are, in addition to (disease specific) numbers of deaths, derived measures such as *standardized mortality* (in several variants), *years of life lost* (ditto) and its counterpart *potential years of life gained*, and, perhaps the best known, *life expectancy*.

Morbidity data, on the other hand, are generally either not available or else not robust. Best recorded are data concerning health care utilization (like costs and hospital days), but these indicators are notoriously affected by changes in medical technology, prices, and attitudes towards health, and are therefore not considered good indicators of morbidity for most diseases. As part II has made clear, nationwide data on disease specific incidence and prevalence are not available for even such a large disease as cardiovascular disease. Other important diseases where morbidity data are scarce are diabetes and respiratory disease. And data on general morbidity, like disability and self-rated health, rely on surveys and face numerous problems of validity. The Dutch national health survey, for example, has a non-response

rate of about 40-45%, and excludes the institutionalized population (Van den Bos, van der Velden, van Sonsbeek, Nusselder, & Lenior, 1994).

Given the available data, researchers have tended to rely on mortality based indicators of population health status, with life expectancy at birth being a first choice. Life expectancy at birth is generally interpreted as the number of years a newborn can expect to live, and is an excellent way to summarize mortality data, but as an indicator of population health status it is one-sided because it ignores morbidity. This would be no problem if mortality and morbidity always moved in the same direction: in that case mortality is a good proxy for morbidity. But as has been shown in part II this is not always true: heart disease mortality has declined while simultaneously the prevalence of heart disease increased, and the trends in colorectal cancer incidence and mortality are diverging.

Mortality can be a good proxy for morbidity when a fixed and known proportion of incident (or prevalent) cases dies. But some diseases cause relatively little or no mortality but protracted periods of morbidity (locomotor and respiratory diseases, for instance), while yet others, in particular cardiovascular disease, combine elements of chronic and acute disease: chronic disease interspersed with acute periods that may have a fatal outcome.

But while mortality based indicators give a biased view of population health status, so do morbidity based ones. The changes in heart disease epidemiology again provide a good example: the increase in prevalence of heart failure indicates a worsening of population health status, a judgment few people would agree with, given the simultaneous decline in acute heart disease mortality.

The realization that population health status should be measured by looking at both mortality and morbidity has prompted researchers to construct indicators that combine the two. Given the plethora of indicators for mortality and especially for morbidity there are many ways to do this. Several indicators that combine both have actually been implemented. Such a combined indicator is a natural candidate for measuring compression and expansion of morbidity. In the remainder of this chapter we will look into the way mortality and morbidity may be combined in one indicator, list some implementation options, look at two implementations in some detail, and define a way to measure compression and expansion of morbidity using elements of both implementations.

Time as a common denominator

Since morbidity and mortality are obviously different things a common denominator is needed in order to combine them in one indicator. The common denominator is found by looking at the way successful mortality based population health status indicators, like *life expectancy* and *years of life lost*, are derived from mortality data.

Mortality is by itself not a very interesting indicator: since all people die the fact as such carries no information about population health status. The crucial point is the age at which mortality occurs, but age-specific mortality is an unwieldy indicator because it consists of a large array of numbers. To derive an indicator the age-specific mortality is converted to the average years of life lived (life expectancy) or number of life years lost. Thus the event of death is converted to time: time lived or lost (Murray, 1996).

The literal interpretation of these indicators raises problems. Life expectancy, as it is commonly understood, is calculated using period data, and therefore life expectancy at birth is *not* the number of years a newborn may expect to live. When the decline in mortality continues, as it presumably will, the newborn may expect to live longer, as has been the case during the 20th century (Murray, 1994).

Another problem is the interpretation of years of life lost. When we calculate years of life lost in a population by multiplying age specific numbers of deaths with the life expectancy at that age, we are implying that every member of the population can expect to live at least the average number of years, which is a contradiction in terms.

We therefore prefer not a literal interpretation, but one based on the observation that *life expectancy* and *years of life lost* are much more sensitive to mortality at younger ages than at older (Shryock & Siegel, 1976). In other words, deaths at younger ages are being considered more severe, and therefore both can be interpreted as mortality indicators that weight for age at death. Under the age-weighted interpretation the application of period mortality data poses no problems, and we do not implicitly perform contradictory calculations with *years of life lost*.

So *life expectancy* and *years of life lost* are age-weighted mortality indicators with time as the unit of measurement. Given this, it is possible to combine morbidity and mortality in one indicator when we express morbidity also in units of time, that are equivalent to the units used for the mortality indicators.

The simplest and probably most frequently applied way to do this is by considering time lived with morbidity equal to time not lived due to mortality. Most calculations of the so called *health expectancy* indicator use this procedure: health expectancy is calculated according to standard life expectancy method, with the difference that years lived with morbidity do not count towards the total years of life lived by the life table cohort (a formal expression for this calculation can be found in the appendix to chapter 16).

Other, more complex methods consider time lived with morbidity only partly equal to time lost due to mortality. The size of this part can be determined by trade-off methods like the *time trade-off* and *person trade-off*. In the former method subjects are presented a fixed period lived in a defined less than full health state, and are asked to decide which part of the period they would be willing to give up when the remainder is then lived in full health (Torrance, 1987). The *person trade-off* method uses a somewhat different perspective, but is basically the same (Murray, 1996).

As the less than full health state becomes more severe, most subjects are willing to give up a larger part of the period lived, and therefore the resulting fractions that divide the fixed period between full health and death, have the dual interpretation of severity weights for the health state under consideration. These methods are used in Quality of Life (QALY) research and to determine disability weights for the Disability Adjusted Life Years (DALYs).

Thus a continuum of population health status indicators emerges, with on the one extreme *life expectancy*, where years lived with morbidity are considered equivalent to years in full health, on the other *health expectancy*, which considers years lived with morbidity equivalent to years lost due to death, and in between indicators like the DALY, where years with morbidity are considered partly equal to years in full health, and for the remainder considered lost, with the allotment depending on the severity of the disease.

Implementation options

In addition to the different ways morbidity is treated there are a number of other implementation options for a population health status indicator. The actual choices made in a particular case should depend on the intended application of the indicator. Applications include monitoring population health status (current level and changes over time), understanding observed changes, estimating the benefits of potential interventions, and resource planning and optimal resource allocation. Depending on the application important choices are:

- *Life table versus real population.* The most important determinant of health care needs is the size and age structure of the population. Applications involving resource planning and allocation should therefore use a real population (which is meant to include projections of the future population). When comparing populations or studying trends in underlying population health status a life table population offers the advantage of abstracting from population size and age structure.
- *Static versus dynamic population.* When time is an issue in the application, for example when not just the size of an effect matters, but its timing as well, a dynamic population is necessary. In general time will be more an issue in applications involving resource planning. In many other cases simpler static method will serve.
- *Prevalence versus incidence/duration.* With a prevalence approach the amount of morbidity in the population is estimated directly, with an incidence/duration approach estimates of incidence and duration are used to calculate the amount of morbidity. In the absence of past trends both methods will yield the same result, but otherwise the prevalence method describes the actual situation, including the legacy from the past, while the incidence/duration approach ignores the past and calculates what the

present would look like if current incidence and duration would have been in force in the past (Crimmins, Saito, & Hayward, 1993). The advantage of the incidence/duration approach is the causal link from incidence to morbidity prevalence and mortality, which allows estimation of intervention effects. In particular in combination with dynamic methods this is a very powerful, but also complex and data hungry method. The prevalence approach is much simpler, and better suited for estimation of current population health status and care needs.

- *Definition of morbidity.* In principle any morbidity measure can be employed: generic measures like disability, self-rated health, or health care utilization; or one or more specific diseases or conditions. Often, and in particular for generic morbidity measures, a threshold value has to be chosen. These choices are inevitably rather arbitrary, and have a large impact on indicator outcome (Murray & Lopez, 1997).
- *Weights for severity of morbidity.* The previous section explained that in order to combine mortality and morbidity time lived with morbidity is made equivalent to time lost due to mortality. This can be done by simply equating the two (giving up all the time lived with morbidity, which has the dual interpretation of a 0-weight), but more complex weighting schedules are also possible with, according to severity, different weights for different health states, diseases, or both.
- *Other weights.* In addition to morbidity weights other weights have been employed to modify time lost due to morbidity and mortality. The original DALY, as developed by Murray and colleagues and applied in the Global Burden of Disease project, uses an age-weighting schedule that intends to stress the time lost between the ages of about 10 and 55, while downplaying losses outside that range (Murray, 1994). In a dynamic analysis (i.e. one that includes calendar time) losses and benefits are often weighted for time preference (*discounted*), causing these losses and benefits to be less important as they occur further away in time (Drummond, Stoddart, & Torrance, 1990). In economic analysis time lost is often weighted for income (the *human capital* method) (Koopmanschap & Rutten, 1993). Many more weighting schedules are thinkable, but few are desirable and sound, and even all three schedules mentioned here have their critics.

Actual implementations

Health Expectancy

From various permutations of the options in the previous section a large number of indicators can be constructed, but two implementations dominate the field: *health expectancy* and DALY. Fortunately they are also good examples, because on a rating scale from simple to complex they occupy

positions far apart, with *health expectancy* being a simple indicator and DALY a complex.

Health expectancy stands for a family of indicators which may differ in details, but the vast majority share the following characteristics (Robine & Ritchie, 1991). A *health expectancy* always uses a life table population, is static, and in most cases employs the prevalence approach (dubbed "Sullivan-method"), a generic morbidity measure, and most calculate the indicator by excluding time lived with morbidity (effectively giving a 0-weight to that time). The prevalence method allows the use of cross-sectional data, for example from a health survey, and the calculation is straightforward (see appendix to chapter 16).

Different definitions of what constitutes morbidity have been used, but a popular generic measure is "disability", resulting in the Disability Free Life Expectancy (DFLE). But disease specific morbidity can also be used, as in the calculation of a dementia-free life expectancy (Ritchie, 1994).

The simplicity of this indicator is very attractive, and makes it particularly suited for inexpensive and quick estimates of current population health status. For many countries estimates of, in particular, the DFLE have been made, but unfortunately comparisons between country estimates are being hampered by differences in the definition of disability, threshold values, and survey methods (Boshuizen & Van de Water, 1994).

The usefulness of the health expectancy indicator, in its most used implementation with cross-sectional prevalences and a 0-1 weighting schedule, for answering the question of compression versus expansion of morbidity is also limited. The ability to make comparisons over time is hampered by the use of the prevalence method. In the prevalence method a stock variable (prevalence) is combined with a flow variable (mortality). Stock variables adjust only slowly to changes in the flow variables that govern their value, and in chapter 16 it is shown, using the heart disease model described in chapter 4, that this characteristic of stock variables leads to spurious trends in a health expectancy indicator based on the prevalence approach.

Problems with the health expectancy indicator also originate from the routinely used 0-weight for time lived with morbidity. This, firstly, makes the indicator very insensitive to changes in severity of disability: only when the changes cause people to cross the threshold value of disability will such changes show up in the indicator. And, secondly, as we show in chapter 17, an improvement of survival with morbidity results as a consequence of this 0-weight in an unchanged health expectancy at birth, and even in a declining one at higher ages, rather than showing an increase. This counterintuitive result follows because the total number of healthy years lived by the life table cohort remains unchanged, while at higher ages the cohort will have more survivors, thus less healthy years on average. When researchers are aware of this problem they can avoid the declining health expectancy at higher age (chapter 17), but as long as the survival improvement is not accompanied by an improvement in health status such that at least part of the years lived with morbidity previously will become healthy years, the health

expectancy indicator with a 0 morbidity weight will not show an increase after survival improvement. This seems a rather awkward characteristic for a population health status indicator to be used for the study of compression and expansion of morbidity.

DALYs

The concept of *Disability Adjusted Life Years (DALY)* was developed with the optimal allocation of scarce resources in mind (World Bank, 1993). It uses a real population structure, incidence/duration methods for the description of a large number of conditions, and morbidity weights by condition and severity. The DALY gives the number of life years lost by the population due to mortality and morbidity, with the latter calculated by multiplying time lived with a specific disease by its weight for disability severity.

Several other attributes distinguish the DALY. One is that future years of life lost are being discounted (by 3%) to weigh for time preference (see part V for more on discounting). A second is the use of a standard life table (model west, level 26) for the calculation of years lost due to mortality. This table has a life expectancy that is higher than observed in any nation to date, and therefore the number of life years lost is systematically over-estimated, not much for western countries and Japan, but quite a lot for sub-Saharan countries.

Murray argues that the use of the same life table worldwide prevents a death at a given age in a high mortality country to be less important than one in a low mortality country, and that equity requires that they should be valued the same (Murray, 1994). This makes sense, as long as the resulting estimate is not interpreted as actual years of life lost. Above we suggested an alternative interpretation of *years of life lost* as an age-weighting schedule for mortality, with death at a younger age being worse. This alternative avoids the interpretation problem.

In addition to the age-weights for mortality by way of the *years of life lost* the DALY uses age-weights for both morbidity and mortality, that intent to count the years lost in the age range of about 10 to 55 as the more important, while playing down losses outside that range (Murray, 1994). In chapter 18 it is shown for the case of mortality that due to combined effects of the two age-weighting schedules not the range between about 10 and 55 years of age is emphasized, but rather that between 0 and 27.

Measuring compression and expansion: Disability Adjusted Life Expectancy

Health expectancy (in the DFLE or similar variants) and DALY constitute just two instances of the set of possible population health status indicators that combine morbidity and mortality. For our purpose of measuring compression and expansion we can construct a suitable indicator and define

compression and expansion in its terms. This indicator is implemented as follows:

- *Life table population.* The indicator is to be used for comparison over time, and therefore should abstract from changes in population size and structure.
- *Disease specific and generic morbidity.* We want to link disease specific changes to population health status outcome, and therefore a disease specific approach is necessary. But simultaneously we need generic morbidity: unless a very large number of diseases is modeled the morbidity described by the diseases will be only part of total morbidity. The solution is to describe the remaining morbidity as generic morbidity.
- *Incidence/duration and prevalence approach.* Our objective is mostly to understand and explain observed, and estimate potential compression and expansion from changes in disease incidence and survival, and therefore we need the causal model provided by the incidence/duration approach. But the remaining, generic morbidity we assume to be constant, and for that part of morbidity the prevalence approach will suffice.
- *Weights for morbidity severity.* The indicator uses morbidity weights by disease and, when changes within diseases are the subject, by severity of disease state, to avoid the problems that the 0-weight for morbidity causes.

In chapter 19 we introduce the Disability Adjusted Life Expectancy (DALE), an indicator that combines aspects of both health expectancy and DALY. It is like a health expectancy in that it uses a life table and an estimate of the generic morbidity measure disability. From the DALY it gets disease specific incidence/duration models and disability weights for morbidity.

With the DALE we can now define compression and expansion of morbidity. We define compression to occur when the life expectancy with disability (LED, the difference between life expectancy and DALE) decreases, and expansion when it increases. Under this definition an increase in life expectancy leads to expansion of morbidity as the DALE increases less (or remains the same or declines). Note that under this definition compression is not necessarily a good thing (and neither expansion a bad one): compression can occur with declining life expectancy and DALE, as long as the life expectancy declines faster.

Often in addition to the above definition, also called absolute compression and expansion, the concept of relative compression and expansion is used (Robine & Mathers, 1993). We say that relative expansion occurs when the ratio of LED and life expectancy (LED/LE) increases, and relative compression when it declines. Note that under this definition relative compression can occur simultaneously with absolute expansion, and absolute compression with relative expansion. This is in contrast with the definition by Robine and Mathers, which is an effort to make these categories mutually exclusive, but which produces difficult to classify cases when the life

expectancy declines (Robine & Mathers, 1993). In general it is best to report both life expectancy and DALE and their changes, and not just the LED, to give a full representation of the changes in population health status.

The DALE, like the DALY, links the epidemiological concepts of disease incidence and survival to the population health status outcome, and can therefore be used to show that recent trends in cardiovascular disease caused an increase in life expectancy and expansion of morbidity, while the abolishment of smoking would result in increasing life expectancy with compression of morbidity. Put differently: the DALE is a population health status indicator fit to analyze the consequences of epidemiological trends and medical interventions in terms of the compression and expansion of morbidity.

But chapter 19 also lists two methodological problems with the DALE, both originating from the use of a bottom-up, disease specific approach. The first problem is the existence of comorbidity. Comorbidity causes problems in two ways: it breaks the one-to-one correspondence between prevalence of disease and prevalence of disability, and in addition it is unclear what disability weights should be applied to the comorbid conditions. In chapter 20 we explore the uncertainty and sensitivity surrounding the calculation of years lived with disability by the DALE life table cohort. We conclude that in many cases the exact value of the disability weight will have only a minor impact on the indicator, and that it might even be justified in aggregate analyses to ignore comorbidity altogether.

The second methodological problem with the DALE as introduced here is the estimation of the "all causes" disability. Murray uses a "brute force" approach for the DALY and the DALE to describe total disability by including a very large number of diseases and conditions (Murray, 1994; Murray & Lopez, 1997). Our DALE computation explicitly describes the disability caused by a limited number of diseases only, and obtains the supposedly constant remainder by subtraction from an "all-causes" estimate. The latter is derived from a Disability Free Life Expectancy in a rather rough procedure, and while the result looks plausible enough, the procedure is certainly open for improvement.

In most cases the DALE can be implemented as a static life table measure, but there is no reason why a dynamic version would be impossible. In chapter 24 a dynamic life table population method is used (among others), which could easily be applied for a dynamic DALE too. Also a real instead of life table population could be used, which would make the indicator very much, if not quite, the same as a DALY approach.

The general conclusion here is that there exists a range of possible implementations for indicators that combine morbidity and mortality. DFLE, DALY and DALE are but three instances of a large set of possible implementations. The main point is to choose or construct the implementation that, given the application at hand, is most appropriate.

References

- Bos, G. A. M., van den Velden, J., van der Sonsbeek, J. L. A., van Nusselder, W. J., & Lenior, M. E. (1994). *Gezonde levensverwachting en chronische ziekten: vergelijking van databronnen en berekeningen*. ISG NIVEL CBS.
- Boshuizen, H., & Van de Water, H. (1994). *An international comparison of health expectancies* (No. TNO-PG 94.046). TNO Preventie en Gezondheid.
- Crimmins, E. M., Saito, Y., & Hayward, M. D. (1993). Sullivan and multistate methods of estimating active life expectancy: two methods, two answers. In J. M. Robine, C. D. Mathers, M. R. Bone, & I. Romieu (Ed.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*(pp. 155-160). London: John Libbey.
- Drummond, M. F., Stoddart, G. L., & Torrance, G. W. (1990). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press.
- Koopmanschap, M. A., & Rutten, F. F. H. (1993). Indirect costs in economic studies: confronting the confusion. *PharmacoEconomics*, 6(4), 446-454.
- Murray, C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ*, 72(3), 429-45.
- Murray, C. J. (1996). Rethinking DALYs. In C. J. Murray & A. D. Lopez (Ed.), *The Global Burden of Disease*(Vol. I, pp. 1-98). Boston: Harvard School of Public Health.
- Murray, C. J. L., & Lopez, A. D. (1997). Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet*, 349, 1347-52.
- Ritchie, K. (1994). *International comparisons of dementia-free life expectancy: a critical review of the results obtained*. Paper presented at the Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.
- Robine, J. M., & Mathers, C. D. (1993). Measuring the compression or expansion of morbidity through changes in health expectancy. In J. M. Robine, C. D. Mathers, M. R. Bone, & I. Romieu (Ed.), *Calculation of health expectancies: harmonization, consensus achieved and future perspectives*(pp. 269-286). London: John Libbey.
- Robine, J. M., & Ritchie, K. (1991). Healthy life expectancy: evaluation of global indicator of change in population health. *British Medical Journal*, 302, 457-460.
- Shryock, H. S., & Siegel, J. S. (1976). *The methods and materials of demography, condensed edition*. Orlando: Academic Press, Inc.
- Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. [Review]. *J Chronic Dis*, 40(6), 593-603.
- World Bank (1993). *Investing in Health*. New York: Oxford Univ Press.

Health expectancy, an indicator for change?

Abstract

Study objective Health expectancy is an increasingly used indicator of population health status. It collapses both mortality and morbidity into a single indicator, and is therefore preferred to the total life expectancy for populations with low mortality but high morbidity rates. Three methods of calculation exist: the Sullivan, double decrement, and multi-state methods. This report aims to describe their relative advantages and limitations when used to monitor changes in population health status over time.

Design The differences between the three methods are explained. Using a dynamic model of heart disease, the effect of the introduction of thrombolytic treatment on the survival of patients with acute myocardial infarction is calculated. The resulting changes in health expectancy are calculated according to the Sullivan and multi-state methods.

Main results As opposed to the double decrement and the multi-state method, the Sullivan method produces spurious trends in health expectancy in response to the change in survival.

Conclusions Estimates of health expectancy in a dynamic situation can be very misleading when based on the Sullivan method, with its attractively moderate data requirements. The multi-state method, which requires longitudinal studies of population health status, is often indispensable.

Introduction

In an article published in the *Journal of Epidemiology and Community Health* of December 1992 Margaret Bone evaluated the international efforts to measure the health expectancy as an index of a population's state of

health, and its application as an indicator of changes in population health.(1) For several years now an international network of researchers has been working to clear up methodological issues, to standardize methods, to encourage the collection of appropriate data, and to gain acceptance for the indicator in health policy.(2)

The health expectancy (or healthy life expectancy) is derived from both mortality and morbidity, and indicates which part of the total life expectancy is spent in good health. Time trends in health expectancy help to determine whether we are improving the nations health, or that we are just being more successful in preventing severely ill people from dying. This property makes health expectancy an important index in the ongoing debate, sparked off by Fries, on the compression or expansion of morbidity.(3)

There is no doubt that the success of the health expectancy indicator is largely due to its intuitive appeal. It seems a straightforward extension of the notion of life expectancy, and is generally interpreted as the average number of years a newborn will live without (serious) disease. It also has attractive flexibility, depending on the definition of "healthy": there have been disease free, disability free, and quality adjusted life expectancy indicators.(2) As with life expectancy itself, however, there is more to the health expectancy indicator than meets the eye.

For one thing, there are three different methods of calculating the health expectancy: the Sullivan, double decrement, and multi-state methods.(4) Each has different data requirements, and produces different results. Most of the studies cited above employ Sullivan's method, named after the researcher that pioneered its use.(5) The popularity of this method has a good reason: it is the least demanding in terms of data requirements. There is a price, however: the Sullivan method gives reliable results only in a static environment.

This article considers the conditions under which simple health expectancy indicators based on Sullivan's method can be used, and those which require the more demanding multi-state method. The introduction of thrombolytic treatment in hospitals and its effect on the health expectancy is used by way of illustration.

Methods

The health expectancy indicator is known in various guises, which differ in their definition of health and how it is measured, but otherwise they are very similar. We will confine ourselves to the disease free life expectancy definition, more particularly to life expectancy free from heart disease, but the argument can be generalized to apply to the whole range of health expectancy indicators.

The health expectancy indicator is clearly an offspring of the standard life expectancy indicator, and employs the same method: a life table. In a standard life table a birth cohort of usually 100,000 people is subjected to mortality probabilities in relation to age. For each age, the total number of years the shrinking cohort has yet to live is calculated, and dividing this total

by the number of people still alive gives the life expectancy for that particular age. The Appendix (A) contains a more formal description of this calculation.

Three ways of determining the health expectancy

Three variations of this procedure may be used to calculate the health expectancy:

1. *Sullivan's method* uses disease prevalence data in relation to age to subtract the number of years the cohort still has to live with disease from the total number of years. The health expectancy is then calculated by dividing this number of healthy years by the number of people alive. As a rule the disease prevalence data are from a cross sectional survey;
2. *The double decrement method* uses disease incidence data: the birth cohort is subjected to both mortality and incidence probabilities, the former corrected for disease specific mortality. From the cohort of people who are neither dead nor ill the life expectancy is calculated according to the standard procedure;
3. *The multi-state method* also uses incidence probabilities to calculate disease prevalence, but in addition allows for one or more disease states including, when applicable, a "cured" state where the cohort may be subject to recurrent disease. It depends on the definition of "healthy" which of these states will count towards the health expectancy, but once this is decided the calculation proceeds in the standard way. In Appendix B an equation is given for health expectancy according to the Sullivan method and the multi-state method with only one disease state, and no cure or recurrent disease.

It is evident that the multi-state method is by far the most demanding in terms of data requirements. In fact only longitudinal studies with a long follow up period and a sufficient number of rounds can provide the necessary detail. Still, it is very attractive because it can capture the natural course of a disease, and can encompass patients who are cured or have intermittent disease free periods.

The prevalence data used in Sullivan's method reflect implicitly this natural course too, but in a very complex way. Prevalence is a stock variable: current prevalence of spinal injuries among 40 year old (former) car drivers covers an accumulated 22 years of car accidents that happened to drivers of as many different ages. As is shown in Appendix C current cross sectional prevalence is a function of a long series of past incidence and mortality rates (and cure rates, when applicable).

Incidence and mortality are flow variables: current acute mortality from spinal injuries among 40 year olds reflects this years' car accidents to drivers of that age only. While the multi-state method uses only flow variables, the Sullivan method employs the flow variable "mortality" to calculate total number of years lived, and the stock variable "prevalence" for the diseased number of years. The inconsistent mix of stock and flow variables may lead to odd results.

The double decrement life table, like the multi-state, uses only flow variables, is therefore consistent, and has much less demanding data requirements. The drawback is that it, in effect, treats disease incidence the same as mortality.

Comparing Sullivan and multi-state methods

To illustrate our argument we shall compare results from the Sullivan method with those from the multi-state method. In order to ensure comparability with the Sullivan method we use the same incidence data and combine these with survival data to calculate a prevalence. In both cases, the possibility of cure is ignored - that is, all patients who have had a myocardial infarction will contribute to the prevalence of heart disease for the rest of their lives.

The model

We use a dynamic population model for ischaemic heart disease to compare the two methods. It is a state-transition model, that owes much to the Coronary Heart Disease Policy Model of Weinstein *et al.*(6) After patients enter the disease model with a first manifestation of heart disease - angina pectoris, an acute coronary event, or heart failure - they are subject to risks of (possibly repeated) events such as an operation or another acute coronary event. Depending on their state when an event occurs, they may be referred to a new state, they may die, or they may remain where they are.

Input includes the incidence of myocardial infarctions in relation to age, based on the nationwide Dutch hospital register.(7) Survival after admission to hospital with myocardial infarction is based on the same register and, for the long term, on the results of the Framingham Study. Combined with an "all other causes" mortality we were able to reproduce very well the observed mortality from myocardial infarction in the Netherlands.(8)

The model is dynamic in the sense that prevalence in each state depends on the prevalence in the previous time period, and on inflow and outflow variables. These inflow and outflow variables can be manipulated to simulate changes in incidence and survival over time. Output options include disease free life expectancies according to the Sullivan method (using the dynamically calculated prevalence) and the multi-state method (using a synthetic prevalence based only on current incidence and survival).

Effect of thrombolysis

For our calculations we use the hypothetical example of the simultaneous introduction of thrombolytic treatment in all Dutch hospitals. The assumption is that all hospitals introduced thrombolysis in 1991, and that none used it before that. We assumed a conservative, 25% reduction in acute, in-hospital deaths from myocardial infarction after thrombolysis, and calculated long term effects on (healthy) life expectancy keeping incidence and all other survival parameters constant.(9)

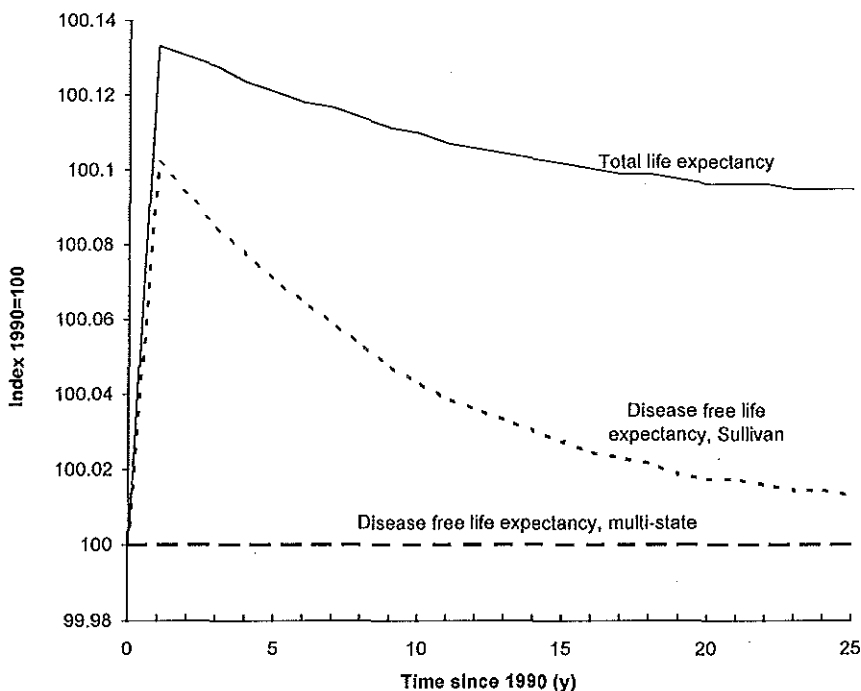


Figure - Total and disease free life expectancy after adoption of thrombolytic treatment in men. Multi-state method versus Sullivan's method.

Results

What can we expect from this intervention? Since thrombolysis has an effect on acute deaths only we expect the total life expectancy to increase suddenly between 1990 to 1991. Since the incidence of myocardial infarction is kept constant, and because all people are considered "heart patients" after a myocardial infarct, the life expectancy free from heart disease should remain unchanged.

From equations (3) and (5) we can see how a change in disease specific mortality will, through the concomitant change in total mortality, instantaneously affect the total number of years lived by the synthetic cohort, but will only gradually be reflected in population prevalence. This results in an abrupt change in the Sullivan health expectancy (up when mortality goes down, and vice versa), followed by a slow approach to the correct value. The synthetic prevalence of the multi-state method (eq. 7) is, however, like the total number of years lived, adjusted instantly.

In the figure we present the results from a 25 year simulation with the model. Total life expectancy increases suddenly, but tapers off somewhat afterwards as the now surviving patients drift into chronic heart disease and

its high mortality. The disease free life expectancy according to the multi-state method remains unchanged, as expected. The health expectancy according to the Sullivan method, however, shows a quite different pattern: it rises initially almost as much as total life expectancy, then starts falling and reaches the multi-state line asymptotically after about 40 years.

Here the Sullivan method shows the disadvantage of mixing stocks and flows. The constant incidence in combination with the lower mortality will, in the end, necessarily produce a higher prevalence. But the prevalence at any age is a function of incidences and mortalities at all lower ages. After a change in incidence or mortality, the prevalence in a population will have reached its new equilibrium value only when all cohorts dating from before the change are extinct.

While the prevalence, and consequently the number of years lived with disease according to the Sullivan method, is still catching up, the lower mortality has already raised the total number of years lived by the life table cohort. Since the health expectancy is calculated using the difference of total and diseased years lived, the fast adjusting total life expectancy initially pulls up the slowly adjusting Sullivan health expectancy.

The Sullivan indicator shows an increase followed by a decrease when it ought to remain unchanged. Only when the prevalences have settled down to the new equilibrium does the Sullivan disease free life expectancy indicator produce no spurious trends. In Appendix D we derive how many years it takes for the Sullivan health expectancy to reach this equilibrium: this is determined by the difference between the highest age in the life table considered (typically something like 95 or so), and the lowest age at which the disease under consideration becomes important (for chronic diseases typically somewhere between 20 and 60). So it turns out that incidence, mortality (and cure, if included) must have been constant for a period from 35 to 75 years, depending on the disease.

Discussion

The results show that the health expectancy according to Sullivan's method can produce misleading results when dynamic effects are present. It is true that our example is geared to show this: a disease free life expectancy indicator of a disease that affects people rather early in life, yet has a relatively good survival, will be much more "off track" than, for instance, an indicator of life expectancy without terminal disease.

We can see from equations (4) and (5) in the appendix which disease characteristics will reduce the deviation from the correct value. The less current prevalence is influenced by past prevalence, the sooner the Sullivan health expectancy will have approached the right value. This occurs, for example, when mortality is high or incidence increases sharply with age.

Current heart disease prevalence is dependent on past prevalence, and heart disease is therefore a good example to provide the clarification the issue apparently still needs. Bone attributes the call for the use of longitudinal data

to the need to take into account explicitly reversals in disability.(1) That is an important point, but, as shown above, not the whole story. Robine and Ritchie are aware of the problem of interpreting a time series of health expectancies according to Sullivan's method, but they contribute this to the combination of period data (current mortality probabilities) with cohort data, which they say currently observed disability prevalence is.(10) But observed disability (or disease) prevalence is just as much period data as observed mortality probabilities: both are influenced by the history of the cohorts that make up the current population, prevalence only more so. Therefore their suggested remedy, to derive disability free life expectancy from period data only, will not work. The problem originates from combining stock and flow, not period and cohort data.

In a more recent article Robine *et al.* point out that health expectancies based on a combination of stock and flow variables pose problems when making between country comparisons.(11) That is undoubtedly true, as the health expectancy is meant to estimate the impact of *current* health risks, not so much the legacy of the past. Although the authors discuss the development of health expectancy over time, however, no mention is made that similar stock and flow problems occur when estimating trends.

It could be argued that the example of the sudden introduction of thrombolytic treatment is a theoretical and disruptive event, while things do not normally show such dramatic dynamics. But while its introduction in clinical practice admittedly will not have happened overnight, an English study shows that thrombolysis took about three years to go from practically zero to a new treatment plateau - long perhaps from a clinical point of view, but short compared with the kind of time lags associated with the Sullivan method (see appendix D).(12)

In addition, although total mortality develops rather smoothly over time, disease specific mortalities show diverse and much more dynamic patterns.(13) For example, male mortality from ischaemic heart disease in the Netherlands, standardized for population structure and indexed for 1950=100, reached a high of 301 in 1972, and had fallen to 191 in 1990 (with some allowance for codification changes over the period).(14) Presumably the underlying morbidity from ischaemic heart disease will not have stayed unchanged either. Given the very long time lags involved, such changes will surely bias Sullivan health expectancy trend estimates.

Most health expectancy estimates published so far are disability free life expectancies, or some comparable health measure that is the compound result of a large number of diseases.(2) Interpreting a time series of Sullivan health expectancies based on such a compound health measure becomes a daunting task indeed. The health expectancy thus measured is at any point in time a function of a large number of disease specific prevalences, each being at some unknown point on the way to a new (and probably shifting) equilibrium. To disentangle this jumble and distinguish true change from delayed adjustments seems impossible.

Recent studies report that health expectancy is increasing less than total life expectancy, or not even increasing at all.(15, 16) But the decline in mor-

tality rates in middle and old age started somewhere in the late 60s or early 70s, and given the response of the Sullivan health expectancy trend estimates to mortality declines - an initial overestimate of health expectancy followed by a long term decline towards the correct value - these estimates will still display a downward trend as a reaction to previous declines in mortality. In combination with current mortality declines the observed trend in Sullivan health expectancies may go either way, but provides no useful information. The observed stagnation of health expectancy, that has accompanied recent increases in total life expectancy may, therefore, very well be an artefact of the Sullivan method.

The double decrement method provides no viable alternative. Although it is methodologically sound, it is far too crude for monitoring population health status by equating disease incidence with death. This means that it is insensitive to changes in disease prevalence or severity that originate from new or better treatment that patients may benefit from or even be cured by.

This leaves the multi-state method. It is consistent because it calculates disease prevalence using only current flow variables, which eliminates the problem of an inherited stock of patients that plagues Sullivan's method. It is potentially subtle enough to track the changes in health status after, for instance, the introduction of thrombolysis, that will not just prevent a number of deaths, but will also limit the damage to the heart for a number of patients who would have lived anyway. To use the multi-state method to its full potential, however, detailed longitudinal data are needed, also on the disease specific level. The kind of model we have employed in this study also gives a better understanding of the dynamics of population health status, and it has similar data requirements. Only when we can monitor population health status *and* explain the direction it is moving in, can we understand what we are monitoring and make informed suggestions about improvement.

Sullivan's method was intended to give an estimate of the health expectancy, using readily available or easily obtained data. In a population with an increasing total life expectancy it will be biased upward, but this may be a disadvantage well worth putting up with. The method should not be pushed beyond this goal, however, because what is a small bias in health expectancy will confound a trend analysis of a time series of health expectancies since the direction and size of the bias are time dependent. So the question "Is health expectancy is a valuable indicator for changes in population health" deserves a conditional "Yes" - the condition being longitudinal studies, that can provide the input for multi-state life tables and dynamic models.

The short cut Sullivan's method provides, as compared to the multi-state method, is a dead end when it comes to the analysis of changes in health expectancy over time.

Appendix

A: Life expectancy

Life expectancy is calculated by submitting a so called "synthetic" birth cohort of 100,000 people to age specific mortality probabilities. The birth co-

hort is called "synthetic" because, and this is important to note, the mortality probabilities used are the current probabilities. A real birth cohort has been submitted throughout life to mortality probabilities in the past that were, as a rule, higher than the current ones, and a cohort of newly borns will face future probabilities that we expect to be lower yet. Because of these changes in mortality, the current life expectancy at age 0 will not be equal to the average number of years to be lived by a cohort of newborns.

If we denote the synthetic cohort with l_a , with a the age index and a_{max} the highest age considered, and the mortality probabilities with q_a , then the life expectancy e_a is given by the total number of years to be lived by the cohort at age a divided by the number alive at that age. Formally:

$$e_a = \frac{\sum_a^{a_{max}-1} 0.5(l_a + l_{a+1}) + l_{a_{max}} e_{a_{max}}}{l_a} \quad 1$$

Using $l_{a+1} = l_a(1 - q_a)$ we can rewrite (1) to:

$$e_a = \frac{\sum_a^{a_{max}-1} l_a(1 - 0.5q_a) + l_{a_{max}} e_{a_{max}}}{l_a} \quad 2$$

Life expectancy is a static estimator: because $l_{a+1} = l_a(1 - q_a)$ for all a any change in the current mortality q_a is instantly reflected in the total number of years lived to its full effect. In fact, the life expectancy estimator assumes that the mortality probabilities have been unchanged for at least a_{max} years.

B: Health expectancy

Health expectancy is calculated by not letting years defined as unhealthy count towards the total number of years to be lived. When P_a is the age specific prevalence of unhealthiness (expressed as a proportion) then the health expectancy h_a can be calculated by:

$$h_a = \frac{\sum_a^{a_{max}-1} l_a(1 - 0.5q_a)(1 - P_a) + l_{a_{max}} e_{a_{max}}(1 - P_{a_{max}})}{l_a} \quad 3$$

Equation (3) can be used for both the Sullivan and the multi-state method, the difference is where the prevalence P_a comes from.

C: Empirical versus synthetic prevalence

The Sullivan method uses empirical prevalence data obtained through a survey or some similar method, while the multi-state method uses the prevalence calculated in the life table itself from the current incidence and mortality probabilities. The cross sectional prevalence for a very simple disease process with incidence I_a and mortality M_a and no cure is given by:

$$P_a^t = (P_{a-1}^{t-1} + I_{a-1}^{t-1})(1 - M_{a-1}^{t-1}) \quad 4$$

Applying (4) recursively to itself, and assuming that prevalence at birth equals zero, we can rewrite (4) as follows:

$$P_a^t = \sum_{j=0}^{a-1} (I_j^{t-a+j} \prod_{k=j}^{a-1} (1 - M_k^{t-a+k})) \quad 5$$

The prevalence as given by (5) is used by the Sullivan method, for the multi-state method the following equations apply:

$$P_a^t = (P_{a-1}^t + I_{a-1}^t)(1 - M_{a-1}^t) \quad 6$$

$$P_a^t = \sum_{j=0}^{a-1} (I_j^t \prod_{k=j}^{a-1} (1 - M_k^t)) \quad 7$$

This multi-state disease prevalence one might call, as an analogue to the life table cohort, a "synthetic prevalence". This synthetic prevalence is, unlike a cross-sectional prevalence, not a stock variable, because, being a function of current incidence and mortality only, it does not depend on past values.

D: Time lags

From (5) and (7) it can be deduced under which conditions the Sullivan and multi-state disease prevalences at age a and time t will be equal:

$$I_j^{t-a+j} = I_j^t \quad \forall j \in [0..a-1]$$

$$M_j^{t-a+j} = M_j^t \quad \forall j \in [0..a-1]$$

To put it differently: past age specific incidence and mortality must have been equal to the current incidence and mortality for, depending on age, up to a years. When a_{min} stands for the lowest age at which incidence occurs this requirement relaxes to $a - a_{min}$ years. Because the calculation of the health expectancy always uses the prevalence of the highest age a_{max} (see eq. 3) this implies that the Sullivan health expectancy equals the multi-state health expectancy when age specific incidence and mortality have been constant for up to $a_{max} - a_{min}$ years. The same applies for cure probabilities, if included.

References

1. Bone MR. International efforts to measure health expectancy. *Journal of Epidemiology and Community Health* 1992;46:555.
2. Mathers CD, Robine JM. Health expectancy indicators: a review of the work of REVES to date. 6th Work-group meeting REVES, International Research Network for Interpretation of Observed Values of Healthy Life Expectancy. Montpellier: REVES, 1992: 116.

3. Fries JF. Aging, natural death, and the compression of morbidity. *N Engl J Med* 1980;303:130.
4. Robine JM, Romieu I, Mathers CD. Multiplication of health expectancy calculations and international comparison problems. WHO/CBS Consultation to develop common methods and instruments for health interview surveys: third consultation. Amsterdam, 1992.
5. Sullivan D. A single index of mortality and morbidity. Health Services and Mental Health Administration (HSMHA) Health Reports 1971;86:347.
6. 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. *Am J Pub Health* 1987;77:1417.
7. LMR (Information Center for Health Care). Files from the National Medical Registration concerning admissions and intervention in 1988. Utrecht: 1990.
8. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J Pub Health* 1994; 84:20.
9. Gruppo Italiano per lo Studio della Streptochinasi Nell Infarto Miocardico (GISSI). Long-term effects of intravenous thrombolysis in acute myocardial infarction: final report of the GISSI Study. *Lancet* 1987;1:871.
10. Robine JM, Ritchie K. Healthy life expectancy: evaluation of global indicator of change in population health. *Br Med J* 1991;302:457.
11. Robine JM, Michel JP, Branch LG. Measurement and utilization of healthy life expectancy: conceptual issues. *Bulletin of the World Health Organization* 1992;70(6):791.
12. Ketley D, Woods KL. Impact of clinical trials on clinical practice: example of thrombolysis for acute myocardial infarction. *Lancet* 1993;342(October 9):891.
13. Mackenbach JP. Mortality and medical care. Rotterdam: Erasmus University Rotterdam, 1988. PhD Thesis.
14. Centraal Bureau voor de Statistiek. Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1. Voorburg, Published yearly.
15. Crimmins E, Sato Y, Ingegneri D. Changes in life expectancy and disability-free life expectancy in the United States. *Pop Develop Rev* 1989;15:235.
16. Mathers CD. Disability-free and handicap-free life expectancy in Australia. Health differentials series no. 1. Canberra: Australian Institute of Health, 1990.

Changes in incidence and survival of cardiovascular disease and their impact on disease prevalence and health expectancy

Abstract

Background. Mortality and incidence rates for heart disease and stroke are declining . All other things being equal this obviously will increase life expectancy, but what happens to disease prevalence depends on the balance between incidence and survival: it may increase, decrease, or remain unchanged. The health expectancy, in turn, depends on the balance between life expectancy and disease prevalence.

Method. A multi-state life table is used to calculate the impact of various combinations of changes in incidence and survival of heart disease and stroke on disease prevalence, life expectancy, and life expectancy without disease.

Results

- *A better survival is not reflected in the health expectancy. Health expectancy at birth will remain unchanged, and at higher ages it will decline. This is a counter-intuitive result, that could (and should) be remedied by giving life years with disease a weight greater than 0.*
- *Both lower incidence and better survival have the effect of relatively shifting the burden of disease from younger ages to older ages. Health expectancy at birth does not reflect such distributional shifts.*
- *The most likely future changes (judging from the recent past) in both heart disease and stroke are a lower incidence and a better survival, but the latter outrunning the first. Results are decreased prevalences*

at younger ages, together with increased prevalences at older ages ones, and especially increased comorbidity. This rightly shows up as an increased health expectancy, but in as far as prevalence is a good indicator it also implies a greater need for care.

Conclusions

- *Health expectancy would be a better indicator for public health if diseased (or disabled) life-years were weighted instead of ignored. This can be done, as the World Bank 1993 report has shown. Another way is using quality adjusted life years.*
- *Policy makers will be interested in the age distribution of effects from interventions. Unfortunately health expectancy at birth is rather insensitive to such distributional changes, and while solutions are conceivable they have the drawback of diminished intuitiveness.*
- *When health expectancy is used in a policy making environment it should be made clear that a higher health expectancy could very well imply higher demand for care.*

Introduction

Health expectancy (or healthy life expectancy) is preferred to life expectancy as an indicator for population health status because it is based not just on mortality, but also includes morbidity. As soon as a country has moved through the epidemiologic transition, that is when the burden of illness has shifted from mostly infectious, acutely fatal disease towards mostly non-communicable, chronic disease, it becomes highly desirable to adopt an estimator of public health that puts less emphasis on (child) mortality and more on (all age) morbidity than life expectancy does.

The work of REVES on the health expectancy indicator - clearing up methodological issues, standardizing methods, encouraging data collection, and gaining acceptance for the indicator in health policy - has met with considerable success.(1) For example, in the Netherlands a recently published comprehensive overview of the health status of the Dutch population used health expectancy as the central summary indicator.(2)

The success of the health expectancy indicator is in large part due, no doubt, to its intuitive appeal. It seems a straightforward extension of the notion of life expectancy, and tends to be interpreted as the number of years someone can expect to live without (serious) disease, followed by a number of unhealthy years until death. This intuitive appeal makes health expectancy an attractive instrument to communicate results from research to policy makers and the general public.

The ability to communicate results to people outside the small circle of directly involved researchers is an important asset. In this paper we investigate the behavior of health expectancy in reaction to a few hypothetical, but in view of the recent past not unlikely developments in the epidemiology of

cardiovascular disease.(3, 4) We look at the results of declining incidence and increasing survival in terms of changes in disease prevalence and health expectancy, and judge the information health expectancy provides on both intuitiveness and usefulness.

Methods

We set up a multi-state life table with two diseases, heart disease (consisting of ischemic heart disease and heart failure), and stroke. States are: neither heart disease nor stroke, heart disease, stroke, comorbidity of heart disease and stroke, and the three absorbing states death from heart disease, stroke and 'all other causes'. Inflow in the disease states depends on incidence probabilities by age, outflow on survival probabilities, also by age, cure is ignored. In all but the absorbing states people are subject to 'all other causes' mortality probabilities, that are independent from disease-specific survivals. Co-morbidity is calculated assuming disease independence.

We use a subset of the input data collected for the stroke and heart disease models that are part of NIMPH: the Netherlands Integrated Model of Public Health. They consist of total mortality probability and disease-specific mortality rates from the Central Bureau of Statistics (5, 6). The incidence probabilities are based on nation-wide hospital register data and regional studies.(3, 4) From incidence and disease specific and total mortality we calculate age- and disease-specific survival probabilities.(7)

We vary incidence of both diseases by multiplying age-specific probabilities with a factor of 0.9 (all ages), and multiply probabilities to die from a specific disease, given prevalence (i.e. 1-survival probability) likewise with 0.8, always comparing outcome with the 1988 base year. We look at age-specific point prevalence and disease-free life expectancy.

Results

Increased survival

First we increase survival with heart disease, while leaving all other parameters at their base value. Figure 1 shows outcome expressed as indices (with 1988=100) of prevalences of heart disease and heart disease-free life expectancy by age. Heart disease prevalence rate increases for all age groups after increased survival, but much more so in older age groups: a relative shift of the burden of disease towards older ages. In numbers this shift is still stronger because a larger part of the initial birth cohort lives to high age.

The health expectancy at birth (and up to about 40) remains constant, as one would like to see from a disease-free life expectancy with disease incidence and mortality from all other causes constant. But for older ages it increasingly declines. This is of course as it should be: the number of years lived without disease in each age interval remains constant, but the older age

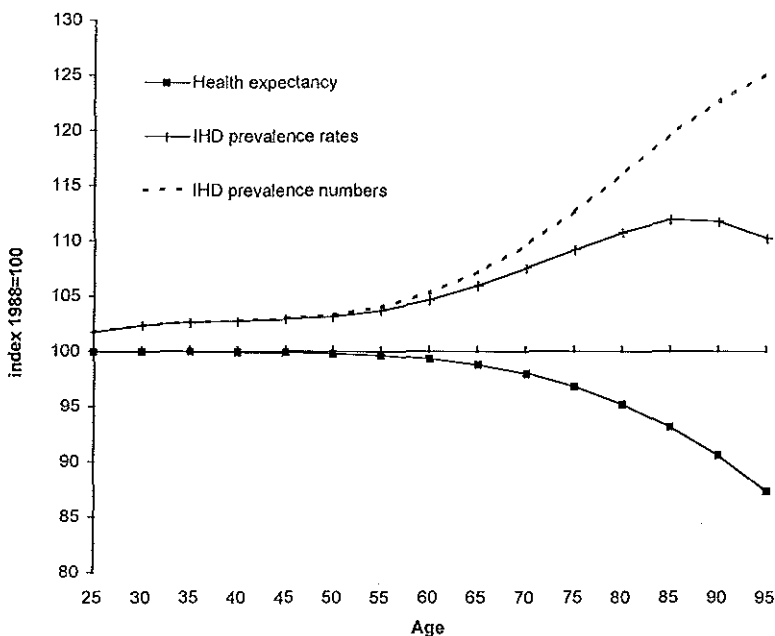


Figure 1 - Health Expectancy and heart disease prevalence, index 1988=100, by age, with constant incidence and increased survival. IHD: heart disease.

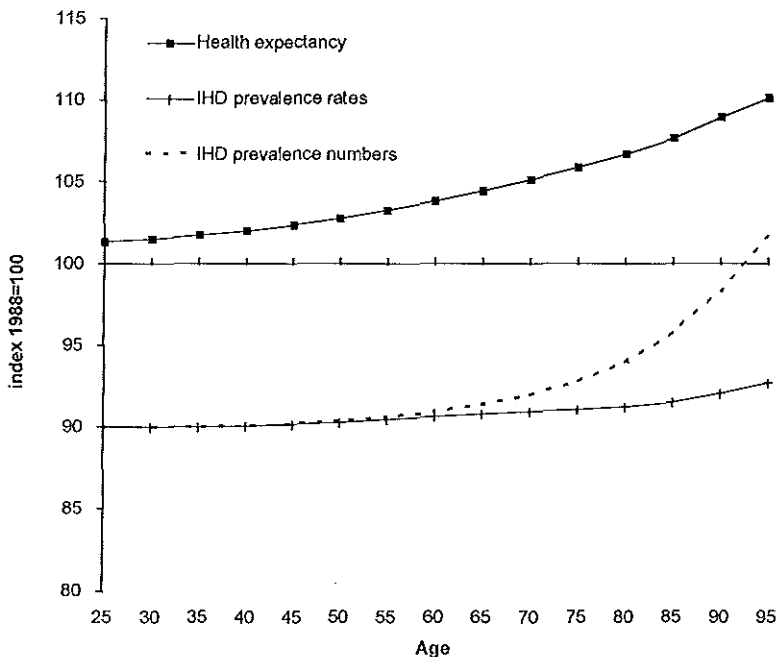


Figure 2 - Health Expectancy and heart disease prevalence, Index 1988=100, by age, with decreasing incidence and constant survival. IHD: heart disease.

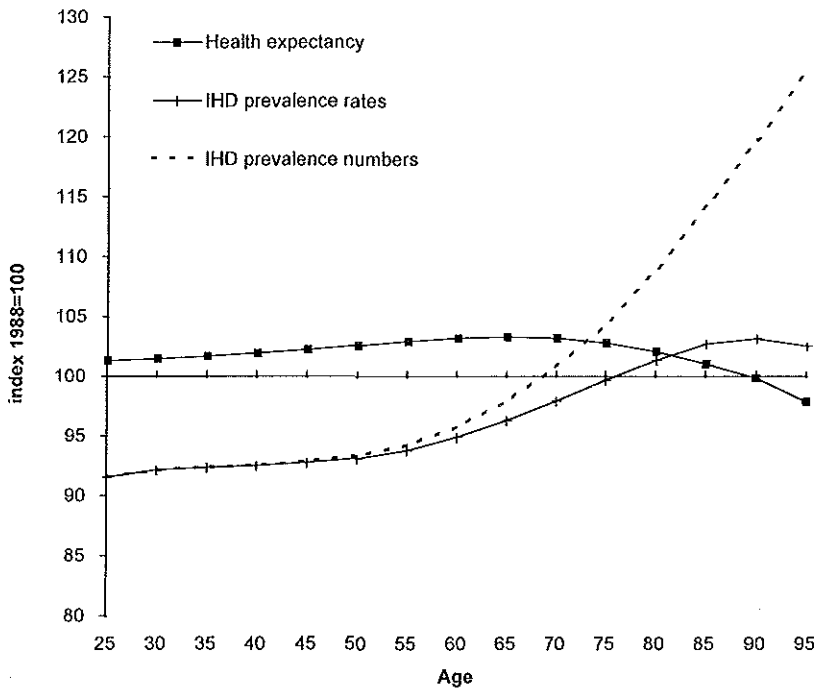


Figure 3 - Health Expectancy and heart disease prevalence, index 1988=100, by age, with decreasing incidence and increased survival. IHD: heart disease.

groups become larger due to the declined mortality. Because the increase is solely in the diseased state the average number of years lived without disease declines. Computationally sound as this may be, it is also highly counter-intuitive that health expectancy at age 65 or so declines when survival with heart disease increases.

Decreased incidence

Secondly we multiplied age-specific incidence with 0.9, again with all other parameters at their base value (Figure 2). The age-specific health expectancy resulting from this behaves like one would expect: higher for all ages, but more so for the elderly where incidence rates are higher. More interesting is age-specific prevalence: while at the lowest ages the prevalence rate stands at a corresponding 90% of the base rate, this percentage edges up a bit with age. The lower prevalence rate in any age group results in a larger group at risk for incidence, which counteracts the lower incidence probability.

As with the increased survival, this again results in a relative shift of the burden of disease towards older age groups. Through the implied lower disease-specific mortality, this shift is more pronounced in prevalence numbers: at the highest age this prevalence even exceeds the base value.

Increased survival and decreased incidence combined

The combined picture of better survival and lower incidence is in Figure 3. Lower incidence is reflected in 5 to 10% lower prevalences (numbers) at young ages, and the increased survival causes up to 25% higher prevalences at older ages. The result is a tilted index line, with a pivot point somewhere between age 65 and 70.

Age-specific health expectancy is surprisingly unaffected by this: it stays within 3.5% of the baseline value, rising until 65 and declining after. The counteracting influences from increased survival and decreased incidence more or less cancel each other out.

Co-morbidity

For stroke prevalence a similar pattern applies as for heart disease: lower incidence and better survival results in a tilted index line with increased prevalence at higher ages. The age-specific co-morbidity of heart disease and stroke, that under the assumption of independence of the diseases is simply the product of the two prevalences (expressed as proportions), shows the same tilted index line, but then in exaggerated form. At higher ages the co-morbidity is about 50% higher than the base line value.

Discussion

Health weights

Health expectancy is one of several efforts to measure the health status of a population by collapsing information on mortality and morbidity into one figure. Other ways include quality-adjusted life-years (QALYs), utility-based life-years, disability-adjusted life-years (DALYs). All these indicators have one point in common: a life-year spend with disease is valued differently from one in good health. The main differences between health expectancy and the other indicators are that firstly, for health expectancy the diseased life-years as a rule get a weight of 0, while the others use something between 0 and 1, and secondly, that health expectancy uses the weights to calculate a healthy life expectancy, and the others as a rule do not.

Another concern about these methods is that there is very little contact between them. For example, the World Bank reports total disability-adjusted life-years lost, an outcome that is very difficult to interpret and makes comparisons between countries impossible, as they admit, while all necessary ingredients for a disability-adjusted life expectancy were available.(8) With some notable exceptions researchers seem happy to pursue their own ways and tend to stress the differences between the methods rather than learn from others and apply what is useful.(9)

The above-reported effect of increased survival on health expectancy illustrates the disadvantage of this tunnel vision. When life-years lived with disease are weighted 0 any increase in survival will have the result shown

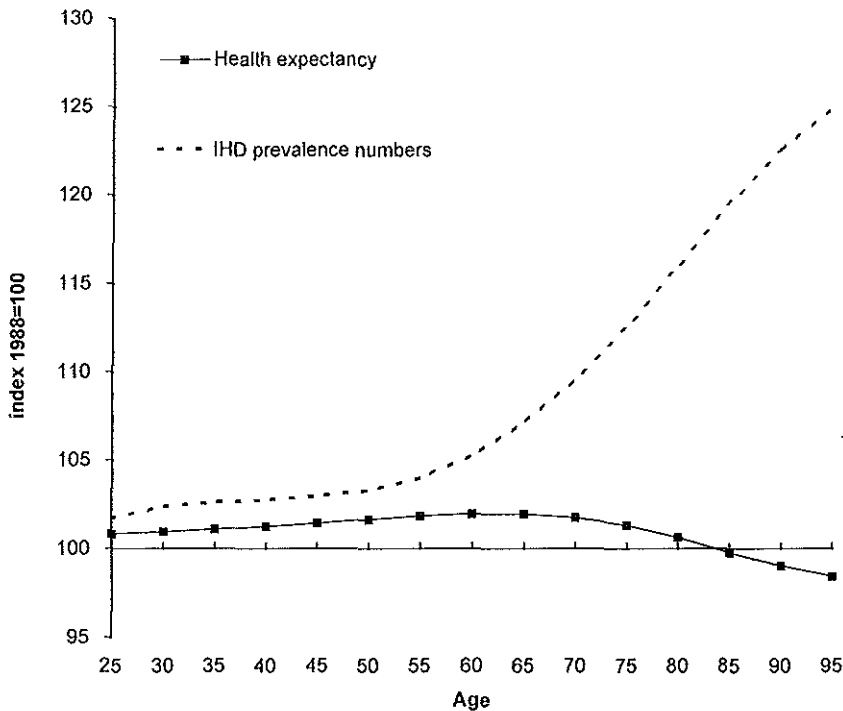


Figure 4 - Health Expectancy (unhealthy years weighted) and heart disease prevalence, index 1988=100, by age, with constant incidence and increased survival. IHD: heart disease.

above: an unchanged health expectancy at birth, but a decreasing one at higher ages. This means that medical interventions in chronic diseases, by definition not able to cure the patients, but life extending and/or quality enhancing, are, as measured by the health expectancy, at best worthless, or even detrimental. For a health status indicator meant to address the problems of measuring health in populations with a high prevalence of chronic disease this seems a rather peculiar characteristic.

There are two ways around the problem. One is to use multi-state life table techniques to look at the age-specific health expectancy of the healthy population at that age. This will at least avoid the decrease of the health expectancy at higher ages, although it will still render any increase in survival worthless. And of course it requires better data than the Sullivan method.

Another way is to use weights. When diseased life-years are weighted more than 0 an increase in survival will show up as a higher health expectancy at birth, a gain that increases up to some age, then starts to diminish and eventually may become a decreased health expectancy for the highest age group (see Figure 4). This outcome much better reflects the population health status after an increased survival: it is an improvement (unless the gained life-

years are so awful they deserve a weight of zero) with the disadvantage that it causes higher disease prevalence, especially at high ages.

Which set of weights is used seems at this point rather immaterial: almost anything is better than the 0-1 scheme routinely employed in health expectancy calculations. At the previous REVES meeting an example of a quality adjusted life expectancy was presented, and the World Bank's DALYs will probably be used to calculate disability adjusted life expectancies.(10, 11) Such cross-fertilization is bound to produce a better health expectancy indicator.

Age distribution

Health expectancy is determined in large part by morbidity and mortality from chronic diseases, so changes in the epidemiology of chronic diseases should ideally have an impact on health expectancy. Cardiovascular disease in the Netherlands (as in many other western countries) has seen declining age-specific mortality, for stroke at least since the fifties, for heart disease since the early seventies. A closer look at the available evidence reveals that this decline is most likely due to a combination of declining incidence and increasing survival.(3, 4)

Declining incidence may be unique to cardiovascular disease, cancer incidence, for example, has been reported up (albeit based on inconclusive evidence).(12, 13) But increased survival probably extends to many other chronic conditions: medical technology has seen more progress in the management of chronic disease than in the cure. Barring major breakthroughs this pattern is likely to continue.

The effect of decreased incidence and increased survival on disease prevalence has been demonstrated above: a relative shift of the burden of disease towards older ages. It has also been demonstrated that the health expectancy indicator is rather unresponsive to such shifts in disease distribution. There are two points we would like to make here:

- If this pattern of a decreasing prevalence at younger and an increasing one at older ages will indeed turn out to be the prevailing development, it would seem that the debate on compression versus expansion of morbidity is being held in too simplistic terms. While morbidity is compressing in the sense that there is less of it at younger age, it is at the same time expanding in the sense that there is more of it at older age. Perhaps we should rethink our terminology.
- Secondly, the unresponsiveness of the health expectancy indicator to such developments is worrying. Most likely there is a way to amend health expectancy to reflect an age distribution shift (like reporting two health expectancies: one at birth until age 75, and one at age 75), but the drawback is diminished intuitiveness. We have seen the same in the compression-expansion debate: the niceties of relative compression and absolute expansion are probably lost to all but the inner circle of debaters. There may be no easy solution here.

Health expectancy and policy making

Health policy has become a major issue in many countries, mostly because of ever increasing health care costs. The health expectancy indicator is likely to become more important for policy making, because it gives a more balanced view of population health status than life expectancy, without getting lost in a myriad of details. This increased use is no doubt in large part due to the work of REVES, but, as we have tried to show here and elsewhere, there is no reason for complacency yet.⁽¹⁴⁾

Use of the health expectancy indicator for policy making is not without danger of course. Like any summary indicator it glosses over a lot of detail, some of it of central interest to the policy maker. For example, if the age distribution shift of disease described above will indeed materialize health care costs will increase because of increased numbers of patients, who are older on average, and have more comorbidity. It may come as a nasty surprise to the policy maker that a virtually unchanged health expectancy will be accompanied by increasing costs.

It should be the responsibility of REVES now that health expectancy has become respectable to further improve on the indicator and to guard against inappropriate and uninformed use.

References

1. Mathers CD, Robine JM. Health expectancy indicators: a review of the work of REVES to date. 6th Work-group meeting REVES, International Research Network for Interpretation of Observed Values of Healthy Life Expectancy. Montpellier: REVES, 1992: 116.
2. RIVM. Volksgezondheid Toekomst Verkenning: de gezondheidstoestand van de Nederlandse bevolking in de periode 1950-2010. Den Haag: Sdu Uitgeverij, 1993.
3. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke Trends in an Aging Population. *Stroke* 1993;24:931.
4. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure: the future rise of heart failure. *Am J of Public Health* 1994;84:20.
5. Centraal Bureau voor de Statistiek. Overlevingstafel naar leeftijd, 1980-1984. *Maandstatistiek bevolking* 1986;(4):37.
6. Centraal Bureau voor de Statistiek. Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1. Published annually.
7. Barendregt JJ, van Oortmarsen GJ, van Hout BA, van den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Mathematical Population Studies* (in press).
8. World Bank. Investing in Health. World Development Report 1993. New York: Oxford University Press, 1993.

9. Robine J-M, Mathers CD, Bucquet D. Distinguishing Health Expectancies and Health-Adjusted Life Expectancies from Quality-Adjusted Life Years. *Am J of Public Health* 1993;83(6):797.
10. Berthelot J-M, Roberge R, Wolfson MC. The Calculation of Health-Adjusted Life Expectancy for a Canadian Province Using a Multi-Attribute Utility Function: A First Attempt. 6th REVES International Workshop. Montpellier, 1992.
11. Murray C. Personal communication.
12. Adami HO, Bergstrom R, Sparen P, Baron J. Increasing cancer risk in younger birth cohorts in Sweden. *Lancet* 1993;341:773.
13. Bonneux L, Van Oortmarsen GJ, Barendregt JJ. Increase in cancer incidence in younger birth cohorts (letter). *Lancet* 1993;341:1409.
14. Barendregt JJ, Bonneux L, Van der Maas PJ. Health Expectancy, an indicator for change? *Journal of Epidemiology and Community Health* 1994;48:482-487.

DALYs: the age-weights on balance

Abstract

In studies for the 1993 World Development Report: Investing in Health Murray et al. developed the Disability Adjusted Life Year (DALY). This article examines one particular aspect of the DALY methodology: the weighting of life years by age. For the quantitative implementation of this notion Murray proposed a general equation to weight life years by age, which specifies that years lived between the ages of 9 and 54 have a weight greater than unity, and years outside that range less than unity. The age-weighted life years are used to calculate "expected years of life lost" (EYLL). Comparison of age-weighted and un-weighted age-specific life expectancies shows that the age range which becomes more important due to weighting is not 9-54 years, but 0-27 years. This happens because the EYLL is an age-weighting system in itself, emphasizing the young. The result of piling one age-weighting systems on top of the other gives an even stronger emphasis on the young than the EYLL generates by itself. Although this is unlikely to upset the results from the Global Burden of Disease study, we do not think it desirable. And it is certainly different from what we were led to expect.

Introduction

In 1994, Murray et al. published a series of 8 articles in the *Bulletin of the World Health Organization* (1-8) on the global burden of disease and on health expenditures and intervention packages, which describe the basic studies carried out for the *World Development Report 1993: Investing in Health* (9). These articles, now also available in book format (10), give an overview of the methodology used to assess the global cost-effectiveness of specific health expenditure packages. The sheer size of the undertaking com-

mands respect. This respect increases considerably when examination of the *Bulletin* papers and background material (11), reveals the care with which this study has been carried out.

The study developed and applied several innovative concepts, one of which - the Disability Adjusted Life Year (DALY) - has attracted much attention and generated controversy. However, use of a non-disease-specific concept like a DALY is inevitable when you want to compare the burden of morbidity (and its sequelae) of different diseases. The controversy concentrates on two particular aspects of the DALY methodology, one of which we will look into - the weighting of life years by age; the other sensitive aspect is discounting of future life years (12).

We must stress that we do not consider age-weighting as such to be inappropriate or wrong; on the contrary, we think it may certainly make sense, depending on the purpose and circumstances. However, precisely because it is controversial it is of great importance to apply it meticulously. One weak point in the methodology as reported by Murray is that the age-weighting eventually produces a different result from what might be expected from the age weight function. To illustrate this we will first briefly describe the DALY methodology, and then apply it using a life-table.

The DALY methodology

The methodology assesses the impact on public health, on both morbidity and mortality, of various packages of health care that might be provided. Most health care measures are disease-specific interventions, either preventive or therapeutic, and this makes it imperative to examine the effects of interventions on a disease-specific level. To compare the impact of the intervention between diseases or health care packages it is necessary to have a non-disease-specific common denominator.

The established common denominator used to be mortality or mortality based measures like life expectancy and potential years of life lost (PYLL). Including morbidity in the public health impact means adopting a non-disease-specific measure to express the burden of diseases and their sequelae. To express the combination of mortality and morbidity in a single measure it has become a (disputed) standard to value one year lived in good health by 1, and a year lived with less than good health by a value between 0 and 1, depending on the severity of the health problem. The years not lived (lost to premature mortality) are assigned a value of 0.

Given the disease- and age-specific mortalities the "expected years of life lost" (EYLL) due to mortality can be estimated by multiplying the number of disease-specific deaths for each age by the life expectancy at that age (the so called "local life expectancy"), and then summing over all deaths. Years lost due to morbidity can be estimated by multiplying the years lived with disease by the value between 0 and 1, summing the thus-weighted years lived, and subtracting this number from the total unweighted number of years lived.

The sum of the years lost due to mortality and morbidity then gives the total burden of this particular disease.

The DALY methodology follows this basic scheme, while adding the following features of its own.

- The objective of measuring the total burden of disease from all (important) diseases combined. This requires making consistent estimates of the incidence, prevalence, and mortality of a great number of diseases, a number large enough to represent the major part of the population's health. Using an incidence, prevalence, and mortality model the researchers forced their estimates to be internally consistent, and while undoubtedly some estimates may be inaccurate (13), this is a major achievement.
- The use of one standard life-table worldwide, and not different regional life-tables. Murray argues convincingly that this is necessary to compare "like with like", and not to value a death at a given age more in a high life expectancy country than in a low one (1).
- The use of disability weights. Six classes of disability severity were distinguished; using expert panels it was decided how many people with a certain disease would as a consequence be disabled, and how these were distributed over the six classes. Needless to say, this procedure has some weaknesses, but in that respect the DALY is no exception among procedures to obtain disability weights, or equivalents such as QALYs (quality-adjusted life years) and utilities (14, 15).
- The use of age-weights. Murray cites both economic and social role arguments to justify valuing lost years of life due to mortality and morbidity differently by age. Intuitively most people will agree that a death of a newborn or a 90-year-old, while being tragic, is less so than that of a 15-year old. The quantitative implementation of this notion is, on the other hand, a matter of much controversy.

Murray proposes the following general formula to weight life years by age:

$$Cxe^{-\beta x}$$

where C and β are constants, x is age, and e is the exponential. For $C=0.16243$ and $\beta=0.04$ (the values used in the study) the age-weight curve is as depicted in Figure 1.

- The discounting of future years. Although in economics it goes without saying that future costs and benefits should be discounted, it is not a generally accepted concept in the health sector that future years of life should be treated similarly.

In the end Murray presents a rather daunting general formula that calculates DALYs lost, combining the lost years due to mortality and morbidity, the age-weights and the discounting (1). Because the lost DALYs are calculated using this general formula, the effects of each step made to get the end result remain obscure. This is illustrated below by taking a closer look at the age-weighting and its effects.

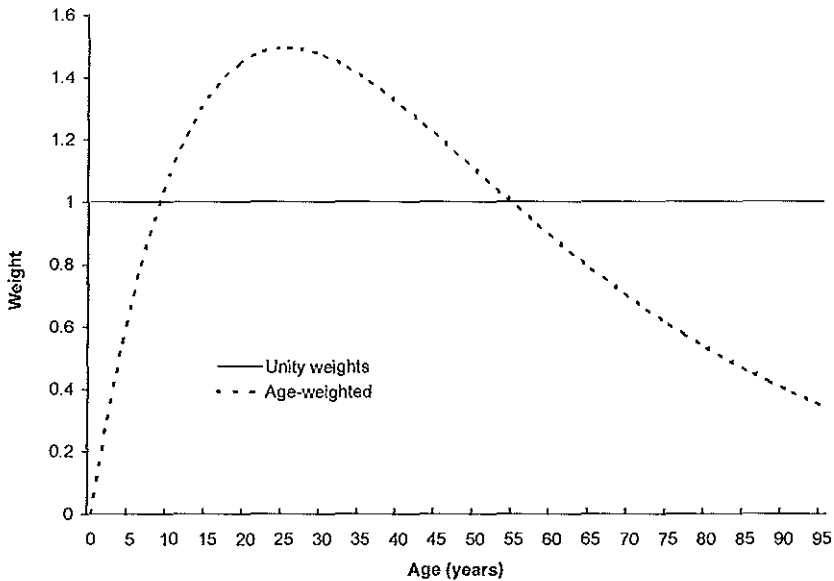


Figure 1 - The age-weight function $Cxe^{-\beta x}$ where x is the age, e the exponential, $C=0.16243$ and $\beta=0.04$.

Which ages count for more?

The age-weight function with the parameters as mentioned above specifies that years lived between the ages of 9 and 54 have a weight greater than unity, and the years outside that range will have a weight less than unity, with a weight of 0 at exact age 0 (newborns). If, for simplicity's sake, we ignore DALYs lost due to morbidity (ie the disability weights) and consider only mortality, the next step is to use the age-weighted life years to calculate age-specific life expectancies that are needed to determine the EYLL. We carried this out using the 1986-90 life-table for Dutch women (16), and Figure 2 shows the resulting age-weighted and unweighted age-specific life expectancies.

The first thing to note is that the impact of age-weighting is much less than Figure 1 would suggest: the weighted and unweighted curves are fairly close. In particular, the rather dramatic zero weight for newborns translates into an only minor downturn of the weighted curve. This is because the EYLL of a newborn reflects all the future life years the baby could expect to live, all of which are weighted more than zero.

Secondly, and importantly, the age range which becomes accentuated due to weighting (the range where the weighted curve is above the unweighted one) has shifted from between 9 and 54 to between 0 and 27 years. This shift towards younger age becomes intuitive when one considers, for exam-

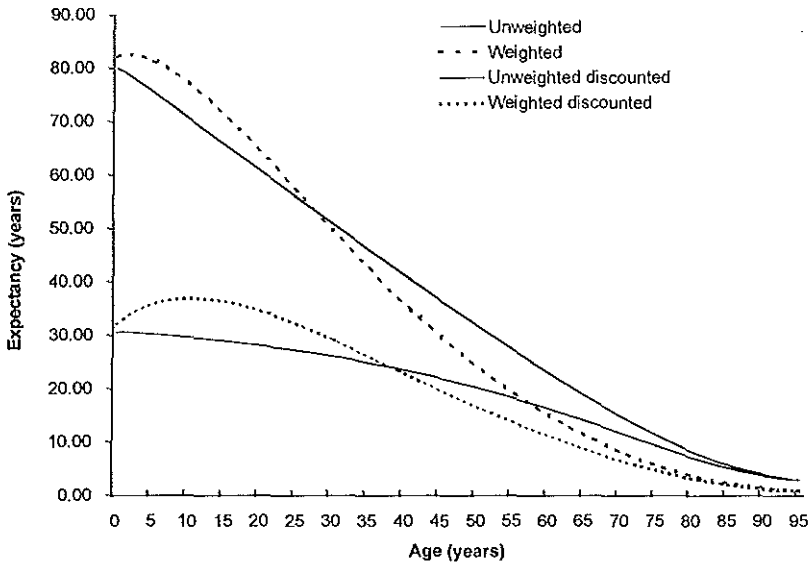


Figure 2 - Age-specific life expectancies for Dutch women, 1986-90. *Unweighted* values the years lived by 1 for all ages, *weighting* is done using the curve of Figure 1, *unweighted discounted* and *weighted discounted* use a discount rate of 3%.

ple, a woman of 40, with a life expectancy of 41 years. According to Figure 1, 14 of those expected years will receive a weight greater than unity, and the remaining 27 a weight less than unity, yielding an average weight of less than unity. Only when, going towards younger ages, the number of expected life years with a weight greater than 1 becomes large enough to counterbalance the number with a weight of less than 1 (for this life-table, at age 27, when life expectancy is about 54), the weighted curve will get above the unweighted one. While the calculation is correct, this result may come as something of a surprise.

Figure 2 also shows age-specific life expectancies with future years discounted, both with and without age-weighting. It is apparent that the impact of discounting, even at such a slight rate of 3%, is much larger than age-weighting. The curve labeled "weighted discounted" is the familiar "DALYs lost due to death" curve (1). The discounting of future years partly redresses the shift towards younger ages: the discounted age-weighted curve is above the discounted unweighted curve in the range from 0-38 years. This is because the discounting mostly affects the life years lived at a higher age, years which are now more easily counterbalanced by the more-than-unity weighted years of middle age.

Discussion

We have shown that the age-weight function emphasizes deaths in the age range 0-27, instead of the range 9-54 years as suggested by Murray. The same will happen for disability when the disability weights are applied to the years lived: disability between 0 and 27 years will be emphasized, and outside that range played down. Adding discounting displaces the emphasized age range to 0-38, but, strictly speaking, this is of no consequence because discounting is concerned with time preference, and not with age. Since Murray's arguments for age-weighting are on the one hand economic (productive years count more), and on the other social (with middle-aged years more important than the extremes), we presume that the shift towards childhood and early adult years is unintended.

It has apparently been overlooked that expressing a death (or disability) as "expected years of life lost" implies age-weighting as well: younger deaths are emphasized. Thus, in the calculation of the DALY, one age-weighting system emphasizing the young is piled on top of another, which emphasizes the middle-aged. The result is an even stronger emphasis on the young, and de-emphasis on the old, than the EYLL generates itself. Whether this is advantageous or not, is debatable (we think not), but it is certainly different from what we were led to expect.

Will our finding upset the results from the Global Burden of Disease study? That is very unlikely, given the rather small impact of the age-weight function on the age-specific life expectancies (see Figure 2). It might even be sensible to abandon the age-weighting function altogether, because of the disproportionately large attention it has received; relative to its impact. If it is retained, on the other hand, some rethinking seems appropriate. With the age-weight function as it is now, Alice's reaction to the Jabberwocky poem comes to mind: "Somehow it seems to fill my head with ideas - only I don't exactly know what they are!" (17).

References

1. Murray, C.J., Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ*, 1994. 72(3): p. 429-45.
2. Murray, C.J., R. Govindaraj, and P. Musgrove, National health expenditures: a global analysis. *Bull World Health Organ*, 1994. 72(4): p. 623-37.
3. Murray, C.J., J. Kreuser, and W. Whang, Cost-effectiveness analysis and policy choices: investing in health systems. *Bull World Health Organ*, 1994. 72(4): p. 663-74.
4. Murray, C.J. and A.D. Lopez, Global and regional cause-of-death patterns in 1990. *Bull World Health Organ*, 1994. 72(3): p. 447-80.
5. Bobadilla, J.L., et al., Design, content and financing of an essential national package of health services. *Bull World Health Organ*, 1994. 72(4): p. 653-62.

6. Murray, C.J. and A.D. Lopez, Quantifying disability: data, methods and results. *Bull World Health Organ*, 1994. 72(3): p. 481-94.
7. Michaud, C. and C.J. Murray, External assistance to the health sector in developing countries: a detailed analysis, 1972-90. *Bull World Health Organ*, 1994. 72(4): p. 639-51.
8. Murray, C.J., A.D. Lopez, and D.T. Jamison, The global burden of disease in 1990: summary results, sensitivity analysis and future directions. *Bull World Health Organ*, 1994. 72(3): p. 495-509.
9. World Bank, *Investing in Health*. World Development Report. 1993, New York: Oxford Univ Press.
10. Murray, C.J.L. and A.D. Lopez, ed. *Global comparative assessments in the health sector. Disease burden, expenditures and intervention packages*. 1994, World Health Organization: Geneva. 196+viii.
11. Jamison, D.T., et al., ed. *Disease Control Priorities in Developing Countries*. Oxford Medical Publications. 1993, Oxford University Press (for the World Bank): Oxford.
12. Anonymous, World Bank's cure for donor fatigue. *Lancet*, 1993. 342(8863): p. 63-64.
13. Gupta, P.C., R. Sankaranarayanan, and J. Ferlay, Cancer deaths in India: is the model-based approach valid? *Bull World Health Organ*, 1994. 72(6): p. 943-944.
14. Richardson, J., Cost utility analysis: what should be measured? [review] *Soc Sci Med*, 1994. 39(1): p. 7-21.
15. Schwartz, S., J. Richardson, and P.P. Glasziou, Quality-adjusted life years: origins, measurements, applications, objections [see comments]. *Aust J Public Health*, 1993. 17(3): p. 272-8.
16. CBS, Overlevingstafels voor vrouwen naar leeftijd, 1986-1990. *Maandstatistiek bevolking*, 1991(12): p. 77.
17. Carroll, L., Through the Looking Glass, in *The Annotated Alice*, M. Gardner, Editor. 1970, Penguin Books: Harmondsworth.

Health Expectancy: from a population health indicator to a tool for policy making

Abstract

This article presents a methodologic framework for a Disability Adjusted Life Expectancy (DALE), an extended and enhanced variety of the common Health Expectancy indicator. The DALE is based on a causal link running from risk factors through diseases and disability to mortality. This causal link allows the DALE to be used for evaluation of potential interventions and analysis of observed trends. The DALE methodology consists of a combination of a multi-state life table with explicitly modelled disease processes and of disease specific disability weights. The article presents two illustrations: the impact on the DALE of trends in cardiovascular disease epidemiology, and the benefits of nonsmoking. Two problems of the DALE are discussed: estimating 'all other causes' disability and the presence of comorbidity.

Introduction

Health policy is concerned with maintaining and, where possible, improving public health, given limited resources. The responsibility of health policy makers is to choose among potential interventions. These choices are necessary because the number of potential interventions is virtually unlimited, while resources are scarce. The dilemma was perhaps exposed most starkly by the Oregon experience, where a list of publicly financed medical interventions under a budget constraint was compiled, albeit that policy makers there in fact evaded the dilemma by passing it on to the public (Sipes-Metzler, 1994).

Part of health policy is on a generic level, like providing sufficient hospital capacity and broad access to medical services. The large majority of decisions, though, are made on the level of diseases and risk factors. For example, should we provide screening for breast cancer to women under 50 years of age? Is the widespread adoption of cholesterol synthesis inhibitors to be recommended, or should we instead concentrate on supplying folic acid to women with the desire to get pregnant?

The diversity of the choices and outcomes makes it difficult to reach a decision, unless outcomes are mapped to some common denominator. Disability-free life expectancy provides such a common denominator, but its properties make it rather unfit for populations with a high level of chronic disease. Improvement of survival for the chronically ill, a major outcome of much of modern medicine, is evaluated by the disability-free life expectancy estimator as being worthless, or even detrimental (Barendregt & Bonneux, 1994). This seems hardly to be in correspondence with the appraisal of both medical professionals and the general public.

In this paper a methodologic framework for a Disability Adjusted Life Expectancy (DALE) is presented. The DALE is based on a causal chain that runs from risk factors to disease incidence, from disease incidence to disease prevalence, from disease prevalence to disability prevalence, and to disease specific and total mortality. This way the impact on Life Expectancy (LE) and DALE of various preventive and curative interventions, and of trends in the epidemiology of specific diseases can be estimated.

A multi-state, multi-disease life table for men is used to describe disease specific incidence, prevalence and mortality from five diseases: cardiovascular diseases, lung cancer, and Chronic Obstructive Pulmonary Disease (COPD). The disease specific disability weights from the Global Burden of Disease project (World Bank and WHO (Murray & Lopez, 1994b)) are applied to calculate disability prevalence and a DALE. The impact on the DALE of trends in heart disease is shown, and the current partly-smoking population is compared with one that hypothetically is smoke-free. Two major problems with the approach are discussed: the estimation of disability from all other causes, and the problem of comorbidity.

Methods

The life table

A multi-state life table with five diseases was set up: Ischemic Heart Disease (IHD), Congestive Heart Failure (CHF), Cerebrovascular Accident (CVA, or stroke), lung cancer, and Chronic Obstructive Pulmonary Disease, including emphysema (COPD). Each disease is described by a continuous-time markov process, and consists of a single prevalent state. There are six absorbing states: death from each specific disease, and from 'all other causes' (Barendregt, van Oortmarssen, van Hout, van den Bosch, & Bonneux, in press).

Transition into the disease states depends on incidence probabilities by age, CHF in addition receives input from IHD. Disease specific mortality depends on survival probabilities, also by age, cure is ignored. In all but the absorbing states people are subject to incidence probabilities for all diseases (except when they are already prevalent of that particular disease), and to the 'all other causes' mortality probabilities, that are independent from disease specific survivals. With the exception of CHF, which prevalence and mortality depend in part on IHD, diseases are assumed to be independent.

The disease processes are instances of incidence-prevalence-mortality models, that, by formalizing relationships between these variables, enforce internal consistency of the data and allow to infer unobserved data (Murray & Lopez, 1994a). Age and disease specific survival probabilities could thus be estimated from incidence and disease specific and total mortality. The transmission rate from IHD to CHF was estimated by making the life table prevalence of IHD equal to observed prevalence.

Disease and mortality data

Input data are a subset of the data collected for the stroke and heart disease models that are part of NIMPH: the Netherlands Integrated Model of Public Health. They consist of total mortality probability and disease specific mortality rates from Statistics Netherlands (Statistics Netherlands, 1991; Statistics Netherlands, Published annually). The incidence probabilities are based on a nation-wide hospital register data and regional studies (Bonneux, Barendregt, Meeter, Bonsel, & van der Maas, 1994; Niessen, Barendregt, Bonneux, & Koudstaal, 1993).

Data on COPD incidence and prevalence are from the Continuous Morbidity Registration (CMR), a regional registration in four general practices (van Weel, van den Bosch, van den Hoogen, & Smits, 1987). Since vir-

Table 1 - Proportions disabled and distribution over severity classes for 5 diseases, and weights for each severity class. From the Global Burden of Disease project.

		Severity class						Average
		I	II	III	IV	V	VI	
	weight	0.096	0.22	0.4	0.6	0.81	0.92	
	prop dis							
IHD		0.725	0.686	0.14	0.14	0.034	0	0.126
CHF		1	0	0.7	0.2	0.1	0	0.294
CVA		1	0.35	0.3	0.15	0.1	0.05	0.306
lung ca	pre-terminal	1	1	0	0	0	0	0.096
	terminal	1	0	0	0	0.4	0.3	0.759
COPD		1	0	0	0.5	0.3	0.2	0.542

tually all Dutch have a GP these data are considered to be representative. Because of small numbers (about 12,000 patients) data from 10 years were averaged.

Disability weights

Table 1 shows the disease specific disability weights from the Global Burden of Disease project for the five diseases. Each disease has a proportion disabled (in our case mostly equal to 1), and this proportion is distributed over six severity classes, with class I being the mildest and VI the worst. The weights cover a wide range: in the mildest class a person is about 10% disabled, in the severest over 90%. This means that a person in one of the mildest classes would probably not normally be classified as 'disabled' when the health state is a dichotomous variable.

Since the disease models in the multi-state life table have just one stage each, the average loss for each disease per unit time due to disability was used instead of the various severity classes. The average was calculated from the proportion disabled, the distribution over the severity classes, and the weight of each class (column 'average' in Table 1). Lung cancer has two different weights, one for the terminal stage and one for the pre-terminal. In our calculations the terminal stage was assumed to last 3 months.

An 'all causes' DALE

An 'all causes' DALE was estimated by taking as a proxy for disability distribution over age the distribution over age of costs of disease per person year. In a previous study, the 1988 Dutch health care costs have been allocated to age, sex, health care sector and disease category, based on comprehensive data on morbidity, mortality and direct costs. For each health care sector total costs were assigned to age, sex and primary diagnosis using utilization of services, the number of patients or the expenditures per group of patients. Most data were extracted from national registers with nearly complete coverage. Health insurance covers 99.8 % of the Dutch population and accessibility of health care is very high. Of all costs 25.2 % could not be allocated to a primary diagnosis, either because of missing information or because costs were non-personal (administration, general public health services, etc). The data used in this study concern only the allocated health care costs. A more complete description of health care data-bases and methods used is given elsewhere (Koopmanschap, et al., 1994).

From these costs by age and the population structure 1988 an age specific cost per person was calculated. This age gradient was used to allocate disability by age to a life table cohort such that the DALE from the life table at age 15 was equal to the Disability-Free Life Expectancy (DFLE) in the Netherlands as estimated on basis of 1989-90 data from the Statistics Netherlands (van den Bos, van der Velden, van Sonsbeek, Nusselder, & Lenior, 1994). Equating the DALE to the DFLE is warranted because the average disability weights of the DALE can be interpreted as the proportion of the diseased considered unhealthy by the DFLE (see discussion).

Cardiovascular trends and nonsmoking

This life table is used to look at the impact of trends in IHD and CHF incidence and survival on LE and DALE, and at the difference between the current partly-smoking population and one that is smoke-free. In both cases it is assumed that an 'all other causes' burden of disability can be created by simply subtracting for each age the disabled years caused by the five diseases from the 'all causes' total. In other words, comorbidity is ignored.

For the heart disease trends a combination of declining incidence and increasing survival that has been observed in the previous decades was used (Barendregt & Bonneux, 1994; Bonneux, Barendregt, Meeter, Bonsel, & van der Maas, 1994). Incidence of both diseases is varied by multiplying age specific probabilities with a factor of 0.9 (all ages), and multiply probabilities to die from a specific disease, given prevalence (i.e. 1-survival probability) likewise with 0.8, always comparing outcome with the 1988 base year. These multiplication factors reflect cumulative changes observed over a ten-year period.

Secondly, a hypothetical smoke-free population was created by using the smoking prevalences and relative risks in Table 2 to estimate nonsmoking incidences. Smoking is a direct risk factor for IHD, CVA, Lung cancer, and COPD, and an indirect one (through IHD) for CHF. The life table was recalculated with the nonsmoking incidences while keeping all else equal.

Results

Table 3 shows the result from the 'all causes' DALE estimation procedure. The LE's show some slight differences, because total mortality probabilities are based on not exactly corresponding time periods. The DALE is calibrated to equal the Disability-Free Life Expectancy at age 15, which gives a DALE at birth of 64.7 years. For most ages the DFLE and DALE are quite similar, and the Disabled Life Expectancies (DLE) are also on a par.

An estimate of the disease specific burden of disability, expressed as number of disabled person years of the life table cohort, is provided in Table 4. The base line estimate shows that IHD, CHF, and CVA impose roughly comparable burdens of disability, lung cancer much less due to its lower incidence and high lethality, and COPD far more. The five diseases together are responsible for about 30% of the total burden of disability.

Table 2 - Prevalence of smoking (%) by age; and Relative Risks (RR) of smoking on incidence for 5 diseases.

Age group	Prevalence of smoking (%)	Disease	RR
0-14	0	IHD	3
15-19	20	Lung ca	10
20-34	39	CVA	2
35-49	42	CHF	1
50-64	39	COPD	25
65+	34		

The top panel also shows the impact of the simultaneous decrease in IHD and CHF incidence and improvement of survival. The burden of disability for CVA, Lung cancer, COPD, and 'all other causes' all go slightly up, due to the decrease in total mortality. IHD is down marginally, while CHF is clearly up. This is explained by the fact that CHF, on top of its improved survival, also acts as end-stage heart disease, which many IHD patients get if their survival of acute coronary events is long enough (Bonneux, Barendregt, Meeter, Bonsel, & van der Maas, 1994). On balance total disability increases by 2 percent.

Figure 1 shows age-specific DALEs (expressed as an index with the base line value equal to 100) for decreased incidence, improved survival, and the combination, again for IHD and CHF. Decreased incidence results in a higher DALE throughout the age range (like a DFLE), improved survival results in an increase at younger ages, and a decrease at older (unlike a DFLE, which shows no effect at birth and a decrease at older ages), and the combination of

Table 3: Age specific Life Expectancy (LE), Disability Free Life Expectancy (DFLE), Disability Adjusted Life Expectancy (DALE), and Disabled Life Expectancy (DLE): outcome of the 'all causes' DALE estimation. StatNeth: Statistics Netherlands; estimate: this paper.

Age	LE		DFLE		DALE		DLE	
	StatNeth	estimate	StatNeth	estimate	StatNeth	estimate	StatNeth	estimate
0	73.3	73.5			64.7			8.9
15	59.3	59.4	51.3	51.3	51.3	8.0	8.1	
30	44.8	44.9	37.0	37.7	37.7	7.8	7.2	
45	30.5	30.7	23.3	24.5	24.5	7.2	6.1	
65	14.2	14.3	9.6	9.5	9.5	4.6	4.7	
75	8.5	8.5	4.6	4.5	4.5	3.9	4.0	
85	4.7	4.7	1.4	1.6	1.6	3.3	3.1	

Table 4: Disease specific, 'all other causes', and total burden of disability in person years of the life table cohort, base line estimate with two variants, indices base line=100.

	IHD	CHF	CVA	Lung ca	COPD	All other	Total
Base line	43086	41576	22871	4014	150449	613053	875051
IHD & CHF trends	42066	43977	23605	4150	154722	622611	891132
index	98	106	103	103	103	102	102
smoke-free	28310	46069	19892	1160	18496	668278	782206
index	66	111	87	29	12	109	89

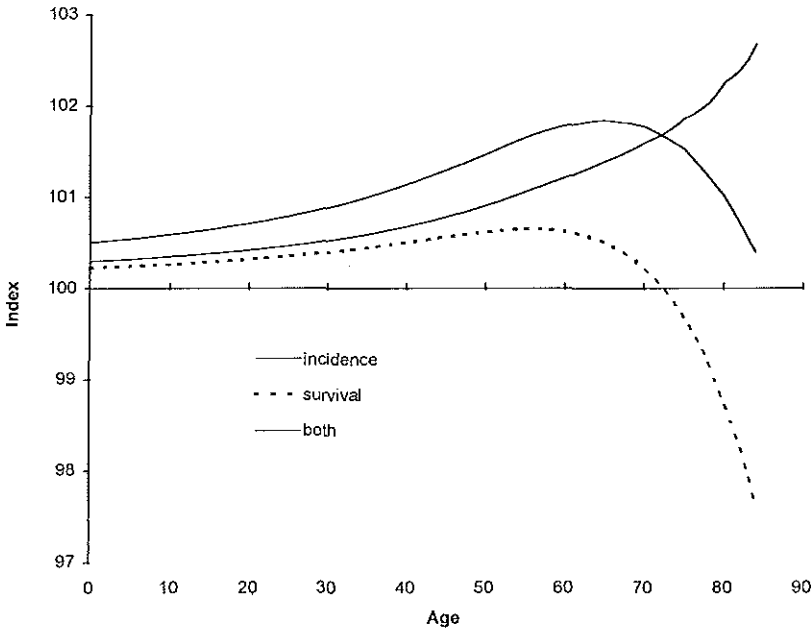


Figure 1 - Disability Adjusted Life Expectancies (DALE) by age after separate and combined decline in incidence and Improvement of survival for heart disease. Index with base line equal to 100.

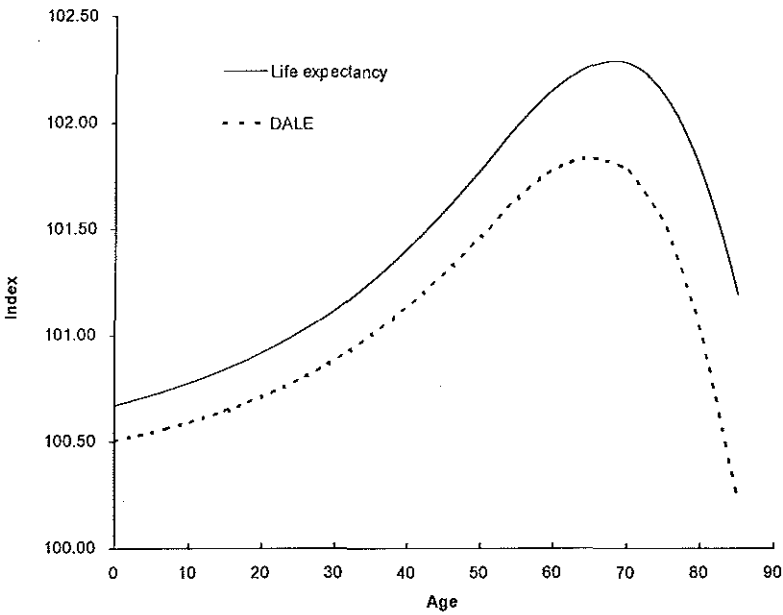


Figure 2 - Life Expectancies (LE) and Disability Adjusted Life Expectancies (DALE) by age after a simultaneous decline in incidence and improvement of survival for heart disease. Index with base line equal to 100.

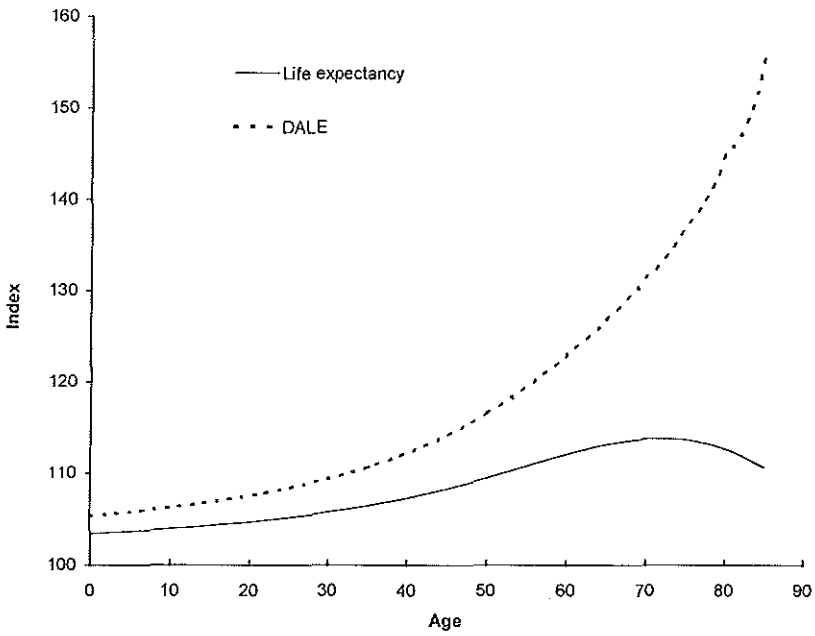


Figure 3 - Life Expectancies (LE) and Disability Adjusted Life Expectancies (DALE) by age without smoking. Index with base line equal to 100.

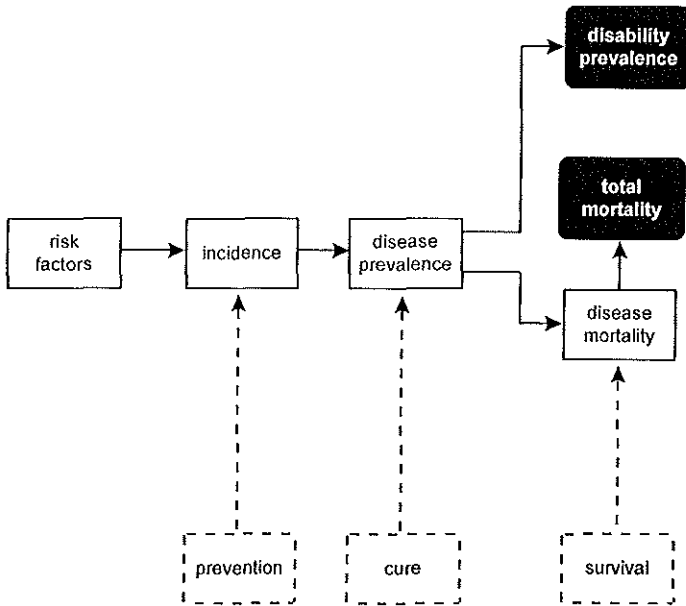


Figure 4 - Elements of the Disability Process. Health Expectancy usually concerns the black boxes, Disability Adjusted Life Expectancy (DALE) includes the white boxes and thereby provides additional policy controls (dashed boxes).

the two shows an increase for most ages, with an maximum around 65, and a slight decrease at the oldest age (much like a DFLE) (Barendregt & Bonneux, 1994).

Figure 2 shows an index line (again with base line value equal to 100) for the LE and DALE after the heart disease trends. Clearly the trends in heart disease result in a higher LE and DALE, but the increase in the DALE is a bit less, due to the improved survival. The recent trends in heart disease therefore seem to have caused an expansion of disability.

The lower panel of Table 4 also gives the results for the hypothetical smoke-free population. Disability caused by COPD virtually disappears, lung cancer is much less, IHD is about a lower, and CVA 10%, but CHF is 10% higher and 'all other causes' 9%. On balance disability is 11% less if smoking would not exist, largely due to the decrease in COPD disability. Figure 3 shows the resulting index lines for the LE and DALE. A smoke-free population would clearly not just have a higher LE, but would gain even more in disability adjusted years: a potential compression of disability exists if we would succeed in further reducing smoking.

Discussion

Health expectancy indicators come in many varieties, but all have at least one aspect in common: they combine information on health states (however defined) with mortality rates into a single indicator. Mostly health state prevalences are used (Sullivan method), in some cases transition probabilities (multi-state life table method), and usually the unhealthy state is assigned a weight of 0, although also ranges between 0 and 1 are sometimes used (Mathers, Robine, & Wilkins, 1994; Wilkins & Adams, 1983).

The DALE methodology presented in this paper differs on two main points from the more customary indicators of health expectancy. The first point is the inclusion of a complete, if simplified, process leading from risk factor exposure and disease incidence to disability and mortality (see figure 4). The second is the use of a much more elaborate and disease specific disability weighing schedule than the 0-1 weights of most health expectancy calculations, which produce estimates of disability-free life expectancy (DFLE).

Estimates of DFLE so far have mostly been of a descriptive nature. Researchers have mostly limited themselves to estimating total mortality and disability prevalence (the black boxes in figure 4), and calculated a DFLE from the two. There are two main applications for such estimates: one to compare between countries (to see which countries are doing better and, perhaps, learn from that), and secondly to compare estimates within a country on different points in time (to estimate trends). Although neither application is without its problems, these are relevant uses (Barendregt, Bonneux, & Van der Maas, 1994; Boshuizen & Van de Water, 1994).

Taking a thus measured DFLE beyond the use as a population health indicator is not possible, however. Neither total mortality nor disability

prevalence allows one to infer what the reasons are for its particular level and the changes therein. A decline in life expectancy, as currently observed in several Middle and Eastern European countries, cannot be explained (much less acted upon) unless it is broken down to its contributing causes of death (injuries and cardiovascular disease seem important contributors) (Feachem, 1994; World Bank, 1993).

Much the same goes for disability. Disability is caused by diseases. Only when the observed level of disability can be attributed to specific diseases does it become possible to determine what policy will be most appropriate for improvement. Potential interventionist health policies are unlikely to be on a generic disability level: that is too unspecific. Health policy concentrates on preventing diseases, curing them, and providing care and extending survival when neither works (affecting the white boxes in figure 4). Much of prevention, almost all cure, and a major part of survival extending care is disease specific. When researchers are satisfied to estimate DFLE using just total mortality and disability prevalence (the black boxes in figure 4) they severely limit the possibilities for policy making.

The second point distinguishing the DALE methodology from the customary indicators of health expectancy, the use of disease specific disability weights, follows from the first. The disease specific approach requires a mapping from disease specific prevalences to disability prevalence. Assigning weights to non-fatal health outcomes is as such no new feature, witness for example the Quality Adjusted Life Years or QALY (Drummond, Stoddart, & Torrance, 1987). The particular mapping employed in this paper was developed for the joint WHO and World Bank Global Burden of Disease project, that was the basis of the 1993 World Development Report *Investing in Health* (World Bank, 1993). The disease specific disability weights were developed in an elaborate and iterative process, involving a large number of international disease experts, repeated internal consistency checks, and independent reviews (Murray & Lopez, 1994a). More than 100 conditions were thus evaluated.

On the basis of these disability weights, Murray and colleagues then went on to develop the DALY (Disability Adjusted Life Years), to be used in cost-effectiveness studies. Citing economic considerations an age-weighting schedule and a discount rate were applied (Murray, 1994). For their specific purpose, a cost-effectiveness study, these are legitimate (if debatable) add-ons (Barendregt, Bonneux, & Van der Maas, 1996), but the build-in economic features render the DALY unfit for the calculation of a DFLE or DALE (Ginneken, 1994).

But the raw disability weights, on the other hand, provide an opportunity to get around some severe limitations of the customary health expectancy estimator. They permit the use of a bottom-up, disease specific approach, which renders the estimator much more powerful. The weights also permit one to avoid the problem of the standard health expectancy, which evaluates an improvement in survival with disability as worthless (at birth) or even detrimental (at higher ages) (Barendregt & Bonneux, 1994). The weights are available for a large number of conditions, including injuries. Currently work

is being carried out on refinement and further validation of the disability weights (Murray, 1996).

One might argue that the elaborate weighting system, with for each disease a proportion disabled, and a distribution over 6 severity classes with specific weights (see Table 1), hampers the interpretation of the results, and therefore makes the DALE less intuitive than the DFLE. But this system seems primarily designed to allow the experts making the evaluations to account for disease heterogeneity. The more elaborate system eventually collapses to just one weight per disease (the column labeled 'average' in Table 1) as long as no changes in disease epidemiology occur.

This single weight can be interpreted as the average disability weight for all people with that particular disease, or as the proportion of people with the disease that is disabled (for those who do not wish to part with the 0-1 weighting schedule). This equivalence between an average weight and a division in two proportions with weights 0 and 1 is analogous to the time trade-off method popular in Quality of Life research (Torrance, 1987). In the latter interpretation the DALE simply becomes an aggregation of disease specific disability-free life expectancies.

This equivalence in principle also permits one to interpret an estimate of an 'all causes' DFLE as an 'all causes' DALE. The problem with this interpretation is that the definitions of what constitutes 'disability' are arrived at along vastly different ways: one through expert opinion on severity of diseases, the other through a population survey on the presence of disability. Although we expect both methods to produce largely corresponding results, this is an untested (but testable) hypothesis.

An estimate of an 'all causes' DALE is a prerequisite to obtaining an 'all other causes' DALE, and cannot be arrived at by simply adding up single disease estimates for two reasons. One reason is the huge number of diseases, for each of which estimates of incidence, prevalence, and mortality would have to be made (and checked for internal consistency), the other is that disabilities caused by single diseases do not simply add up due to comorbidity.

In the present study the comorbidity problem and the non-additivity that follows from it is ignored, and an 'all other causes' DALE is estimated by subtracting the disability contributions of the five diseases from the 'all causes' disability. The existence of comorbidity implies that there are fewer people with any disease than the sum over all disease prevalences. Since one can have a specific disability only once, this would imply that there is also less disability, but for the fact that comorbidity in most cases seems to make disability more severe (Verbrugge, Lepkowski, & Imanaka, 1989). So on balance the bias on our 'all other causes' DALE estimate could go either way. This problem with the DALE methodology is as yet unsolved, although the work of Nusselder *et al.* (1994) suggests a possible solution.

Another problem with the DALE methodology is data: data requirements far exceed those of a DFLE, in particular a Sullivan-type DFLE. While disease specific mortality rates will often be available, this is certainly not true for disease specific incidence and prevalence rates. Lack of epidemiologic data will in many cases thwart estimation of a DALE.

Despite the loose ends and higher data requirements the DALE methodology is well worth pursuing. Earlier attempts to estimate the influence of specific diseases on health expectancy relied on combined cause elimination and Sullivan methods (Mathers, 1992; Nusselder, van der Velden, Lenior, van Sonsbeek, & van den Bos, 1994). While this certainly produced illuminating results, cause elimination remains a crude and unrealistic method. Crimmins *et al.* (1994) used a multi-state life table technique to look at the effect of changing mortality and morbidity on health status and life expectancy. Unfortunately they operationalized morbidity in terms of functional ability, leaving the readers wondering in what way the imputed changes in morbidity and mortality might be brought about. Stoto and Durch (1991) tried to estimate the impact of the US Healthy People 2000 targets on life expectancy and healthy life expectancy. Because of the lack of a methodologic framework linking risk factors and preventive (and, one should add, curative) services to health status measures and mortality they simply used causally unrelated changes in mortality and disability in a Sullivan life table. Their results are therefore less than convincing.

The DALE methodology represents such a framework that links risk factors, prevention and therapy to disability and mortality. The current model is very simple, but could be extended by allowing more risk factors, levels of exposure, and multi-stage disease models. For example, when improvements in disease survival occur through decreased mortality in acute phases, patients tend to amass in the severest disease stage because all preceding stages act as risk factors for the severest stage (Bonneux, Barendregt, Meeter, Bonsel, & van der Maas, 1994; Niessen, Barendregt, Bonneux, & Koudstaal, 1993). Such a change in disease epidemiology might have an important effect on associated disability. But of course the drawback of multi-stage disease models is that data requirements multiply.

The DALE is more powerful than the customary health expectancy indicator because it explains disability and mortality through the epidemiology of their underlying causes, instead of just describing them as such. Because the DALE encompasses the complete process from risk factors to diseases, to disability and mortality, it provides policy makers with a much enhanced array of potential controls to manipulate.

For the same reason it may help researchers to better understand the changes in population health status. Our exercises with heart disease trends and a smoke-free population are a case in point. Increased survival with chronic disease is not restricted to heart disease: improved care for acutely ill patients tends to keep patients alive longer across a large array of chronic conditions. Just like with heart disease improved survival will tend to amass patients in the end-stage of the disease, with the severest disability. Unless medicine finds ways to stop the chronic diseases from progressing, or health promoters can convince people to give up smoking, an expansion of disability seems inevitable.

References

- Barendregt, J. J., & Bonneux, L. (1994). *Changes in incidence and survival of cardiovascular disease and their impact on disease prevalence and health expectancy*. In: Mathers C, McCallom J, Robine J-M, eds. *Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES)*, Canberra, February 1994. Canberra: Australian Institute of Health and Welfare: AGPS, 345-354
- Barendregt, J. J., Bonneux, L., & Van der Maas, P. J. (1994). Health Expectancy, an indicator for change? *Journal of Epidemiology and Community Health*, 48, 482-487.
- Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1996). DALYs: the age-weights on balance. *Bull World Health Organ*, 74(4), 439-443.
- Barendregt, J. J., van Oortmarsen, G. J., van Hout, B. A., van den Bosch, J. M., & Bonneux, L. (in press). Coping with multiple morbidity in a life table. *Mathematical Population Studies*.
- Bonneux, L., Barendregt, J. J., Meeter, K., Bonsel, G. J., & van der Maas, P. J. (1994). Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J of Public Health*, 84, 20-28.
- van den Bos, G. A. M., van der Velden, J., van Sonsbeek, J. L. A., Nusselder, W. J., & Lenior, M. E. (1994). *Gezonde levensverwachting en chronische ziekten: vergelijking van databronnen en berekeningen*. ISG, NIVEL, CBS.
- Boshuizen, H., & Van de Water, H. (1994). *An international comparison of health expectancies* (No. TNO-PG 94.046). TNO Preventie en Gezondheid.
- Crimmins, E. M., Hayward, M. D., & Saito, Y. (1994). Changing mortality and morbidity rates and the health status and life expectancy of the older population. *Demography*, 31(1), 159-75.
- Drummond, M. F., Stoddart, G. L., & Torrance, G. W. (1987). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press.
- Feachem, R. (1994). Health decline in eastern Europe. *Nature*, 367(6461), 313-4.
- van Ginneken, J. K. (1994). *Potential of the Global Burden of Disease project for determining health expectancy*. In: Mathers C, McCallom J, Robine J-M, eds. *Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES)*, Canberra, February 1994. Canberra: Australian Institute of Health and Welfare: AGPS, 170-181.
- Koopmanschap, M., van Roijen, L., Bonneux, L., Bonsel, G., Rutten, F., & van der Maas, P. (1994). Costs of diseases in an international perspective. *Eur J Public Health*, 4(4), 258-264.
- Mathers, C. D. (1992). *Estimating gains in health expectancy due to elimination of specified diseases. Paper presented to the 5th meeting of the International Network on Health Expectancy*. Ottawa.
- Mathers, C. D., Robine, J. M., & Wilkins, R. (1994). *Health expectancy indicators: recommendations for terminology*. In: Mathers C, McCallom J, Robine J-M, eds. *Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES)*, Canberra, February 1994. Canberra: Australian Institute of Health and Welfare: AGPS, 34-41.

- Murray, C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ*, 72(3), 429-45.
- Murray, C. J. (1996). Rethinking DALYs. In C. J. Murray & A. D. Lopez (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*. Cambridge MA: Harvard University Press.
- Murray, C. J., & Lopez, A. D. (1994a). Quantifying disability: data, methods and results. *Bull World Health Organ*, 72(3), 481-94.
- Murray, C. J. L., & Lopez, A. D. (Eds.). (1994b). *Global comparative assessments in the health sector. Disease burden, expenditures and intervention packages*. Geneva: World Health Organization.
- Niessen, L. W., Barendregt, J. J., Bonneux, L., & Koudstaal, P. J. (1993). Stroke Trends in an Aging Population. *Stroke*, 24, 931-939.
- Nusselder, W. J., Velden, J., van der, Lenior, M. E., Sonsbeek, J. L. A., van, & Bos, G. A. M., van der (1994). *The effect of elimination of selected chronic diseases on disability-free life expectancy: compression or expansion of morbidity? Preliminary results*. In: Mathers C, McCallom J, Robine J-M, eds. *Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES)*, Canberra, February 1994. Canberra: Australian Institute of Health and Welfare: AGPS, 355-365.
- Sipes-Metzler, P. R. (1994). Oregon Health Plan: ration or reason. *J Med Philos*, 19(4), 305-14.
- Statistics Netherlands (1991). Life tables for men by age, 1986-1990. *Maandstatistiek bevolking*(12), 76.
- Statistics Netherlands (Published annually). *Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1*. Voorburg: CBS.
- Stoto, M. A., & Durch, J. S. (1991). National Health Objectives for the Year 2000: The Demographic Impact of Health Promotion and Disease Prevention. *Am J Public Health*, 81, 1456-1465.
- Torrance, G. W. (1987). Utility approach to measuring health-related quality of life. [Review]. *J Chronic Dis*, 40(6), 593-603.
- Verbrugge, L. M., Lepkowski, J. M., & Imanaka, Y. (1989). Comorbidity and its impact on disability. *Milbank Q*, 67(3-4), 450-84.
- van Weel, C., van den Bosch, W. J. H. M., van den Hoogen, H. J. M., & Smits, A. J. A. (1987). Development of respiratory illness in childhood - a longitudinal study in general practice. *J R Coll Gen Pract*, 37, 404-408.
- Wilkins, R., & Adams, O. B. (1983). Health Expectancy in Canada, Late 1970s: Demographic, Regional, and Social Dimensions. *Am J Public Health*, 73, 1073-1080.
- World Bank (1993). *Investing in Health*. New York: Oxford Univ Press.

Disability Adjusted Life Expectancy and comorbidity: exploring uncertainty and sensitivity

Abstract

The Disability Adjusted Life Expectancy (DALE) consists of a combination of a multi-state life table with explicitly modeled disease processes, together with disease specific disability weights. The explicitly modeled disease processes cause comorbidity to be a problem, in two different ways. One is that due to comorbidity disabilities caused by single diseases do not simply add up. And second, it seems that in most cases comorbidity makes disability more severe.

We present an analysis to explore quantitatively the impact of uncertainty about the severity of disability in the case of comorbidity, and of the sensitivity of the estimate of intervention effects and the aggregate level of disability in the population for various assumptions on comorbidity prevalence. We use a multi-state life table with 5 diseases and 'all other causes' disability, and do an intervention on smoking prevalence.

We list the elements of the estimate that are uncertain, and look at the sensitivity of the outcome to variations in the values of these uncertain elements. We show that the disability weight in the case of comorbidity is but one of a range of uncertain elements, and that the sensitivity of the estimated years lived with disability to variations in the comorbidity weight is relatively small. We conclude that for the estimation of the DALE and of the impact of interventions on it other uncertain variables than the disability weight of comorbidity deserve priority for research. The use of a simple rule to assign a disability weight for comorbidity (like the highest weight of the constituent diseases), or even ignoring comorbidity altogether hardly affects DALE estimates.

Introduction

The Disability Adjusted Life Expectancy (DALE) is an extended and enhanced variety of the Health Expectancy indicator. The underlying concept is the establishment of a causal link running from risk factors through diseases and disability to mortality. This causal link allows the DALE to be used for evaluation of potential interventions and analysis of observed trends. The DALE methodology consists of a combination of a multi-state life table with explicitly modelled disease processes, together with disease specific disability weights (Barendregt, Bonneux, & van der Maas, in press).

A major problem of the DALE methodology is the presence of comorbidity, in two different ways. One is that due to comorbidity disabilities caused by single diseases do not simply add up. The existence of comorbidity implies that there are fewer people with any disease than the sum over all disease prevalences. Since one can have a specific disability only once, this implies that in the aggregate there is less disability than the sum over all single disease disability. On the other hand, and this is the second problem, it seems that in most cases comorbidity makes disability more severe (Verbrugge, Lepkowski, & Imanaka, 1989). Estimates of total disability that ignore comorbidity will therefore be too high due to the first problem, and too low due to the second problem, and how the balance is of these countervailing tendencies is not obvious.

In this paper we present an analysis to explore quantitatively the impact of comorbidity on the estimated number of years lived with disability. We list the elements of the estimate that are uncertain, and look at the sensitivity of the outcome to variations in the values of these uncertain elements. We conclude on the relative importance of the disability weights in case of comorbidity, as compared to the contributions from the other sources.

Methods

DALE methodology

We use the same multi-state life table as in our earlier DALE calculations, with the exception that now comorbidity is modelled explicitly (Barendregt, Bonneux, & van der Maas, in press). The table consists of five diseases: Ischemic Heart Disease (IHD), Congestive Heart Failure (CHF), Cerebrovascular Accident (CVA, or stroke), Lung Cancer, and Chronic Obstructive Pulmonary Disease, including emphysema (COPD). Each disease is described by a continuous-time markov process, and consists of a single, age specific prevalent state (Barendregt, van Oortmarssen, van Hout, van den Bosch, & Bonneux, in press). There are six absorbing states: death from each specific disease, and from 'all other causes'.

Transition into the disease states depends on incidence probabilities by age, CHF in addition receives input from IHD. Disease specific mortality depends on survival probabilities, also by age, cure is ignored. In all but the ab-

sorbing states people are subject to incidence probabilities for all diseases (except when they are already prevalent of that particular disease), and to the 'all other causes' mortality probabilities, that are independent from disease specific survivals. With the exception of CHF, which prevalence and mortality depend in part on IHD, diseases are assumed to be independent.

Input data are a subset of the data collected for the stroke and heart disease models that are part of NIMPH: the Netherlands Integrated Model of Public Health. They consist of total mortality probability and disease specific mortality rates from Statistics Netherlands (Statistics Netherlands, 1991; Statistics Netherlands, Published annually). The incidence probabilities of cardiovascular diseases are based on a nation-wide hospital register data and regional studies (Bonneux, Barendregt, Meeter, Bonsel, & van der Maas, 1994; Niessen, Barendregt, Bonneux, & Koudstaal, 1993).

Lung Cancer incidence is from a regional study (Eindhoven Cancer Registry). Data on COPD incidence and prevalence are from the Continuous Morbidity Registration (CMR), a regional registration in four general practices (van Weel, van den Bosch, van den Hoogen, & Smits, 1987). Since virtually all Dutch have a GP these data are considered to be representative. Because of small numbers (about 12,000 patients) data from 10 years were averaged.

Table 1 shows the disease specific disability weights from the Global Burden of Disease project for the five diseases (Murray & Lopez, 1994). Each disease has a proportion disabled (in our case mostly equal to 1), and this proportion is distributed over six severity classes, with class I being the mildest and VI the worst. The weights cover a wide range: in the mildest class a person is about 10% disabled, in the severest over 90%. This means that a person in one of the mildest classes would probably not normally be classified as 'disabled' when the health state is a dichotomous variable.

Table 1 - Proportions disabled and distribution over severity classes for 5 diseases, and weights for each severity class. From the Global Burden of Disease project.

		Severity class						Average
		I	II	III	IV	V	VI	
	weight	0.096	0.22	0.4	0.6	0.81	0.92	
	prop dis							
IHD		0.725	0.686	0.14	0.14	0.034	0	0.125466
CHF		1	0	0.7	0.2	0.1	0	0.294
CVA		1	0.35	0.3	0.15	0.1	0.05	0.3061
lung ca	pre-terminal	1	1	0	0	0	0	0.096
	terminal	1	0	0	0	0.4	0.3	0.759
COPD		1	0	0	0.5	0.3	0.2	0.542

Since the disease models in the multi-state life table have just one stage each, the average loss for each disease per unit time due to disability was used instead of the various severity classes. The average was calculated from the proportion disabled, the distribution over the severity classes, and the weight of each class (column 'average' in Table 1). Lung cancer has two different weights, one for the terminal stage and one for the pre-terminal. In our calculations the terminal stage was assumed to last 3 months.

An important aspect of the DALE methodology is the estimation of an 'all causes' DALE. Because the five diseases only describe a limited part of the total number of years lived with disability by the cohort we need an estimate of the disability caused by 'all other causes'. This disability from 'all other causes' is estimated by subtracting the disability modeled from an 'all causes' disability. The 'all causes' disability is estimated using the Disability-Free Life Expectancy (DFLE) in the Netherlands at age 15 as based on 1989-90 data from Statistics Netherlands (van den Bos, van der Velden, van Sonsbeek, Nusselder, & Lenior, 1994). To distribute the associated disability over age we assumed that health care costs are a good proxy for disability, and that disability therefore shows the same distribution over age as total health care costs per person year. More details are in (Barendregt, Bonneux, & van der Maas, in press).

Comorbidity

In our earlier DALE exercises we ignored comorbidity: for each disease modeled we calculated years lived with disability, and the sum over the diseases was subtracted from the 'all causes' disability to obtain an 'all other causes' estimate (all done age specifically). Taking explicitly account of comorbidity complicates the calculations considerably: the number of combinations, given 5 diseases, is $2^5 - 1 = 31$. Many of these will have vanishing small prevalences, though, in particular the comorbidities of three or more diseases. To simplify matters we therefore look at the combined prevalences of any two diseases only. Heart disease is modeled as Ischemic Heart Disease and Heart Failure, and we assume that people with both are in fact Heart Failure patients. By means of these simplifications the number comorbid conditions reduces to 9, in addition to the 5 single disease conditions.

The prevalence of the comorbid conditions is calculated by assuming independence, and the disability weight is assumed equal to the highest weight (worst condition) of the constituent diseases. Both assumptions are subject to the uncertainty and sensitivity analysis (see below).

Uncertainties

The calculation of a DALE with explicit attention to the impact of comorbidity involves several elements, all surrounded by at least some uncertainty. These elements are:

- The estimate of the years lived with disability from 'all causes'. Uncertainty is caused by the procedure to derive this from the Dutch estimate of DFLE, and by the estimation of the latter itself.
- Disease prevalences. The estimates of disease prevalences are based on a variety of sources, made internally consistent by using an incidence-prevalence-mortality model (Barendregt, Bonneux, & van der Maas, in press; Murray, 1994). In the life table the prevalences are the result of incidence and survival rates, we will vary only the former to examine the influence of different prevalence estimates (the effects of varying survival rates are very similar for our current purposes).
- Comorbidity prevalence. When diseases share a risk factor, they will most likely be clustered instead of being independently distributed over the population. Clustering will increase the number of patients with comorbidity.
- Disability weights. The disability weights from the GBD-project were arrived at by expert opinion, and are without doubt subject to uncertainty.
- Disability weights for comorbidity. The baseline assumption is that the weight in case of comorbidity is equal to the highest weight of the constituent diseases, but it may very well be lower or, more likely, higher.
- These are the uncertain variables in case of the DALE estimate. When looking at the impact of an intervention on the DALE there is an additional uncertainty:
- The impact of the intervention on disease epidemiology.

All uncertain variables are varied around the baseline value with +20%, +10%, -10%, and -20%. The intermediate values allow to check for non-linearities. In all cases we compare the outcomes of disease specific disability, comorbidity disability, 'all other causes' disability, and 'all causes' disability with the baseline values.

To look at the impact of an intervention we created a hypothetical smoke-free population by using the smoking prevalences and relative risks in Table 2 to estimate nonsmoking incidences (Barendregt, van Oortmarssen, van Hout, van den Bosch, & Bonneux, in press). Smoking is a direct risk factor for IHD, CVA, Lung cancer, and COPD, and an indirect one (through IHD) for CHF. For each of the uncertainty variants the life table

Table 2 - Prevalence of smoking (%) by age; and Relative Risks (RR) of smoking on incidence for 5 diseases.

Age group	Prevalence of smoking (%) [*]	Disease	RR [†]
0-14	0	IHD	3
15-19	20	Lung ca	10
20-34	39	CVA	2
35-49	42	CHF	1
50-64	39	COPD	25
65+	34		

^{*} Source: (Foundation of Public Health and Smoking)

[†] Source: (van de Mheen & Gunning-Schepers, 1996)

was recalculated with the nonsmoking incidences while keeping all else equal. To make the outcomes of the interventions comparable they are expressed relative to these different DALE baseline estimates of the uncertainty variant.

Results

All results are for the male population only. Table 3 shows in the upper panel years lived with disability by the life table cohort for the five diseases, 'all other causes', and 'all causes', both while ignoring comorbidity and with comorbidity modeled. Disability attributed to specific single diseases is of course lower when comorbidity is modeled: depending on the disease by 15 to 44%. Comorbidity under the baseline assumptions accounts for almost 20% of the disability caused by the five diseases. Total disability caused by the five diseases is 8% lower when comorbidity is taken into account, making the 'all other causes' estimate 3% higher.

Figure 1 shows differences (expressed as percentages) in years lived with disability as a consequence of 20% higher values for the uncertain variables. Baseline results are represented by the Y-axis, and positive values designating more disability, and vice versa.

Results are virtually symmetrical for higher and lower values. A 20% higher estimate of the total years lived with disability has no influence on disease specific and comorbidity disability, but it increases the 'all other causes' disability with 28%. Increased disease incidence (and hence prevalence and mortality) tends to lower disability attributable to the other single diseases, because it tends to increase comorbidity. Two exceptions exist: higher IHD incidence increases the disability burden of CHF because IHD acts as a risk factor for CHF. And increased lung cancer incidence lowers

Table 3 - Disease specific, all other, and total burden of disability in person years of the life table cohort, base line estimates (upper panel) and impact of smoking intervention (lower panel) for models with and without comorbidity.

Baseline	IHD	CHF	CVA	Lung ca	COPD	Comorb	All other	Total
no comorb	43,000	42,000	23,000	4,000	150,000	0	624,000	886,000
comorb	37,000	32,000	14,000	2,000	110,000	48,000	644,000	886,000
%	-15	-23	-39	-44	-27	-	3	0
Intervention								
no comorb	28,000	46,000	20,000	1,200	18,000	0	682,000	796,000
%	-34	11	-13	-71	-88	-	9	-10
comorb	27,000	43,000	15,800	870	14,600	8,000	707,000	817,000
%	-26	35	13	-61	-87	-83	10	-8

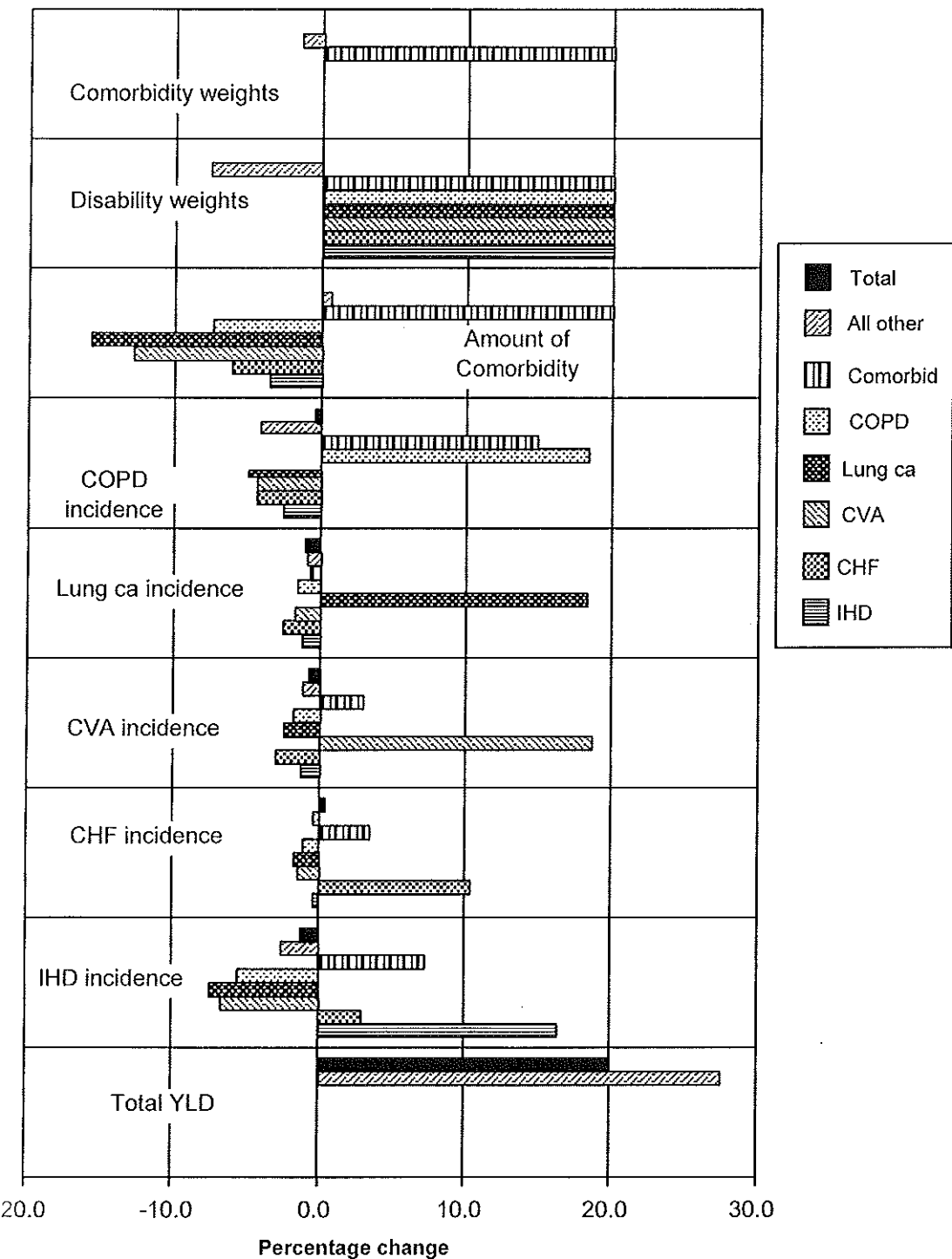


Figure 1 - Sensitivity (% difference with baseline, including comorbidity) of years lived with disability estimates for 20% higher values of uncertain variables. Positive values designate more disability, and vice versa.

comorbidity disability because of its high lethality. Increased disease incidence also increases total disability attributable to the five diseases, consequently the disability from 'all other causes' becomes lower, in particular with IHD and COPD.

Comorbidity prevalence has a rather strong influence on the disability attributable to single diseases, but not much on total disability caused by the five diseases: 'all other' causes disability remains virtually unchanged. The disability weights have a marked influence on disability caused by the five diseases, and therefore also on the 'all other causes' estimate. The weight for comorbidity, on the other hand, influences only the amount of disability from comorbidity, and has only little influence on the 'all other causes' disability.

Table 3 also shows (lower panel) the difference explicit modeling of comorbidity makes when an intervention is made. When comorbidity is ignored the abolishment of smoking lowers disability of all diseases in the life table, with the exception of CHF. Disability from 'all other causes' increases 9% because of longer life expectancy, and total disability declines by 10%. Modeling comorbidity changes this, primarily because the intervention reduces comorbid disability by over 80%. Part of this reduction in comorbid disability becomes single disease disability. The reductions in single disease disability after smoking abolishment are therefore smaller (COPD, lung cancer, IHD), change into an increase (CVA), or the existing increase becomes larger (CHF). 'All other cause' disability increases a bit more, and total disability a bit less.

In figure 2 the sensitivity of the intervention outcome to 20% higher values of the uncertain variables is presented: the results are again virtually symmetrical, therefore the results for 20% lower values are not shown. For each variant we first recalculated new baseline values of disease specific and 'all other' disability, and next looked at the post-intervention disability. The relative impacts of the interventions are compared with the one of table 3, and the difference is again expressed as a percentage, with the baseline relative impact represented by the Y-axis, and positive values designating more remaining disability and less impact, and vice versa.

Results go in the expected direction. Higher total disability before the intervention makes it less effective, because 'all other' disability is higher, and will substitute more of the disability avoided by the intervention.

Individual disease incidence changes all show the same pattern (with the notable exception of CHF disability after IHD incidence change): higher original incidence causes the intervention to be relatively less effective for other diseases. This outcome is due to several effects. One is that the other diseases get a lower baseline disability, because total mortality in the baseline is higher, and a larger part of the disability is assigned to comorbidity. The second effect is that with a larger baseline comorbidity disability more of this disability will become single disease disability due to the intervention. And the third effect is that the intervention avoids more mortality, and therefore more disease substitution will occur, with consequently higher disease specific disability. On balance intervention impact, as measured by total disability, is slightly larger.

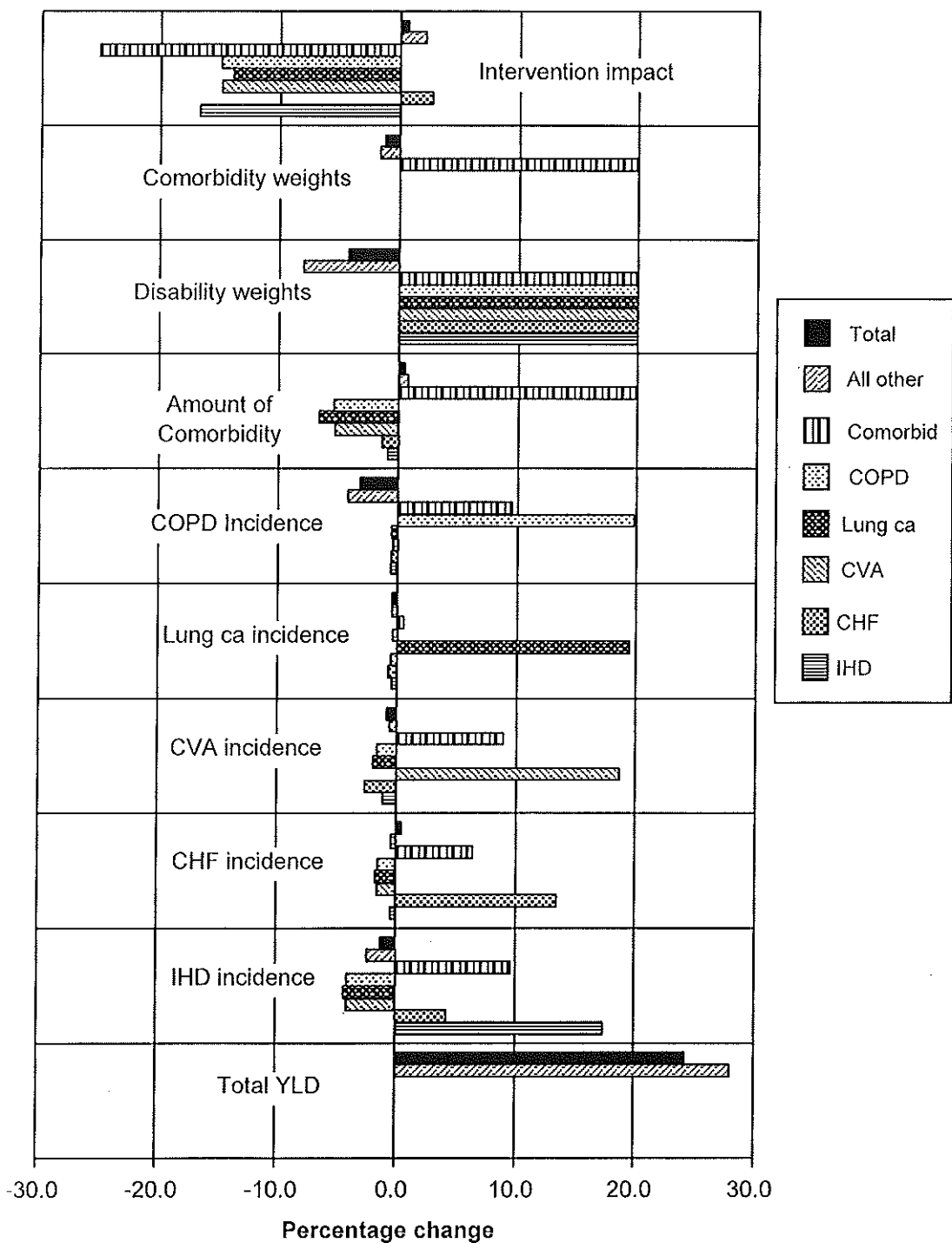


Figure 2 - Sensitivity of the relative impact of the smoking intervention (% difference with the relative impact in table 3) for 20% higher values of uncertain variables. Positive values designate less impact, and vice versa.

An increased amount of comorbidity makes the intervention on balance slightly less effective because 'all other' disability is higher: with higher comorbidity prevalence less of total disability is attributed to the explicitly modeled disease. Increased disability weights gives a rather more effective intervention, because more disability is explained by the smoking-related diseases. Higher comorbidity weights causes the pre- and consequently post-intervention comorbidity disability to be higher, but otherwise has only a small impact. And finally, a higher intervention impact causes the disease specific and comorbidity disability to be much lower, but because of the increased life expectancy the total number of years lived with disability by the life table cohort is slightly higher.

Discussion

In the DALE methodology, as presented in a previous paper, two major problems exist: the estimation of disability from all causes, and the existence of comorbidity. (Barendregt, Bonneux, & van der Maas, in press) These problems are related: an 'all causes' disability cannot be arrived at by simply adding the disability caused by individual diseases because of comorbidity. The same goes for obtaining an estimate of 'all other' disability by subtracting disease specific disability from an 'all causes' disability.

In the previous paper we ignored the problem, in the present one we try to get an impression of the magnitude of the error this introduced, relative to the errors due to the other elements in the estimation. To this end we modeled comorbidity explicitly in the multi-state life table, compiled a list of all the elements in the estimation procedure that are surrounded by uncertainty, and performed a sensitivity analysis of the outcomes to variations in these elements. All outcomes are expressed in 'years lived with disability', which form the basis of the DALE calculation.

First we compared the estimates of years lived with disability with and without explicit comorbidity (table 3). Disease specific years lived with disability are considerably lower when comorbidity is taken into account, but the largest part of these disability years are assigned to comorbidity. On balance disability explained by the diseases including comorbidity is 7.5% lower, causing the 'all other' disability estimate to be 3% higher. This lower explained disability and higher 'all other causes' disability in turn make the smoking intervention less effective: total disability years decline by 8% instead of 10% after smoking abolishment.

The list of uncertain elements in the DALE estimation procedure includes the 'all causes' disability estimate, disease epidemiology, clustering of diseases that determines the prevalence of comorbidity, the disability weights from the Global Burden of Disease project, the disability weights in case of comorbidity, and, when evaluating the smoking intervention, the impact of smoking abolishment on disease epidemiology.

The results of the sensitivity analysis in figures 1 and 2 show that the disability weights used for comorbidity have an impact on the amount of

disability due to comorbidity, and through that on the amount of disability explained by the diseases, and 'all other' disability. The estimation of years lived with disability (figure 1) is most sensitive to respectively the estimate of total disabled years, of the disability weights, the amount of comorbidity, and the epidemiology of the individual diseases. The influence of different comorbidity weights is relatively small.

For the intervention impact (figure 2) the result is similar, with the qualification that the effect of different comorbidity weights on total disability after the intervention is, although still small, relatively larger. Because higher disability weights for comorbidity make the total amount of disability explained by the diseases modeled higher, the outcome on total disability amounts to an increased effectiveness of the intervention.

The overall conclusion is that the comorbidity weights are just one of a range of uncertain elements in the DALE estimation, and one with a relatively small effect on the outcome too. This despite the fact that the diseases modeled are mostly high prevalence diseases (the exception is lung cancer), with consequently a large amount of comorbidity. But even with these high prevalence diseases the amount of total disability explained is only about 27%, leaving a large amount of 'all other causes' disability. When the part of total disability explained rises, through the modeling of more diseases, the impact of the comorbidity weights will presumably become larger, because comorbidity prevalence will be higher.

Nevertheless, given the current number of diseases modeled, comparing the present results with the earlier estimates (table 3) suggests that in this case ignoring comorbidity altogether will not have a large impact on the estimate of total disability caused by the diseases, or on the disability avoided through an intervention.

Ignoring comorbidity is an attractive option, because it avoids complex modeling. Obviously it is not a feasible option when comorbidity itself or the disability it causes is the subject of study, but when the focus is on disability caused by diseases and on total disability it seems an appealing choice. In case comorbidity is included explicitly, without being the study subject itself, simple assumptions on comorbidity disability weights will serve, because the impact on DALE estimates will be minor.

References

- Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1997). Health Expectancy: from a population health indicator to a tool for policy making. *Journal of Aging and Health*. (in press)
- Barendregt, J. J., van Oortmarssen, G. J., van Hout, B. A., van den Bosch, J. M., & Bonneux, L. (1997). Coping with multiple morbidity in a life table. *Mathematical Population Studies*, 7(1).

- Bonneux, L., Barendregt, J. J., Meeter, K., Bonsel, G. J., & van der Maas, P. J. (1994). Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J of Public Health*, 84, 20-28.
- van den Bos, G. A. M., van der Velden, J., van Sonsbeek, J. L. A., Nusselder, W. J., & Lenior, M. E. (1994). *Gezonde levensverwachting en chronische ziekten: vergelijking van databronnen en berekeningen*. ISG,NIVEL,CBS.
- Eindhoven Cancer Registry (1991). *Cancer incidence and survival 1975-1987 in Southeastern Netherlands*. Comprehensive Cancer Center South.
- Foundation of Public Health and Smoking (1993). *Annual Report 1992*. The Hague.
- van de Mheen, P. J., & Gunning-Schepers, L. J. (1996). Differences between studies in reported relative risks associated with smoking: an overview. *Public Health Reports*, 111, 420-426.
- Murray, C. J. (1994). Quantifying the burden of disease: the technical basis for disability-adjusted life years. *Bull World Health Organ*, 72(3), 429-45.
- Murray, C. J. L., & Lopez, A. D. (Ed.). (1994). *Global comparative assessments in the health sector. Disease burden, expenditures and intervention packages*. Geneva: World Health Organization.
- Niessen, L. W., Barendregt, J. J., Bonneux, L., & Koudstaal, P. J. (1993). Stroke Trends in an Aging Population. *Stroke*, 24, 931-939.
- Statistics Netherlands (1991). Life tables for men by age, 1986-1990. *Maandstatistiek bevolking*(12), 76.
- Statistics Netherlands (Published annually). Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1. Voorburg: CBS.
- Verbrugge, L. M., Lepkowski, J. M., & Imanaka, Y. (1989). Comorbidity and its impact on disability. *Milbank Q*, 67(3-4), 450-84.
- van Weel, C., van den Bosch, W. J. H. M., van den Hoogen, H. J. M., & Smits, A. J. A. (1987). Development of respiratory illness in childhood - a longitudinal study in general practice. *J R Coll Gen Pract*, 37, 404-408.

Update and summary

This part has looked at population health indicators in general, and at three instances (Health Expectancy, DALY and DALE) in particular. The research field is vivid, and work and debate on the matters addressed in the articles reproduced here continues. In this chapter we give an update of the latest results and summarize our conclusions.

Health Expectancy

In chapter 16 we explain that measuring trends in health expectancy with an indicator based on a prevalence measure of morbidity encounters a problem because of the combination of stock and flow variables. The seriousness, if not the existence, of this problem is disputed. Some authors argue that the prevalence based indicator only fails in extreme circumstances, and that under circumstances typical of normal populations the prevalence based indicator is acceptable for trend monitoring (van de Water, Boshuizen, Perenboom, Mathers, & Robine, 1995). We hold that the example of the change in survival after myocardial infarction through the adoption of thrombotic therapy constituted one of these extreme circumstances, and that other large changes in incidence and survival are known to have occurred recently (Barendregt, Bonneux, & van der Maas, 1995).

Mathers and Robine showed that a prevalence based health expectancy indicator was able to largely follow an incidence/duration indicator even after trends in the incidence of disability (Mathers & Robine, 1997). However, the simulation that supported this claim artificially removed the aspect which caused most trouble in the simulation we present in chapter 16: the effect that a change in survival (flow variable) immediately changes the estimate of life expectancy, but only slowly morbidity prevalence (stock variable) (Barendregt, Bonneux, & van der Maas, 1997).

In addition we argue that in the compression and expansion discussion the variable of interest is not health expectancy as such, but the difference between life expectancy and healthy life expectancy. Since this difference is much smaller than health expectancy but affected by the same absolute bias, the relative bias becomes considerable. Observed changes over time in life expectancy with morbidity are usually small, and could therefore be affected to a large extent by this bias (Barendregt, Bonneux, & van der Maas, 1997; Mathers, 1994).

Summarizing our findings on health expectancy as a Population health status indicator we can say that it is attractive because of its uncomplicated methodology and data requirements, that estimates of current Population health status will be acceptable, but that methodological problems exist when it is used to evaluate changes over time or within disease processes (which is what compression and expansion of morbidity is about). Solutions for these problems can be found (for example using an incidence/duration approach, or adopting more elaborate morbidity weights), but carry the cost of increased complexity and data requirements.

DALYs

In chapter 18 we have shown that the age weighting schedule of the DALY, contrary to appearances, emphasizes years lost due to mortality between the ages of 0 and 27. Chapter 18 also claims that the same is true for years lost due to disability. In a response Murray and Lopez point out this depends on the duration of the disability: for chronic disability the emphasized age range would indeed be 0 to 27, but as the disability duration declines this range changes to the 10 to 55 range. They give the example of a disability with a 5-year duration, that stresses ages between 7 and 52 (Murray & Lopez, 1996).

We make two observations here. The first is that this makes the age-weighting schedule for the years lost due to disability a rather complex function of the duration of disability, which, given different durations of disability, results in a different schedule for each condition. The aggregate age weighting in the DALY thus becomes an amalgam of a large number of different age-weight schedules, with, because it is dominated by the years lost due to mortality, the highest weight probably at very young age. This amounts to a regrettable loss of transparency for the rather intuitive original idea that life years lost at very young or old ages are not as severe as at adolescence and middle age.

The second observation is that the shift towards younger ages of the emphasized age range is the consequence of first applying the age-weighting schedule, and next the *years of life lost* calculation. This order of calculation may make sense when *years of life lost* are indeed interpreted as years not lived. But when *years of life lost* is interpreted as an age-weighting schedule itself it does not describe a number of years that should have been lived at specific ages, but simply a weight to be applied to a death at a certain age.

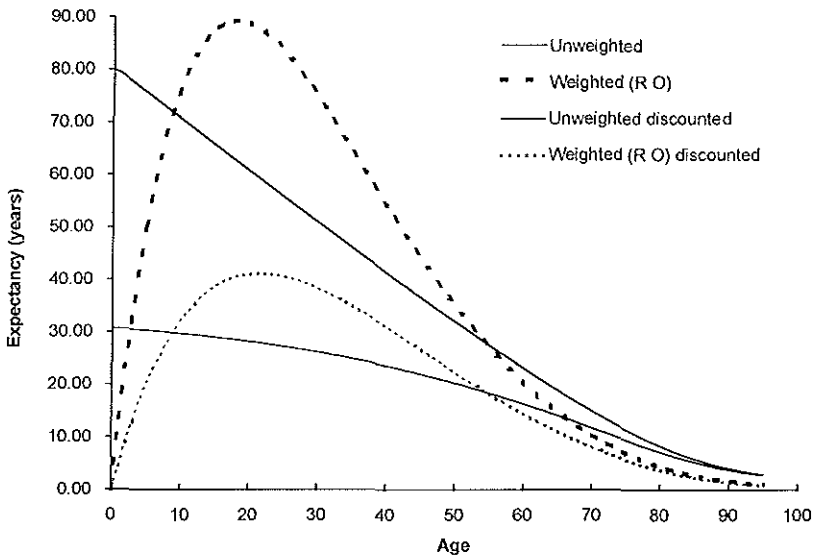


Figure - Age specific life expectancies for Dutch women (1986-1990), without the additional age weight schedule, and with the additional schedule applied after the years of life lost calculation (denoted with "R.O.", or reverse order). Both discounted (3%) and undiscounted variants are shown.

Combining the two age-weighting schedules would then mean multiplying the two weights for that age, or, which amounts to the same, first calculate *years of life lost*, and then apply the age-weighting schedule. The resulting curve is in the figure (compare the figures in chapter 18), which seems to reflect much better the original idea of the age weights.

A similar approach could be used for disability: apply only the age-weight of the age at incidence to future not age-weighted life years lived with disability. While this would solve the problem of a multitude of different age-weighting schedules, it requires a different interpretation of the age-weights than was used in the original DALY. At present Murray and Lopez do not seem inclined to make any changes (Murray, 1996; Murray & Lopez, 1996).

Summarizing we conclude that the DALY is a complex population health status indicator, even more complex than would seem at first glance. It is geared to be an indicator for optimal allocation of resources, witness its use of a real population and discounting. This use also necessitates a disease specific approach, and the establishment of a chain of causation running from disease incidence to prevalence, disability prevalence, and mortality.

DALE

A major problem with the Disability Adjusted Life Expectancy (DALE), as presented in this part, remains the estimation of 'all other disability'. The DALE describes only a few diseases and their disability explicitly, and uses a generic 'all other disability'. To estimate this 'all other disability' an estimate of 'total disability' is needed. In the absence of an estimate based on the disability weights from the Global Burden of Disease project, an alternative method was presented in chapters 19 and 20. An estimate of 'total disability' was obtained by setting the DALE equal to the Disability Free Life Expectancy (DFLE), based on the Health Survey.

In a recent article Murray and Lopez presented DALE estimates for each of the eight world regions they distinguish, using the full set of conditions studied in the Global Burden of Disease study (Murray & Lopez, 1997). The male life expectancy in the Established Market Economies (EME) is 73.4, virtually the same as the one we obtained in chapter 19. The DALE for the EME, on the other hand, was estimated at 67.4, while our DFLE-based DALE is set at 64.7.

In order to see whether a higher DALE would change the results of chapter 19, we recalculated the effect of smoking abolition with the higher DALE estimate. It turns out that the abolition of smoking in that case would have a stronger beneficial effect. When total disability is lower, while the disability of smoking related diseases remains the same, 'all other' disability will be lower. With a lower 'all other' disability the substitution effect of the increased life expectancy will be lower, and the compression becomes stronger. While this outcome is reassuring, it also stresses that for this kind of assessment a direct estimate of the Dutch DALE is highly desirable.

In summary we conclude that the DALE, although not without problems that still await solutions, is an attractive indicator of population health status, that has already found wider application. The attractiveness is based on several characteristics. The DALE, as a health expectancy, allows comparisons over time and place. It is simpler than the DALY because it does not use the age-weighting schedule and discounting, but it retains the essential property of linking specific diseases to overall population health status. We expect to see more applications of the DALE in the future.

References

- Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1995). Reply to Van de Water et al. *J Epidemiol Comm Health*, 49, 330-331.

- Barendregt, J. J., Bonneux, L., & van der Maas, P. J. (1997). How good is Sullivan's method for monitoring changes in population health expectancies? (letter). *J Epidemiol Comm Health*, 51, 578.
- Mathers, C. D. (1994). *Health Expectancies in Australia 1993: preliminary results*. Paper presented at the Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.
- Mathers, C. D., & Robine, J.-M. (1997). How good is Sullivan's method for monitoring changes in population health expectancies. *J Epidemiol Comm Health*, 51, 80-86.
- Murray, C. J. (1996). Rethinking DALYs. In C. J. Murray & A. D. Lopez (Ed.), *The Global Burden of Disease* (Vol. I, pp. 1-98). Boston: Harvard School of Public Health.
- Murray, C. J., & Lopez, A. D. (1996). The incremental effect of age-weighting on YLLs, YLDs, and DALYs: a response [comment]. *Bull World Health Organ*, 74(4), 445-6.
- Murray, C. J. L., & Lopez, A. D. (1997). Regional patterns of disability-free life expectancy and disability-adjusted life expectancy: Global Burden of Disease Study. *Lancet*, 349, 1347-52.
- van de Water, H. P. A., Boshuizen, H. C., Perenboom, R. J. M., Mathers, C. D., & Robine, J. M. (1995). Health expectancy: an indicator for change? *J Epidemiol Comm Health*, 49, 330.

PART V

*Population health status,
prevention,
and health care costs*

Introduction to Part V

Introduction

Compression and expansion of morbidity concern in the first place the future health of populations, but in the debate the thought about the associated costs of health care has never been far away. In the expansion scenario the pandemic of chronic disease is expected to cause an explosion of health care costs, while on the other hand compression, in particular the compression to be achieved by successful prevention, is expected to offer potential savings (Fries, et al., 1993). In this part we take a look at the issue of compression, expansion, and health care costs, without in any way pretending to be comprehensive. Indeed, we primarily want to show that relations between population health status and costs of care are not as straightforward as they may seem, and that this is a topic of research in need of further attention.

Chapters 23 and 24 both look at the prevention of disease and its impact on health care costs. Prevention of disease offers an interesting case: there is little doubt that expansion of morbidity through increased survival with chronic disease will cause an increase in health care costs. But with prevention it may go either way: on the one hand successful prevention will lower morbidity, and hence tend to lower health care costs, but simultaneously it may extend length of life, and thus increase costs. The net effect depends on the balance between these opposing tendencies (and of course on the cost of prevention itself, which may be substantial but, for simplicity's sake, is being ignored).

Chapter 23 makes it clear that a disease specific approach in these matters is necessary, even if only with the broad disease categories applied there. Within the framework of an extended cause elimination life table, extended because it eliminates diseases not just as a cause of death but as a cause of costs as well, it is shown that the effects of elimination differ by

disease. The disease categories can be distinguished in two classes: fatal and non-fatal diseases. The general conclusion is that there only a few cases of largely non-fatal diseases in which elimination of the disease will actually lower health care costs.

In chapter 24 we look at the relation between health care costs and the largest avoidable risk factor for population health status: tobacco smoking. The impact of smoking on health care costs is examined by estimating three life tables, one for a mixed smokers/non-smokers' population, one consisting of smokers, and one of non-smokers. In addition we use a dynamic method to estimate the effects of smoking cessation on costs.

Results show that smokers are more expensive when alive, but a non-smokers' population would be more expensive than the current mixed population of smokers and non-smokers. Nevertheless, given a short enough follow-up period and a high enough discount rate, it would be economically beneficial if smoking is abolished.

Both these chapters use an interdisciplinary approach: methods from epidemiology, demography and economics are employed. The combined use of these methods is a source of much confusion, witness the referee comments we collected on these two papers. We will briefly review the main issues.

Time and age

In this part the demographic technique of the life table is used to allow for substitution of causes of death: elimination of a cause of death or a risk factor will, in the end, not result in fewer deaths. All people die, but when a cause of death is eliminated or lowered they will do so from other causes, and at a higher age. The life table allows to calculate at what age.

The life table is a well-established technique, with various extensions such as the cause elimination life table of chapter 23 and the multi-state life table we use in chapter 24 (Schoen, 1987). Two types of life table can be distinguished: the 'generation' life table, and the 'period' (or 'current') life table. The generation life table follows a single birth cohort using historical data, thus reproducing the actual mortality experience of that cohort. Interesting as this may be, for purposes of describing the mortality experience of the current population, this will not do.

The period life table therefore uses currently observed mortality probabilities, which makes it better suited to describe current mortality, but at the same time makes its interpretation less clear-cut. The period life table has two, mutually exclusive interpretations, depending on how the age axis is looked upon. One interpretation is that the life table represents a single cohort, followed into the future while it ages, the second that it stands for an entire population at one point in time, the so called 'stationary population' interpretation (Shryock & Siegel, 1976). In the latter interpretation the age axis represents just age, in the former both age and time. Describing an entire

population through time requires separate age and time axes, for example a series of life tables, one for each point in time.

Much of the methodological confusion mentioned originates from the application of discounting, in particular in combination with life table methods. Discounting is an economic technique to help in the evaluation of future costs and benefits, monetary or otherwise. This evaluation is hampered by the propensity to enjoy money (or goods) now rather than later, and to pay later rather than now (Drummond, Stoddart, & Torrance, 1990). This propensity, called 'time preference' by economists, makes it difficult to decide which of two different streams of future costs and benefits is the most attractive.

To decide which of these different streams of future costs and benefits is preferred the streams are represented by their Present Discounted Values (PDV's): a weighted sum of the future costs and benefits, with an exponentially lower weight as the cost or benefit is further away in time (Branson, 1979). The calculation requires assumptions to be made about how strong the time preference is (expressed in the value of the discount rate) and how long the follow up must be (the time horizon, see the appendix for the equation used). The procedure makes, assuming that time preference is adequately represented by such an exponential function, the two future streams of costs and/or benefits that differ in timing comparable, and allows to choose which is preferred.

More than one referee of the papers in chapters 23 and 24 criticized us for not using discounting in our life table estimates. We think they are mistaken, and for two reasons:

- 1) The research questions are about estimation: 'what would health care costs be if a disease or smoking would not exist'. Applying discounting to such research questions confuses estimation and evaluation.
- 2) The research questions are about populations. This implies that we use the stationary population interpretation of the life tables, with the age axis representing only age and not time. Applying discounting would therefore amount to correcting for age preference (costs made at old age are less important), not time preference.

The only exception is the last research question of chapter 24, which looks at what would happen with population health care costs after an intervention would succeed to convince all current smokers to quit. To answer this question a dynamic population model is used, with separate age and time dimensions, and not a life table. To see whether the resulting stream of future health care costs is more attractive than the one without the intervention, we then apply discounting, and conclude that, despite the eventually higher health care costs, the intervention is indeed economically attractive with reasonable values for discount rate and time horizon.

Appendix

A decision maker is confronted with two investment projects that differ both in timing of future costs and revenues as in amounts. The problem for the decision maker is which project, if any, to choose. To be able to make this decision the two streams are converted to Present Discounted Values (PDV's) using:

$$PDV = \sum_{t=0}^{t=t_H} \frac{P_t}{(1+r)^t}$$

Where:

- t : time, with $t=0$ the present.
- t_H : time horizon of evaluation.
- P_t : amount of cost or benefit at time t .
- r : discount rate.

The decision maker maximizes profit when he picks the project with the highest PDV, and never chooses a project with a negative PDV (because that project would make a loss). Note that there are two unknowns in the equation: the time horizon t_H and the discount rate r . While in a commercial enterprise obvious values for these unknowns may exist, like amortization period and market rate of interest, this is less obvious for medical interventions (Drummond, Stoddart, & Torrance, 1990).

References

- Branson, W. H. (1979). *Macroeconomic Theory and Policy*. (2nd ed.). New York: Harper & Row.
- Drummond, M. F., Stoddart, G. L., & Torrance, G. W. (1990). *Methods for the Economic Evaluation of Health Care Programmes*. Oxford: Oxford University Press.
- Fries, J. F., Koop, C. E., Beadle, C. E., Cooper, P. P., England, M. J., Greaves, R. F., Sokolov, J. J., & Wright, D. (1993). Reducing health care costs by reducing the need and demand for medical services. The Health Project Consortium. *N Engl J Med*, 329: 321-5.
- Schoen, R. (1987). *Modelling Multigroup Populations*. New York and London: Plenum Press.
- Shryock, H. S., & Siegel, J. S. (1976). *The methods and materials of demography, condensed edition*. Orlando: Academic Press, Inc.

Preventing fatal diseases increases health care costs: a cause elimination life table approach

Abstract

Objectives: *To demonstrate that elimination of fatal diseases will increase health care costs.*

Design: *Mortality data from vital statistics are combined with health care spending in a cause elimination life table. Costs are allocated to specific diseases through the various health care registers.*

Setting and subjects: *The population of the Netherlands, 1988.*

Main outcome measures: *Health care costs of a synthetic life table cohort, expressed as life time expected costs.*

Results: *The life time expected health care costs for 1988 in the Netherlands were £ 56,600 for men and £ 80,900 for women. Eliminating fatal diseases, such as coronary heart disease, cancer or chronic obstructive lung disease increases health care costs. Major savings will be achieved only by eliminating non-fatal disease, such as musculoskeletal diseases and mental disorders.*

Conclusion: *The aim of prevention is to spare human beings from avoidable misery and death, not to save money on the health care system. In low mortality countries, eliminating fatal diseases by successful prevention increases health care spending because of the medical expenses during added life years.*

Introduction

In low mortality countries, health care costs - which are already substantial - are on the rise. Health promotion, based on the simple idea that by preventing illness, illness-related costs will be prevented, has been hailed as the solution by some.(1) However, it is doubtful whether eliminating fatal diseases would cause a decrease in chronic morbidity. Health care needs terminate at death. Surviving means ageing, the strongest determinant for diseases such as osteoarthritis, osteoporosis and related fractures, cognitive decline or loss of vision or hearing.

In demography, the effects of eliminating diseases are studied by cause elimination life tables.(2) In such life tables, a cause of death is eradicated, and the life expectancy recalculated. By linking disease-specific health care costs to the life table population, the consequences of having eliminated a disease as a cause of death, as well as of a source of costs, can be estimated.

Methods

In a previous study, all health care costs in the Netherlands in 1988 (39,800 million Dutch guilders (*f*), \pm 11 400 million pounds sterling (£) at the exchange rate of 1988, for 14.8 million inhabitants) were allocated to age, sex, health care sector and primary diagnosis, based on comprehensive data on morbidity, mortality and direct costs.(3-5) Some 25.2% of all costs were unable to be allocated to a primary diagnosis, either through a lack of information or because of the non-personal nature of these costs (administration, general public health services, etc).

The health care costs are linked to the Dutch period life table for men and women of 1986-1990.(6) The expected costs represent the total costs of the life table cohort during their lifetime. These costs are divided by the initial size of the life table cohort, i.e. "the population at birth", to yield the lifetime expected costs for an individual (see appendix). The interpretation of life time expected costs is analogous to the interpretation of life expectancy. The imaginary life table cohort is subjected to the unchanging costs of 1988 and the unchanging death rates of 1986-1990 until extinction. The life expectancy is the sum of all the years a life table person is expected to live, the life time expected costs are the sum of all the health care costs that person is expected to incur during these life years. To calculate the effect of eradication, a specific disease is eliminated both as cause of death and as cause of costs: the cause elimination life table recalculates life expectancy and lifetime expected costs as if the eliminated disease had never existed. A life table assumes that persons lived on average for a half year in the year of death. In the year that an eliminated cause of death would have occurred, persons are now at risk of death from other causes during the entire year (instead of a half year), and remain fully at risk of other causes for the added life years. Only allocated costs are considered in the cause

elimination life tables. Because cause elimination life tables are interpreted as stationary populations before or after the elimination of a disease,⁽⁷⁾ the costs are not discounted. Cause-specific mortality data are available for 5 year age groups till the age of 84.⁽⁶⁾ After the age of 85, we assume the cause-of-death ratio (deaths from the specific cause divided by all deaths) and the costs to remain constant with rising age. This assumption underestimates the cost of added life years in the very old. Specific causes of death are underregistered at older ages and health care costs increased steadily with age.^(3-5, 8)

Results

Life expectancy in the Netherlands in 1986-1990 was 73.5 years for men and 80.0 years for women. The lifetime expected costs at birth for all health care totalled *f* 198,000 (£ 56,600) for men and *f* 283,000 (£ 80,900) for women. Some *f* 155,000 (£ 44,300; men) and *f* 219,000 (£ 62,600; women) were allocated to a primary diagnosis (see table).

The table shows the inverse relationship between fatality and costs: the highly lethal coronary heart diseases, causing nearly 19% of all deaths, cause only 2.7% of all health care costs. Mental disorders, including psychiatric diseases, mental handicaps and dementia, are together responsible for only 0.6% of all deaths, but consume 26% of the allocated health care budget. Eliminating coronary heart disease would substantially increase the burden on the health care budget, as this would save few costs but add a considerable number of life years. Indeed, life expectancy would increase by about 1.9 years (2.5%), while costs would jump 6%. On the other hand, eliminating dementia would cause no noticeable change in life expectancy, but would save 6% on the health care budget.

Diseases whose elimination augments health care costs substantially are coronary heart disease, cancer and chronic obstructive lung diseases: the present targets of health promotion. The savings yielded by eliminating stroke and heart failure-related costs outweigh the costs associated with gains in life expectancy. Cancer is more fatal among men than among women; consequently, the elimination of cancer would add more life years, and therefore more costs, among men.

Eliminating accidents and other unnatural causes of death adds life years and saves costs, in the light of both the high burden of morbidity and mortality. But the table shows that the largest gains are to be achieved through the elimination of mental disorders and musculoskeletal diseases: limited sources of lost life years, but major sources of costs.

Discussion

This paper shows that lengthening life generally will increase health care needs, particularly needs for long term nursing care, as most life years are

Table

All deaths from 1986-1990 (6) and all allocated costs (in 1000 pounds) from 1988 (5) caused by various disease groups and the effects of elimination of these disease groups on life expectancy and lifetime expected costs. The numbers refer to the international classification of diseases, 9th revision.(14) The percentages in the life expectancy columns show the relative change as a consequence of elimination.

MEN	all deaths (1986-90)		all allocated costs (1000 £; 1988)		life expectancy after elimination		Life time expected costs (£) after elimination	
none	330717	(100%)	3755	(100%)	73.5	(100%)	44011	(100%)
cardiovascular diseases (390-459)	131979	(39.9%)	480	(12.8%)	78.8	(7.1%)	46301	(5.2%)
coronary heart disease (410-414)	69624	(21.1%)	153	(4.1%)	76.0	(3.4%)	46674	(6.0%)
stroke (430-438)	24524	(7.4%)	126	(3.4%)	74.2	(0.9%)	43477	(-1.2%)
heart failure (428-429)	9571	(2.9%)	51	(1.4%)	73.8	(0.4%)	43762	(-0.6%)
cancer (140-208)	101309	(30.6%)	233	(6.2%)	77.5	(5.3%)	47657	(8.3%)
lung cancer (162)	36859	(11.1%)	49	(1.3%)	74.8	(1.8%)	45683	(3.8%)
colorectal cancer (153-154)	9510	(2.9%)	26	(0.7%)	73.8	(0.4%)	44201	(0.4%)
respiratory disease (460-519)	29257	(8.8%)	196	(5.2%)	74.3	(1.1%)	43257	(-1.7%)
chronic obstructive lung dis. (490-496)	20154	(6.1%)	77	(2.0%)	74.1	(0.7%)	44150	(0.3%)
unnatural causes of death (E800-E999)	15782	(4.8%)	205	(5.5%)	74.4	(1.2%)	42617	(-3.2%)
traffic accidents (E800-E848)	5212	(1.6%)	58	(1.5%)	73.9	(0.5%)	43734	(-0.6%)
mental disorders (290-316)	1406	(0.4%)	1025	(27.3%)	73.6	(0.06%)	33256	(-24.4%)
dementia (290)	870	(0.3%)	84	(2.2%)	73.6	(0.02%)	42636	(-3.1%)
musculoskeletal diseases (710-739)	1120	(0.3%)	300	(8.0%)	73.6	(0.04%)	40812	(-7.3%)

WOMEN

none	298569	(100%)	4930	(100%)	80.0	(100%)	62292	(100%)
cardiovascular diseases (390-459)	126988	(42.5%)	502	(10.2%)	85.0	(6.3%)	68964	(10.7%)
coronary heart disease (410-414)	48279	(16.2%)	79	(1.6%)	81.5	(2.0%)	65994	(5.9%)
stroke (430-438)	35609	(11.9%)	201	(4.1%)	81.0	(1.3%)	62263	(-0.05%)
heart failure (428-429)	12163	(4.1%)	71	(1.4%)	80.3	(0.4%)	62176	(-0.2%)
cancer (140-208)	76177	(25.5%)	282	(5.7%)	83.4	(4.2%)	66339	(6.5%)
lung cancer (162)	5650	(1.9%)	10	(0.2%)	80.2	(0.3%)	62717	(0.7%)
colorectal cancer (153-154)	10642	(3.6%)	32	(0.6%)	80.3	(0.5%)	62810	(0.8%)
respiratory disease (460-519)	19152	(6.4%)	154	(3.1%)	80.5	(0.7%)	62101	(-0.3%)
chronic obstructive lung dis. (490-496)	7928	(2.7%)	79	(1.6%)	80.2	(0.3%)	62340	(0.1%)
unnatural causes of death (E800-E999)	11409	(3.8%)	280	(5.7%)	80.5	(0.7%)	59183	(-5.0%)
traffic accidents (E800-E848)	2168	(0.7%)	39	(0.8%)	80.1	(0.2%)	62045	(-0.4%)
mental disorders (290-316)	2687	(0.9%)	1220	(24.8%)	80.0	(0.08%)	46853	(-24.8%)
dementia (290)	2492	(0.8%)	301	(6.1%)	80.0	(0.07%)	57185	(-8.2%)
musculoskeletal diseases (710-739)	3084	(1.0%)	486	(9.8%)	80.1	(0.1%)	56165	(-9.8%)

added to old age. This is not a bad thing; prevention can hardly be blamed if it reaches its target, and lowers mortality. Life saving therapy, such as antibiotic treatment of severe infections or oral rehydration in severe diarrhoea, has the same consequences. Indeed, we firmly believe that primary prevention, such as an effective anti-tobacco policy, is an excellent buy.

But that does not imply that no bill has to be paid. Our paper contradicts popular belief that prevention might 'prevent' health care costs.(1) Acute medical costs may be averted, but the considerable needs for long term nursing care of frail elderly can but increase.

Eliminating causes in a life table demonstrates an unquestionable truth: we all have to die. If we eliminate a specific cause of death, we simply die later from another. In the meantime we grow older, become generally more disabled and need more care.(9) In the Netherlands, cardiovascular diseases and cancer were jointly responsible for nearly 70% of all deaths, yet accounted for a mere 17% of all health care costs, whereas the largely non fatal diseases of the brain, joints and bones, causing under 2% of all deaths, generated 35% of all costs (see table). If fatal diseases are eliminated, health care costs during the added life years swamp the savings yielded by the eliminated disease, even if the intervention is radical and without extra costs. At any age, the imaginary population of the cause elimination life table spends less health care resources per person, yet total costs increase because more persons remain alive, surviving to older ages where chronic morbidity and demands for health care are highest (see figure).

As the debate has focused on health care costs,(1) we took into account only medical costs. The non-medical costs of added life years, such as pensions and non-medical care for the elderly, would far outweigh any non-medical costs of disease and death.(10)

In view of the fact that our cost data are comprehensive, we were able to consider all health care spending, including long term nursing care.(3-5) Previous studies have shown that in the US, payments for acute medical care are higher in the last year before death, irrespective of age.(11) Our results do not contradict these findings, but demonstrate the high burden of chronic care for non-fatal diseases, which are not determined by death.

Cause elimination life tables are simple mathematical models, based on the assumption of the independence of diseases. Only primary diagnoses are taken into account, while co-morbidity is ignored. In real life, many diseases are not independent, and death is often the end of a complex process. However, insights are garnered from parsimonious theoretical experiments precisely because they simplify a complicated reality. Taking all relevant disease interactions into account would increase data needs and model complexity intolerably, but would not change the conclusions in a meaningful way. Indeed, only when a disease process can be postponed without postponing mortality can morbidity and health care needs be 'compressed' by prevention.(12)

From a humanitarian point of view, life is preferable to death and health to illness.(13) The aim of health care is not to save money, but to

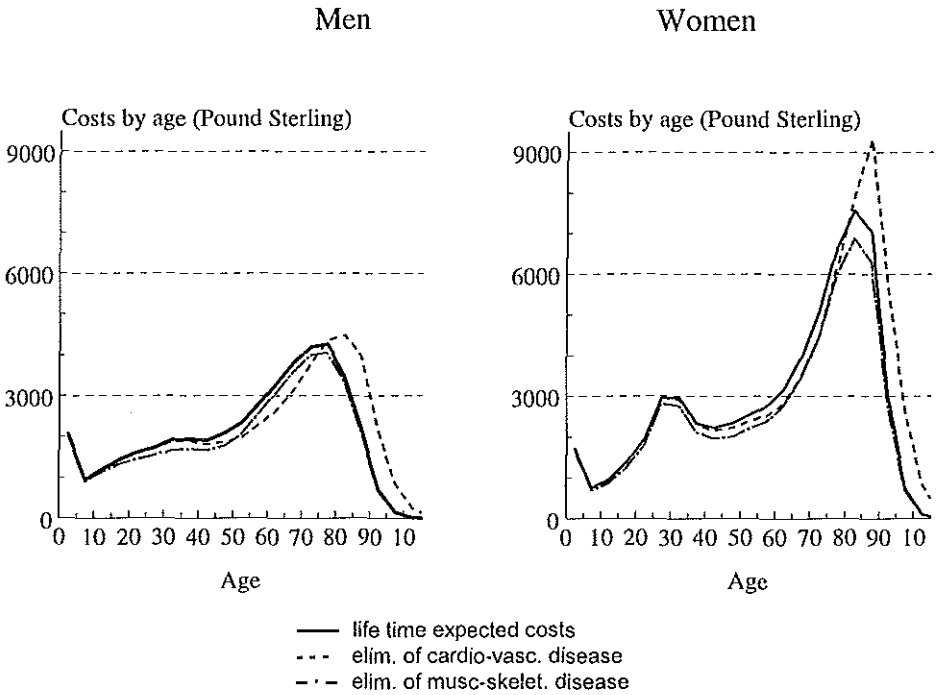


Figure - Costs by age in a life table of men and women, subjected to death rates and to disease allocated costs from 1988. The area under the lines are the life time expected costs. The dashed lines show the effects of elimination of two major disease groups, one fatal (cardiovascular disease), and one non fatal (musculoskeletal affections).

save human beings from preventable suffering and death. Moreover, the medical costs of added life years are trivial. Life extension would cost \pm £ 890 to £ 1400 per life year added, which few would consider unacceptable. But if prevention is used as an argument for constraining future health care expenditures, the medical expenses in the added life years are not insignificant and cannot be ignored. There is no evidence that health care costs are increasing because citizens live unhealthier lives. In fact, quite the contrary would appear to be the case.

We have become increasingly successful at postponing mortality until advanced ages. Old age, however, is associated not only with impending death but also with dementia, social isolation, osteoarthritis, hip fractures, loss of vision and hearing. Even humble progress in disease prevention would have a tangible impact. Any potential savings on health care costs would be a nice icing on that cake.

Appendix

Life time expected costs (LEC) are calculated by

$$LEC = \sum_x \left(\frac{C_x}{N_x} * L_x \right)$$

where C_x is the cost of all causes in the age interval $(x, x+1)$. N_x is the number of persons in the age interval $(x, x+1)$ and L_x , the person-years lived in the age interval $(x, x+1)$.

Eliminating a cause of death is based on the actuarial assumption that persons dying from the specific cause in the age interval $x, x+1$ were considered to have been at a 0.5 year risk of dying from all other causes. The risk of dying at age x from all other causes, adjusted for competing causes, ($q_{x,-\alpha}$) is therefore

$$q_{x,-\alpha} = \frac{d_{x,-\alpha}}{l_x - \frac{1}{2}d_{x,\alpha}}$$

where $d_{x,\alpha}$ is the number of deaths in age interval $(x, x+1)$ due to the cause α , $d_{x,-\alpha}$ is the number of deaths in age interval $(x, x+1)$ due to all other causes than α and l_x is the number of survivors to age x (of the all causes life table cohort). The life table cohort is at risk for this adjusted force of mortality from other causes until extinction.(2)

The expected lifetime costs at birth after elimination of cause α ($LEC_{-\alpha}$) are given by

$$LEC_{-\alpha} = \sum_x \left(\frac{C_{x,-\alpha}}{N_x} * L_{x,-\alpha} \right)$$

where $C_{x,-\alpha}$ represents the costs in the age interval $(x, x+1)$ made for all other diseases than α and $L_{x,-\alpha}$ gives the life years lived in the age interval $(x, x+1)$ after elimination of cause α .

References

1. Fries JF, Koop CE, Beadle CE, Cooper PP, England MJ, Greaves RJ, et al. Reducing health care costs by reducing the need and demand for medical services. *N Engl J Med* 1993;329:321-5.
2. Namboodiri K, Suchindran CM. Life table techniques and their application. Orlando: Academic Press, 1987:92-109.

3. Koopmanschap MA, van Roijen L, Bonneux L. Cost of illness in the Netherlands [in Dutch]. Department of Public Health: Erasmus University Rotterdam, 1991.
4. van Roijen L, Koopmanschap MA, Bonneux L. The cost of illness [in Dutch]. *Ned Tijdschr Geneesk* 1992;136:74-80.
5. Koopmanschap MA, van Roijen L, Bonneux L, Bonsel GJ, Rutten FFH, van der Maas PJ. Costs of diseases in an international perspective. *Eur J Public Health* 1994;4:258-64.
6. Statistics Netherlands. Deaths by causes of deaths, age and gender. Central Bureau of Statistics, Voorburg. Published annually.
7. Shryock HS, Siegel JS. The Life Table. The methods and materials of demography. San Diego/New York: Academic Press Inc. 1976:249-71.
8. Doll R, Peto R. Sources of bias in estimating trends in cancer mortality, incidence and curability. *J Natl Cancer Inst* 1981;66:1270-1281.
9. Nusselder WJ, Van der Velden K, Van Sonsbeek JLA, Lenior ME, Van den Bos GAM. The elimination of selected chronic diseases in a population: the compression and expansion of morbidity. *Am J Public Health* 1996;86:187-94.
10. Russell L. Is prevention better than cure? Washington DC: The Brookings Institution, 1986:116.
11. Lubitz J, Beebe J, Baker C. Longevity and Medicare expenditures. *N Engl J Med* 1995;332:999-1003.
12. Fries JF. Aging, natural death, and the compression of morbidity. *N Engl J Med* 1980;303:130-5.
13. Rose G. The strategy of preventive medicine. Oxford: Oxford University Press, 1992:4.
14. World Health Organisation. International statistical classification of disease, injuries and causes of death. 9th revision. Geneva: World Health Organisation, 1977.



The health care costs of smoking

Abstract

Background: Although smoking cessation is desirable from a public health perspective, its consequences with respect to health care costs are still debated. Smokers have more diseases than nonsmokers, but nonsmokers live longer and can incur more health costs at advanced ages. We analyzed health care costs for smokers and nonsmokers and estimated the economic consequences of smoking cessation.

Methods: We used three life tables to examine the effect of smoking on health care costs - one for a mixed population of smokers and nonsmokers, one for a population of smokers, and one for a population of nonsmokers. We also used a dynamic method to estimate the effects of smoking cessation on health care costs over time.

Results: Health care costs for smokers at a given age are as much as 40 percent higher than those for nonsmokers, but in a population in which no one smoked the costs would be 7 percent higher among men and 4 percent higher among women than the costs in the current mixed population of smokers and nonsmokers. If all smokers quit, health care costs would be lower at first, but after 15 years they would become higher than at present. In the long term, complete smoking cessation would produce a net increase in health care costs, but it could still be seen as economically favorable under reasonable assumptions of discount rate and evaluation period.

Conclusions: If people stopped smoking, there would be savings in health care costs, but only in the short term. Eventually, smoking cessation would lead to increased health care costs.

Introduction

Smoking is a major health hazard, and since nonsmokers are healthier than smokers, it seems only natural that not smoking would save money spent on health care. Yet in economic studies of health care it has been difficult to show who uses more dollars - smokers, who tend to suffer more from a large variety of diseases, or nonsmokers, who can accumulate more health care costs because they live longer.

The Surgeon General reported in 1992 that "the estimated average life-time medical costs for a smoker exceed those for a nonsmoker by more than \$6,000."¹ On the other hand, Lippiatt estimated that a 1 percent decline in cigarette sales increases costs for medical care by \$405 million among persons 25 to 79 years old.² Manning et al. argue that although smokers incur higher medical costs, these are balanced by tobacco taxes and by smokers' shorter life spans (and hence their lower use of pensions and nursing homes).³ Leu and Schaub showed that even when only health care expenditures are considered, the longer life expectancy of nonsmokers more than offsets their lower annual expenditures.⁴

We have analyzed comprehensively the health care costs of smoking. In doing so we have distinguished between the assessment of differences between smokers and nonsmokers and the assessment of what would happen after interventions that changed smoking behavior. Would a non-smoking population have lower health care costs than one in which some people smoke? Are antismoking interventions economically attractive? We sought to answer these questions and to determine the consequences for health policy.

Methods

Analysis of Smokers and Nonsmokers

We examined the effect of smoking in the general population (a mixture of smokers and nonsmokers). We studied the incidence, prevalence and mortality associated with five major categories of disease - heart disease, stroke, lung cancer, a heterogeneous group of other cancers, and chronic obstructive pulmonary disease (COPD). We used data on these diseases, in addition to mortality from all other causes, in an extension of the standard life table, the multistate life table, that includes multiple health states, such as "alive, healthy" and "alive, with heart disease"^{5,6}

Differences in the frequency of the smoking-related diseases between smokers and nonsmokers are commonly expressed as rate ratios. Using these rate ratios, the prevalence of smoking in the population, and the age- and sex-specific disease incidence of the smoking-related diseases in the mixed population of smokers and nonsmokers, we can estimate the incidence of the diseases separately among smokers and nonsmokers.⁶

Table 1
Prevalence of smoking*
(in percent)

Age (yr)	Smokers	
	Male	Female
0-14	0	0
15-19	20	20
20-34	39	37
35-49	42	36
50-64	39	27
65+	34	13

* Data are averages for 1988-1992 in the Netherlands.¹²

Assuming that the relative survival of persons with these diseases is the same among both smokers and nonsmokers, two additional life tables can be calculated - one for smokers and one for nonsmokers. The three life tables differ with regard to the incidence of the smoking-related diseases and therefore in their associated prevalence, disease-specific mortality, and overall mortality. Because of the difference in mortality, more people remain alive in the life table for nonsmokers than in the table for smokers, particularly in the older age groups, and there are corresponding differences in life expectancies.

In constructing the life tables, we used epidemiological data on the incidence and prevalence of the diseases⁷⁻¹⁰, data on mortality from Statistics Netherlands¹¹, data on smoking (Table 1)¹², and rate ratios from an overview of the international literature.¹³ We tested the sensitivity of the analysis by recalculating the life-tables with excess risks (rate ratio -1) that were 50 percent higher and 50 percent lower (Table 2).

The medical costs we used were based on a study that allocated total health care costs in the Netherlands in 1988 (39.8 billion guilders, or \$19.9 billion against the present exchange rate) to categories of age, sex, and disease.¹⁴ We used the Dutch population of 1988 and the prevalence rates of the smoking-related diseases from the life table for mixed smokers and nonsmokers to estimate the cost per case of disease according to age and sex. The remaining costs were assigned to "per capita costs for all other diseases" (in categories according to age and sex) by dividing the costs by the number of people in the category in question. Using the per capita costs

Table 2
Rate Ratios and Sensitivity Range
Associated with Five categories of
Disease

Category	Rate Ratio (Sensitivity range)*	
Heart	3	(2-4)
Lung ca	10	(5.5-14.5)
Stroke	2	(1.5-2.5)
Other ca [†]	2	(1.5-2.5)
COPD [‡]	25	(13-37)

* Rate ratios refer to the rate of the disease in smokers as compared with nonsmokers. The lower and upper bounds of the sensitivity range were calculated as $1+0.5(RR-1)$ and $1+1.5(RR-1)$, respectively, where RR denotes the rate ratio.

† This category includes neoplasms except for stomach, colorectal, lung, breast, prostate, and skin cancers, and benign tumors.

‡ COPD denotes chronic obstructive pulmonary disease.

for each disease and the "all other disease" costs we calculated the health care costs for the populations included in the three life tables.

Assessment of the Effect of Complete Smoking Cessation

The estimated health care cost derived from the life table of nonsmokers can be seen as an estimate of the cost of health care if no one ever smoked. It does not provide an estimate of the health care costs if all smokers stopped smoking. In the latter case, the size of the elderly population would initially be the same as in the mixed population of smokers and nonsmokers. For it to become similar in size to the elderly population among nonsmokers, in which more elderly people are alive, would take several years, even if mortality declined rapidly.

To describe the epidemiological changes and the changes in the population over time, a dynamic model is needed. For this purpose, we needed a series of linked life tables, one for each point in time, with the population at a given age (a) and time (t) depending on the population at age $a-1$ and time $t-1$, and on incidence of disease and the associated mortality between $t-1$ and t . We used the Prevent Plus computer program, which is designed to evaluate interventions concerning risk factor dynamically.^{6,15}

This dynamic analysis produces a projection of future health care costs. To assess the economic attractiveness of an intervention that would make smokers quit, these costs are compared with those expected when no intervention is made. One difficulty in such an evaluation is the fact that most people prefer to receive benefits as soon as possible, and to postpone payments. Economists call this phenomenon "time preference"^{16,17}, and it is taken into account by discounting the future benefits and costs - that is, those further away in time are given lower weights in the overall evaluation.

The degree of time preference is expressed in the discount rate. Typical values range from 0 to 10 percent, with 0 percent meaning that there is no discounting and no time preference and 10 percent meaning that there is strong time preference. Since there is no generally agreed-upon discount rate, we used various rates (0, 3, 5 and 10 percent) in evaluating the intervention.

A second difficulty in evaluating future costs and benefits is deciding how far into the future the analysis should go. There is no generally agreed-upon duration of follow-up time in this type of analysis. For each projection of discounted costs and benefits, we therefore report the duration of follow-up at which the benefits and costs expected in the future exactly balance each other (the break-even year) - the point at which carrying out the intervention is neither more nor less economically attractive than not doing so.

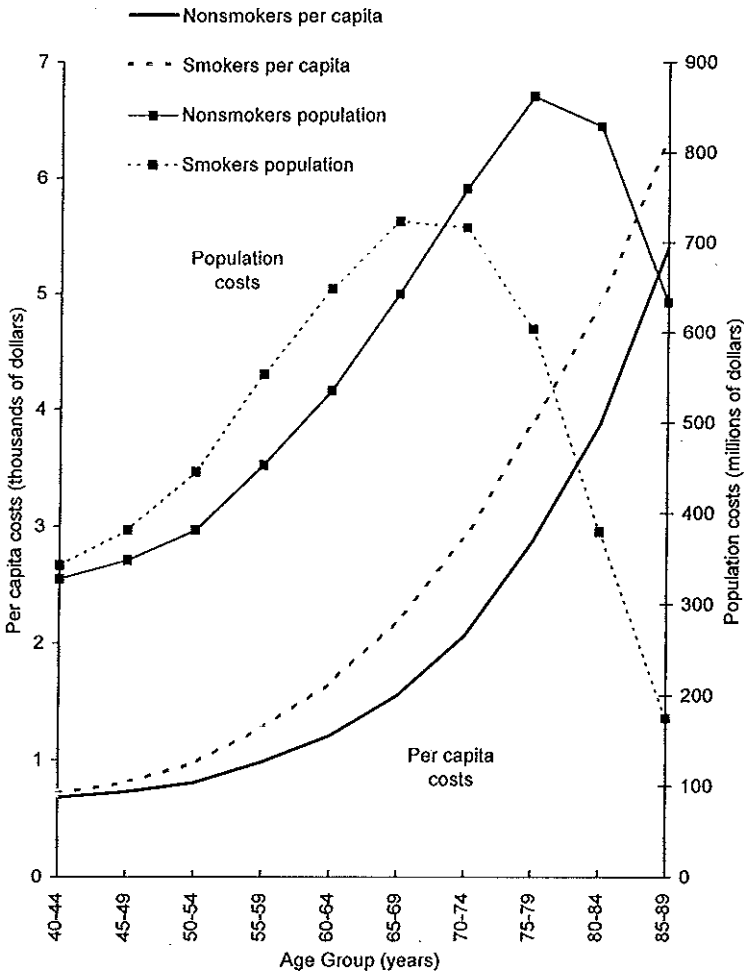


Figure 1 - Estimated Annual per Capita Health Care Costs for Dutch Men in 1988 and for the Male Population in a Life Table, According to Age and Smoking Status. Per capita health care costs for women in the same age groups are very similar to those for men.

Results

Figure 1 shows the annual per capita health care costs for male smokers and nonsmokers 40 to 89 years old, in 5-year age groups (the costs for women in the same age groups are very similar). Per capita costs rise sharply with age, increasing almost 10 times from persons 40 to 44 years of age to those 85 to 89 years of age. In each age group, smokers incur higher costs than nonsmokers. The difference varies with the age group, but

among 65-to-74-year-olds the costs for smokers are as much as 40 percent higher among men and as much as 25 percent higher among women.

However, the annual cost per capita ignores the differences in longevity between smokers and nonsmokers. These differences are substantial: for smokers, the life expectancies at birth are 69.7 for men and 75.6 for women; for nonsmokers the life expectancies are 77.0 and 81.6 (these life table estimates agree very well with the empirical findings of Doll et al.¹⁸). This means that many more nonsmokers than smokers live to old age. At age 70, 78 percent of male nonsmokers are still alive, as compared with only 57 percent of smokers (among women, the figures are 86 percent and 75 percent); at age 80, men's survival is 50 percent and 21 percent, respectively (among women, 67 percent and 43 percent).

These differences in the numbers of elderly people have a profound effect on the health care costs for the population, as Figure 1 shows. In the younger age groups, in which mortality even among smokers is quite low, a population of smokers has higher health care costs than a population of nonsmokers, in the groups of men 70 to 74 and over (and those of women 75 to 79 and over), the lower per capita cost of the nonsmokers is outweighed by the greater number of people remaining alive.

As Figure 1 shows, the nonsmoking population as a whole is more expensive than the smoking population. The area between the curves in which the smokers have higher health care costs than the nonsmokers is smaller than the area between the curves in which the nonsmokers have higher health care costs than the smokers. This is shown in greater detail in Table 3, where the total health care costs for the mixed, the smoking, and the nonsmoking population are presented according to disease category.

All the smoking-related diseases (with the notable exception of stroke among men) are associated with higher costs in a population of smokers, and lower costs in a population of nonsmokers. This relation is particularly strong for the diseases with the highest excess risk: lung cancer and COPD. However, in the mixed population of smokers nonsmokers, smoking-related diseases account for only 19 percent of total costs among men and 12 percent of total costs among women, and the costs of all the other diseases have precisely the opposite relation. In a population of smokers the costs associated with all the other diseases are less than those in the mixed population: 14 percent less for men and 18 percent less for women. Among nonsmokers, the costs of all the other diseases are higher: 15 percent for men and 7 percent for women.

The risk for the diseases not related to smoking is considered equal for smokers and nonsmokers, but the nonsmoking population lives longer and therefore incurs more costs due to those diseases, particularly in old age, when the costs are highest. On balance, the total costs for male and female nonsmokers are 7 percent and 4 percent higher, respectively, than for a mixed population, whereas for smokers the total costs are 7 percent and 11 percent lower.

Table 3 also shows that changing the assumptions about the excess risk associated with smoking-related diseases by as much as 50 percent in

Table 3

Health Care Costs for the Three Populations Studies with Life Tables, According to Sex and Disease Category, with the Ratios of the Costs for Smokers and Nonsmokers to Those for the Mixed Population Containing Both.

Sex and Disease Category*	Population			Smokers: Mixed Population	Nonsmokers: Mixed Population
	Mixed	Smokers	Non-smokers		
	<i>millions of \$</i>			<i>cost ratio (sensitivity range)†</i>	
MEN					
Heart	525	676	371	1.29 (1.18-1.35)	0.71 (0.84-0.61)
Stroke	416	390	428	0.94 (0.91-0.97)	1.03 (1.06-0.98)
Lung ca	114	211	33	1.85 (1.76-1.87)	0.29 (0.44-0.22)
Other ca	226	254	203	1.12 (1.05-1.18)	0.90 (0.97-0.83)
COPD	165	275	23	1.67 (1.72-1.63)	0.14 (0.23-0.11)
Other	6360	5463	7284	0.86 (0.89-0.84)	1.15 (1.09-1.18)
Total	7806	7270	8342	0.93 (0.95-0.92)	1.07 (1.04-1.09)
WOMEN					
Heart	386	538	330	1.39 (1.24-1.48)	0.86 (0.93-0.79)
Stroke	510	571	502	1.12 (1.04-1.19)	0.98 (1.01-0.96)
Lung ca	23	70	9	3.10 (2.61-3.32)	0.39 (0.55-0.31)
Other ca	297	387	264	1.30 (1.16-1.42)	0.89 (0.95-0.84)
COPD	102	254	20	2.48 (2.45-2.45)	0.20 (0.31-0.15)
Other	9358	7653	10013	0.82 (0.87-0.79)	1.07 (1.04-1.09)
Total	10676	9473	11138	0.89 (0.92-0.87)	1.04 (1.03-1.05)

* COPD denotes chronic obstructive pulmonary disease.

† The sensitivity range was calculated with the lower and upper bounds of rate ratios in Table 2. A lower rate ratio reduces the difference between smokers and nonsmokers in the incidence, prevalence, and mortality from smoking-related disease. Therefore, nonsmokers avert fewer cases of smoking-related disease (leading to lower savings) but simultaneously gain less in life expectancy (leading to lower added costs from "other" diseases). For most smoking-related diseases and "other" diseases, lower rate ratios make the difference in costs smaller.

either direction does not change the conclusion, except in the case of stroke. The age-related increase in incidence is steepest for stroke, and there is also an age-related increase for stroke in the cost per case; therefore the health care costs associated with stroke are the most sensitive to changes in life expectancy.

Because of the costs of other diseases, the population of nonsmokers has higher health care costs, partly because these costs increase with age. To test the sensitivity of the analysis to this age-related increase, we recal-

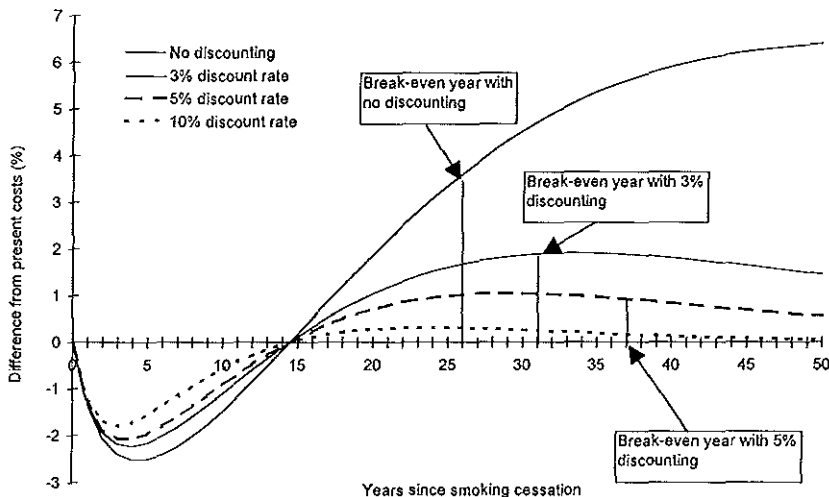


Figure 2 - Percent changes in total health care costs for the male population after smoking cessation, as determined in a dynamic analysis, according the number of years since cessation, with no discounting and with three discount rates.

The labels show the "break-even" years, when the cost and benefit of the intervention balance each other. Shorter follow-up times make smoking cessation attractive economically, and longer follow-up makes it unattractive. With 10 percent discounting, the break-even year is later than 50 years.

culated the three life tables, keeping the health care costs associated with "all other disease" at the 65-to-69-year-old level for people over the age of 65. The costs for the mixed population and for the nonsmoking population became virtually the same, and those for the smoking population were still the smallest, albeit by a small margin.

Figure 2 shows what the economic consequences would be if all smokers stopped smoking. After this abrupt change, the total health care costs for men (the "no discounting" curve) would initially be lower than they would have been (by up to 2.5 percent), because the incidence of smoking-related diseases among the former smokers would decline to the level among nonsmokers. Prevalence rates start to decline, costs decline, and the intervention shows a benefit. With time, however, the benefit reverses itself to become a cost. The reason is that along with incidence and prevalence, smoking-related mortality declines, and the population starts to age. Growing numbers of people in the higher age groups mean higher costs for health care. By year 5, the benefit derived from the presence of the new nonsmokers starts to shrink, and by year 15 these former smokers are producing excess costs. Eventually a new steady state is reached in which costs are about 7 percent higher - the difference between the mixed and nonsmoking populations.

Figure 2 shows the consequences of discounting the projected costs and benefits by various percentages. It is apparent that discounting, even at a rate as low as 3 percent, has a huge impact, and this impact becomes greater as the costs become more distant in time.

Having all smokers quit becomes economically attractive when the future benefits are larger than the future costs or, in terms of Figure 2, when the area below the X axis is bigger than the area above it. From the figure it is clear that this depends heavily on the duration of follow-up time considered and on the discount rate. With a shorter evaluation period and higher discount rates, stopping smoking looks economically more attractive. With a longer evaluation period and lower discount rates, quitting smoking loses its economic advantages. The break-even year, when the initial benefit is exactly balanced by the eventual cost, occurs after 26 years of follow-up when there is no discounting, after 31 years with 3 percent discounting, and after 37 years with 5 percent discounting. At 10 percent discounting, the break-even year occurs after more than 50 years and may not occur at all.

Discussion

This study shows that although per capita health care costs for smokers are higher than those of nonsmokers, a nonsmoking population would have higher health care costs than the current mixed population of smokers and nonsmokers. Yet given a short enough period of follow-up and a high enough discount rate, it would be economically attractive to eliminate smoking.

Some earlier studies have had differing results, partly because many have focused on costs attributable to smoking. From rate ratios and the prevalence of smoking in a population, the proportion of the total number of cases of a disease that can be attributed to smoking, - the population attributable risk - can be calculated.¹⁹ Given the costs according to disease, one can calculate the costs attributable to smoking.²⁰ For instance, in the life table population of mixed smokers and nonsmokers about 8 percent of total health care costs among men and almost 3 percent of total costs among women can be attributed to smoking. Attributable costs, however, can be interpreted as a potential saving only when the diseases do not affect mortality. In the case of most smoking-related diseases, reductions in smoking reduce mortality, creating new opportunities for morbidity from other diseases in the years of life gained.

Other studies of this subject estimate lifetime health care costs, taking the differences in life expectancy into account, and find that smokers have higher medical costs.^{3,21,22} In our study, lifetime costs for smokers can be calculated as \$72,700 among men and \$94,700 among women, and lifetime costs among nonsmokers can be calculated as \$83,400 and \$111,000, respectively. This amounts to lifetime costs for nonsmokers that are higher by 15 percent among men and 18 percent among women.

The studies cited above apply discounting to the lifetime cost estimate. Because costs incurred at older ages are discounted more, this approach reduces lifetime costs for nonsmokers more than those for smokers. For example, when one applies discounting to our life tables for smokers and nonsmokers, smokers have higher health care costs when the discount rate is at least 4.5 percent in men or at least 5.5 percent in women. We disagree with this approach, however. Discounting should be used for purposes of evaluation and should not be applied in a descriptive context, such as the estimation of lifetime costs.

Our analysis is not very sensitive to substantially different values in the rate ratios. Neither is it very sensitive to the age-related increase in the cost of "all other diseases"; that is, an increase that is less steep in the United States than in the Netherlands will not lead to different conclusions. Including additional smoking-related diseases could change results only if they generate morbidity and costs without raising the excess risk of mortality. There may be some of these conditions, such as cataracts, but they are unlikely to change outcome. For example, in our data all eye diseases, most of which are not related to smoking, account for about 1 percent of total health care costs.

This study relied on rate ratios from epidemiologic studies to express the differences between smokers and nonsmokers. To the extent that the rate ratios do not describe these differences sufficiently, the results will be affected. For example, the much lower cost for lung cancer among female smokers than among male smokers (Table 3) is hard to explain physiologically. But as long as the smokers have higher rates of lung cancer than the nonsmokers, such shortcomings of the data will not affect the overall conclusions.

The results of this study illustrate the ambiguities in any economic method of evaluation. Even a well-designed study of this type is marred by inevitable arbitrariness concerning what costs to include, which discount rate to apply, and what duration of follow-up to use. There are differences of opinion - on the discounting of lifetime costs, for example, and the evaluation of long-term effects.^{23,24} Recent efforts at standardization will remedy some of the arbitrariness²⁵⁻²⁷, but fundamental problems with the method still remain.

Finally, with respect to public health policy, how important are the costs of smoking? Society clearly has an interest in this matter, now that several states are trying to recoup Medicaid expenditures from tobacco firms and the tobacco companies have agreed to a settlement. Yet we believe that in formulating public health policy, whether or not smokers impose a net financial burden ought to be of very limited importance. Public health policy is concerned with health. Smoking is a major health hazard, so the objective of a policy on smoking should be simple and clear: smoking should be discouraged.

Since we as a society are clearly willing to spend money on added years of life and on healthier years, the method of choice in evaluating medical interventions is cost-effectiveness analysis, which yields costs per

year of life gained. Decision makers then implement the interventions that yield the highest return in health for the budget.²⁸ We have no doubt that an effective antismoking policy fits the bill.

References

1. MacKenzie TD, Bartecchi CE, Schrier RW. The human costs of tobacco use (2). *N Engl J Med* 1994; 330:975-80.
2. Lippiatt BC. Measuring medical cost and life expectancy impacts of changes in cigarette sales. *Prev Med* 1990; 19:515-32.
3. Manning WG, Keeler EB, Newhouse JP, Sloss EM, Wasserman J. The taxes of sin. Do smokers and drinkers pay their way? [see comments]. *JAMA* 1989; 261:1604-9.
4. Leu RE, Schaub T. Does smoking increase medical care expenditure? *Soc Sci Med* 1983; 17:1907-14.
5. Schoen R. *Modelling Multigroup Populations*. New York: Plenum Press, 1987.
6. Barendregt JJ, van Oortmarsen GJ, van Hout BA, van den Bosch JM, Bonneux L. Coping with multiple morbidity in a life table. *Mathematical Population Studies* (in press).
7. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure. *Am J of Public Health* 1994; 84:20-28.
8. Niessen LW, Barendregt JJ, Bonneux L, Koudstaal PJ. Stroke Trends in an Aging Population. *Stroke* 1993; 24:931-939.
9. Eindhoven Cancer Registry. Cancer incidence and survival 1975-1987 in Southeastern Netherlands: Eindhoven, the Netherlands: Eindhoven Cancer Registry, 1991.
10. van Weel C, B van den Bosch WJHM, van den Hoogen HJM, Smits AJA. Development of respiratory illness in childhood - a longitudinal study in general practice. *J R Coll Gen Pract* 1987; 37:404-408.
11. Overledenen naar doodsoorzaak, leeftijd en geslacht, serie B1. Voorburg, the Netherlands: Centraal Bureau voor de Statistiek, 1989.
12. Foundation of Public Health and Smoking. Annual Report 1992: The Hague, 1993.
13. van de Mheen PJ, Gunning-Schepers LJ. Differences between studies in reported relative risks associated with smoking: an overview. *Public Health Reports* 1996; 111:420-426.
14. Koopmanschap M, van Roijen L, Bonneux L, Bonsel G, Rutten F, van der Maas P. Costs of diseases in an international perspective. *Eur J Public Health* 1994; 4:258-264.
15. Gunning-Schepers LJ, Barendregt JJ, Van der Maas PJ. Population interventions reassessed. *Lancet* 1989;479-481.
16. Drummond MF, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, England: Oxford University Press, 1990.

17. Murray CJ. Rethinking DALYs. In: Murray CJ, Lopez AD, eds. *The Global Burden of Disease*. Vol. I. Boston: Harvard School of Public Health, 1996:1-98.
18. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994; 309:901-11.
19. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic research: principles and quantitative methods*. London: Lifetime Learning Publications, 1982.
20. Shultz JM, Novotny TE, Rice DP. Quantifying the disease impact of cigarette smoking with SAMMEC II software. *Public Health Reports* 1991; 106:326-333.
21. Hodgson TA. Cigarette smoking and lifetime medical expenditures. *Milbank Q* 1992; 70:81-125.
22. Viscusi WK. Cigarette taxation and the social consequences of smoking: National Bureau of Economic Research Working Paper 4891, 1994.
23. Krahn M, Gafni A. Discounting in the economic evaluation of health care interventions. *Med Care* 1993; 31:403-18.
24. West RR. Discounting the future: influence of the economic model. *J Epidemiol Comm Health* 1996; 50:239-244.
25. Mason J, Drummond M. Reporting guidelines for economic studies. *Health Economics* 1995; 4:85-94.
26. Drummond MF, Jefferson TO, on behalf of the BMJ Economic Evaluation Working Party. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996; 313:275-283.
27. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
28. Murray CJ, Kreuser J, Whang W. Cost-effectiveness analysis and policy choices: investing in health systems. *Bull World Health Organ* 1994; 72:663-74.

PART VI

The evolution of senescence

Introduction to Part VI

This paper presents a different outlook to the same problem of the future developments of morbidity. The basic propositions of Fries are that life span is fixed, and that chronic degenerative diseases can be postponed till after the age of 'natural death' (Fries, 1980). Mortality can not decrease anymore, because the population of the USA (and that of the Netherlands) has almost reached the 'natural limits', but degenerative disease can. Others have pointed to the many contradictions in Fries' paper (Manton, 1982). Fries predicted the end of the mortality decline in a period when mortality rates were declining sharply, even at an unprecedented tempo among the elderly and the oldest old. He stated that the population of USA had nearly reached their 'natural' life span, while at the same time he declared that 80% of all causes of death were from chronic degenerative diseases, thus preventable. 'Natural' death was not to be caused by degenerative disorders (but cellular senescence), while those dying at extreme old age suffer characteristically from a wide range of degenerative diseases.

The propositions of Fries should be turned into questions:

- Is life span fixed, and if so, how and at what age?
- What is the relation between chronic degenerative disease and 'natural' death? Why increase both exponentially by age?

These questions are hard to answer with human data. The long life expectancies we enjoy these days are unprecedented in history, and the extreme old are born in a period when mortality was high and life expectancy, although increasing, was still very short. But senescence and senescent death is ubiquitous in all animal species kept under sheltered conditions (Medawar, 1952). Since Darwin, we know that there has to be a good reason for such an ubiquitous characteristic. Indeed, as non-aging individuals seem to have a serious reproductive advantage above aging individuals in

'the struggle for life', there has to be a very good reason indeed why natural selection has never weeded out senescence (Kirkwood, 1985).

In chapter 26, we will review the main evolutionary theories of senescence. The common foundation of these theories is that natural selection has not been concerned with old age, as we did rarely live long enough to get old. We will die anyway, so survival is only useful insofar our offspring benefits. Maintenance mechanisms are tuned to be good enough to get through the normal expectation of life 'in the wild', and to be able to raise healthy offspring, but not so good as to last forever (Kirkwood, 1996).

The evolutionary theory of aging gives answers to the central question of the future evolution of morbidity:

- Is life span fixed? Yes, but it is not fixed by 'hard' programmed limits, but by 'soft' probabilistic limits, a result of genetic variance and stochastic damage (Manton & Tolley, 1991). There is a large variation around an 'optimal' life span (Kirkwood, 1985). Senescent death is not the end result of a genetic program, but of neglect of energy consuming maintenance systems. As a consequence, the numbers of genes and gene interactions involved has to be tremendous, and 'genetic engineering' of the end of life seems an unlikely option (Kirkwood, 1996).
- Is senescent morbidity independent of senescent mortality? This is an obviously untenable assumption. The human organism is a complex multi-component system, each component having its own aging rate (Manton, 1982). These rates are optimised by the same life history, but which still allows large variation in these rates. Senescent death will be determined by the fastest aging rate of a life sustaining component, senescent disability by all components which age faster but are disabling, not fatal. If an effective intervention delays the failure of a life sustaining component, the aging rates of the disabling components may or may not be delayed. If not, disability will replace death.

We conclude that the evolutionary theory as most 'proximal' biological explanation of senescence has to be the background of any future expectation of morbidity in low mortality countries. As long as this theory is standing, forecasts of human (disability adjusted) life expectancy which are inconsistent with the main predictions are to be rejected.

References

- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *N Engl J Med*, 303, 130-5.
- Kirkwood, T. (1996). How can we live forever? *Br Med J*, 313, 1571.

- Kirkwood, T. B. L. (1985). Comparative and evolutionary aspects of longevity. In C. E. Finch & E. L. Schneider (Eds.), *Handbook of the biology of aging* (pp. 27-44). New York: Van Nostrand Reinhold.
- Manton, K. G. (1982). Changing concepts of morbidity and mortality in the elderly population. *Milbank Q/Health Society*, 60, 183-244.
- Manton, K. G., & Tolley, H. D. (1991). Rectangularization of the survival curve. Implications of an ill posed question. *J Aging Health*, 3, 172-93.
- Medawar, P. B. (1952). *An unsolved problem in biology*. London: Lewis.

The expiry date of man

Abstract

In industrialised countries, mortality and morbidity are dominated by age related chronic degenerative diseases. The health and health care needs of future populations will be heavily determined by these conditions of old age. Two opposite scenarios of future morbidity exist: morbidity might decrease ('compress'), because life span is limited, and the incidence of disease is postponed. Or morbidity might increase ('expand'), because death is delayed more than disease incidence.

Optimality theory in evolutionary biology explains senescence as a by-product of an optimised life history. The theory clarifies how senescence is timed by the competing needs for reproduction and survival, and why this leads to a generalised deterioration of many functions at many levels. As death and disease are not independent, future morbidity will depend from duration and severity of the process of senescence, partly determined by health care, palliating the disease severity but increasing the disease duration by postponing death. If morbidity might compress, health care needs will surely expand.

In this article, we discuss expectations of future mortality, morbidity and health care costs in developed countries with high levels of health and health care. The use of such expectations include estimating future demands for health care, identifying the types of services to be provided (e.g. hospital care, nursing care, or other types of institutional care) and determining the prevalence and range of disabilities that exist in a population, to mention a few.¹ Examples abound of the uses of forecasts of the size and health status of the population for health policy planning. Particularly the expected increases in chronic conditions of old age such as dementia, hip fractures, or heart failure are a matter of concern.²⁻⁴ The projected growth of the elderly population, those aged 75 and older, caused both by high birth rates in the past and increased survival now and in the future is the main determi-

nant of escalating needs for long term care.³ Since nearly all future residents of nursing homes and sufferers from disabling diseases of old age between now and 2060 have already been born, forecasts of the size of the health care needs of the elderly population rely heavily on estimates of future survival, and the health status of those surviving. Such long term forecasts of future morbidity and mortality to be based on an understanding of age related disorders. Why does ageing, or better senescing, happen?⁵

The hallmark of senescence is the progressive increase in age specific death rates following puberty, observed in all mammals and many other species kept under conditions ideal for survival.⁶⁻⁸ Underlying this progressive increase is a generalised deterioration in a broad spectrum of physiological and metabolic functions.^{9,10} These physiological decrements leave the organism increasingly vulnerable to a variety of intrinsic and extrinsic factors that may cause disease, disability and death. The evolutionary theory of ageing explains senescence as a process tailored by natural selection in evolutionary time.

In this paper we will first explain the fundamentals of evolutionary theory. Then we discuss consequences for future public health.

A history of life and death

Neo-Darwinian evolutionary theory states that the human genome is the result of the forces of natural selection through evolutionary time-scales.¹¹ The evolutionary success of a species is determined by the "fitness" of its individuals, the increase or decrease of numbers of descendants through successive generations.^{12,13}

"Genetic neglect"

Figure 1 shows the mortality history of Dutch men and women in the twentieth century. While the life expectancy increased tremendously, the age dependent mortality changes are remarkably similar in all periods.⁵ Child mortality is highest immediately after conception, and drops to low levels at the end of childhood and the beginning of puberty. Even in the darkest of ages and the harshest of conditions, all our forefathers and -mothers succeeded to reach adulthood and have at least one child. Not a single one in our long line of ancestors died in infancy or childhood. Genes commanding juvenile survival must be very good indeed.¹⁴

But after the onset of puberty the cumulative probability of both successful reproduction and (violent) death is increasing. In natural conditions the mortality hazard is high, and relatively constant over age. The probability that a person would be dead – from hunger, injury, infection or predation – increases sharply by age, long before old age is reached. The power of natural selection, weeding out detrimental mutations, is consequently decreasing. Genes protecting survival at older ages are not selected for, because ageing is

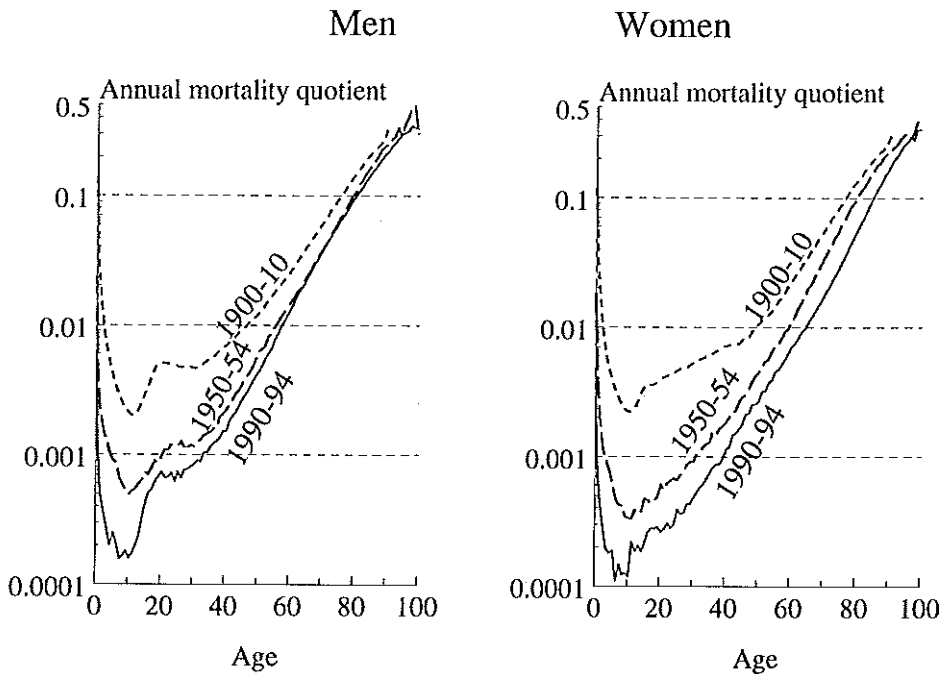


Figure 1 - Mortality rates by sex, age and 3 periods in the 20th century.

rare, and inconsequential to production and survival of offspring.⁶ Evolution does not care about elderly.

This is the central pillar of any evolutionary theory of ageing: with or without senescence, the length of life is limited.^{6,15,16} In natural conditions mortality from hunger, injury, infection or predation is so high, that few individuals survive into old age.

A body is a disposable good

There is an additional reason why senescence exists. Millions of years of ruthless competition have optimised the life history of a species.¹⁶⁻¹⁸ Food (energy) is nearly always in short supply. Given similar physiologies and environmental constraints, the organism that makes the most of scarce resources, will have the stronger lineage of successful descendants. There are choices to be made between investing resources in reproduction or survival.^{9,15,16}

Some species invest in much offspring early in life, but at the price of high mortality and short lifespans. Others adopt a caring strategy: they live long, but reproduce late and little. *Homo Sapiens* is the ultimate carer among mammals. His very long life span is related to his very long youth, which again is related to the large size of his brains.^{7,19,20} The larger the brain is, the longer education and parental care is needed to make use of its full potential.⁷ But after offspring is able to fend for itself, the parent is disposable from the point of view of natural selection.

The alternative name of the optimality theory of senescence, the "disposable soma" theory, refers to the disposable goods from economy.^{9,16} An optimal product takes into account the expected duration of use of that product. It is a waste of resources to invest in increased durability beyond that duration. For animals that is the natural life expectancy in the wild.²¹ Organisms who succeed in diverting this energy to more successful reproduction will outcompete the more wasteful. Fitness, i.e. successful offspring, not survivorship, determines evolutionary success.

Optimality theory is consistent with other, non-evolutionary, theories of ageing. The evolution theory gives an 'ultimate' explanation, why stochastic "wear and tear" and/or somatic damage such as oxidative stress by free radicals cause senescence.²¹⁻²³ Lasting damage is not inevitable: living beings are dynamic systems, capable of maintenance and repair. Damage happens because maintenance and repair become increasingly less effective at older ages, when the body is disposable.

Theories of adaptation

The previous explanations of senescence are called non-adaptive. Animals are not selected 'because' they senesce. Senescence is a by-product of lack of selection pressure and an optimal life history. Alternative theories state that senescence can be adaptive; they are based on an argument of crowding. "Groups" are selected for senescence, because the older individuals are discarded, and leave room for the healthy young. First, there is a circularity in this argument: the reason why the old are unhealthy and subfertile is because they senesce. If the circularity is excluded by stating that the old are as healthy and as fertile as the young, it is hard to see how these groups might be disadvantaged, compared to groups with unhealthy elderly.

Based on theories of adaptation the existence of "suicide genes" can be predicted, genes which actively programme the end of life. This is for the same reason unlikely: there would be strong selection pressure for mutations deleting these genes. At least in the short term, these immortal (healthy and fertile) 'escape mutants' would outcompete the ageing mortals in numbers of offspring. Such escape mutants have never been observed in nature.

Disability and death from evolutionary perspective

Optimality theory states that the age dependent occurrence of degenerative diseases is timed by the life history of our ancestors. Energy consuming maintenance systems at multiple levels, from subcellular DNA-repair to normal functioning of the brain, will assure more or less the same life span.³⁻¹⁰ This implies that diseases caused by senescence are timed by the same process. The distinction between fatal and disabling conditions only makes sense in human societies: in natural conditions, a disabled hunter has few chances for survival. If one of the systems is set too "low", and fails consistently too

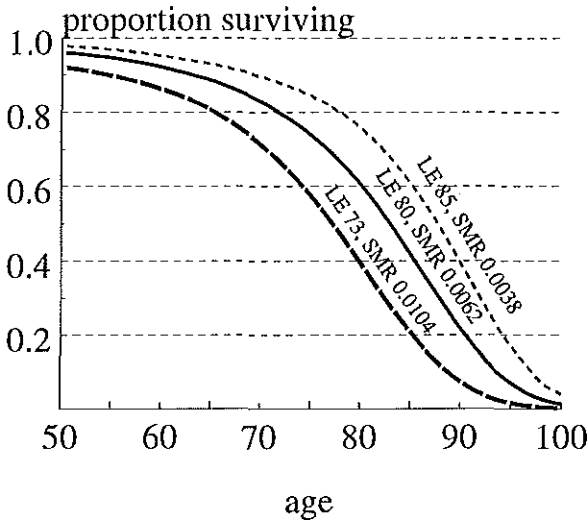


Figure 2 - Survival in period life tables of Dutch women in 1950-54 (life expectancy 73 y), in 1985-89 (life expectancy 80 y) and after projecting the age specific mortality decline between 1950-1989 till 2020-24 (life expectancy 85 y). SMA is the age standardized mortality rate.

soon, natural selection will tend to increase the level at which the system is set. If persons prone to early cancer have less offspring or less opportunity to rear them successfully, their genes will disappear. But conversely, if one of the systems is set too "high", and would fail always after the others, natural selection will tune it down. The saved resources are, in terms of natural selection, better used to boost fitness. The optimal balance between spending "enough" and spending "too much" in survival is not very sharp.²¹ Selection pushes survival to an optimum, but the pressure is weaker the more this optimum is approached. This allows for considerable variance of senescence rates. Not all people age at the same rate, and not all systems of an individual deteriorate at the same rate.

The future of death

As mentioned in the introduction, estimates of the future size of the population of elderly in the foreseeable future are nearly entirely determined by their survival. Figure 2 shows the stationary life table populations of women of two periods, 1950-54 (age standardised mortality 0.0104 y^{-1} , life expectancy 73 year), 1985-89 (age standardised mortality 0.0062 y^{-1} , life expectancy 80 year) and of a hypothetical period with a life expectancy of 85 year (age standardised mortality 0.0038 y^{-1}). This hypothetical life expectancy and mortality are reached when the mortality would decline at the same (age specific) rate as between 1950-54 and 1985-89, which would correspond to the period 2020-24. The stationary life table population assumes that all birth

and mortality rates remain constant at the level of a certain period. In 1950-54, when the life expectancy was 73 years, 6.6% of the life table population was over 75 year and 1.2% over 85 year old. In 1985-89, at a life expectancy of 80 y, 10.5% was over 75 and 3.1% over 85. In 2020-24, at a projected life expectancy of 85 y, an estimated 14.0 % would be over 75 and 5.3 % over 85. In the Netherlands in 1994, the population of 75 and older comprised 5.5% of the population; this fraction is relatively low as a consequence of the large post-World War baby boom cohort. However, these 5.5% consumed 28.8 % of all health care costs.²⁴ In 2020 the survivors of the baby boom cohorts will start reaching 75 y, their ultimate size being determined by the mortality decline in elderly populations, as mortality among the young and adult is already very low. To be valid, any prediction of further declines in mortality rates should therefore involve a theory about senescence.

Based on evolutionary theory, we can make predictions about senescent death. Death is not 'programmed' by death genes, but the result of a general neglect of the genetic material at older ages, both by lack of historical selection pressure and by imperfect maintenance. Life span is not fixed by 'hard' programmed limits, but by 'soft' probabilistic limits, a result of genetic variance and stochastic accumulation of damage.²⁵ Major breakthroughs in 'genetic engineering' of senescent mortality are very unlikely, as the number of genes involving maintenance at multiple levels is large and the potential for synergy immense.²⁶ Opportunities for slowing the rate of senescence are to be sought in decreasing the rate of accumulation of damage, which is identical to a familiar concept to public health practioners and epidemiologists: diminishing exposure to risk factors.

The question how much mortality reduction can be obtained by further risk factor reductions is not without answers. With the notable exception of smoking²⁷, simple risk factor models building on classical epidemiological techniques show but modest extensions of life expectancy by risk factor reduction, in the order of magnitude of one to a few years.^{28,29} More complex multivariate models, using longitudinal data of the Framingham Heart Study, arrived at similar results.³⁰ Predictions of 'natural' life expectancies of more than 95 years and life spans of over 130 y are based on few and hard to verify observations of extinct cohorts born before 1880 and many assumptions about the 'tail' of the distribution of deaths at ages above 110.³¹ The probabilistic nature of the end of life, predicted by evolutionary theory, foresees a long tail of scarce survivors reaching extreme old age, but the population life expectancy is little sensitive to these small numbers.

Another approach to the problem of an uncertain future is to study the recent past, and extrapolate the changes needed to reach a certain life expectancy.³² Figure 1 shows the evolution of mortality in the twentieth century by age (since 1900), and figure 3 by cause of death (since 1950) in the Netherlands.

During the first half of the century, mortality declined predominantly during child- and adulthood of both men and women, the changes in middle age being more modest. In the third quarter of the century, the difference

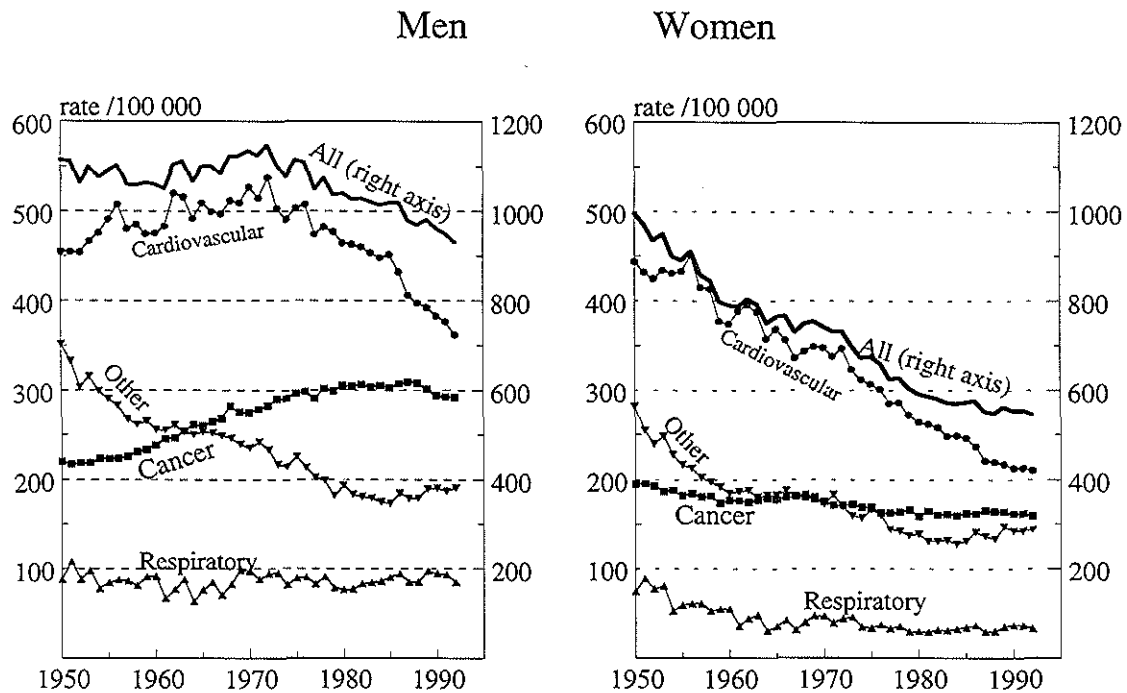


Figure 3
Mortality rates by sex, period
and causes of death since
1950.

between men and women is astonishing. Men experienced increasing cardiovascular disease and cancer mortality, caused by coronary heart disease and smoking induced respiratory cancers, while among women death rates decreased from all causes. The most recent period shows again striking changes between men and women, but now in opposite directions. Among men, cardiovascular mortality is going down rapidly, and cancer mortality seems to have reached its peak, and starts declining now. But among women, the mortality decline is levelling off since 1980; from 1950 till 1980, age standardized mortality decreased by 1.5% per year, from 1980 till 1992 by 0.6% per year. The decrease in cardiovascular mortality slowed down.³³ Cancer and respiratory disease mortality have been stagnating since long, and all other causes are increasing, as among men. To increase the actual life expectancy among women from 80 (80.3 year in 1994) till 85 year, all cardiovascular mortality (the only declining cause of death since 1980) has to be eradicated, or alternatively the present downward trend in mortality from all causes (an annual decline of 0.6%) has to persist for the next 85 years.

Obviously, the recent mortality experience of Dutch women does not necessarily mirror the experience of all women of the developed world. Indeed, one study, reconstructing mortality and census counts of four racial/ethnic populations in the USA, showed that female immigrants of Asian/Pacific origin have reached a life expectancy of even more than 85 years.³⁴ But first, migrant populations are a healthy selection of the population of origin, even more so because USA immigration laws only allowed the highly educated and healthy. Second, moribund immigrants may return to their country of origin to die, biasing death counts. Third, exaggeration of age is prevalent in most human societies; the majority of Asian migrants are foreign-born, and their exact birth-year is not known.^{35,36} Fourth, the actual period life table is the result of the low mortality hazard of modern time, applied to the selected survivors of past (higher) mortality. An earlier demographic transition³⁷ and a lower historical mortality might explain the stagnating increase of life expectancy, also observed in other highly developed countries such as the Scandinavian countries.³⁸ Higher (period) life expectancies are possible in societies moving more rapidly through the epidemiological transition, but only for a while. As the birth cohorts exposed to the high pre-transitional mortality rates die out, the high mortality selection early in life will disappear.

The recent mortality experience of elderly in 27 developed countries showed a steady and unprecedented decline between the 1960s and 1980s.³⁶ However, this mortality decline was observed in cohorts of survivors exposed to 19th centuries' high child mortality and mostly before reaching life expectancies of 80 and higher. The (relative) mortality decline, particularly among women, tapered off with age, as expected when the end of life is programmed by a 'soft' probabilistic limit.²⁵

The life expectancy of Dutch women of 85 in 2020-24 (figure 2) was reached by projecting this historically steep mortality decline between 1950-54 and 1985-89 among women, and ignored largely the levelling off in the most recent period. If we project the period 1985-1994, it will take nearly a

century before Dutch women reach 85 years. So for the foreseeable future (female) life expectancies above 85 seem unlikely in the Netherlands. In selected populations, and in populations moving rapidly through the demographic transition, higher period life expectancies are possible at the end of the transition, but only for a while. While among hunter gatherers grandparents are useful to their genetic offspring because of the prolonged period of juvenile dependence³⁹, the maintenance schedule of *Homo sapiens* does not foresee great-grandparents.

Compressing disability, expanding health care needs?

The mortality trends from the recent past among Dutch women are insufficient to reach a life expectancy of 85 year within a half century. However, even moderate increases in life expectancy will sharply increase numbers of elderly, as we showed previously. The next question is how disabled these future populations will be.

The question if and how senescent disability among these elderly will change depends basically on the equilibrium of three forces. The first force slows the rate of senescence, decreases incidence, and 'compresses' senescence until old age. The second force decreases the mortality hazard, increases disease duration and expands senescence. Disability and death at old age are timed by the same process, and successfully postponing death from one cause will increase disability from other causes. And the often forgotten third force 'compresses' morbidity, not by reducing the numbers but by reducing the severity of disease: by successfully preventing, treating or palliating disability. There is consequently a paradoxical relation between disability and health care need. Actual disability is the disability not prevented and not palliated. One of Fries' prime examples of compression of disease, stroke, has been bought in part by health care, namely treatment of hypertension, which costs 1.3 % of the total Dutch health care budget.^{24,40,41} Historically, it is reasonable to assume that the epidemiological transition causes both decreasing mortality and decreasing disability as most of the transition is caused by prevention of infectious diseases.^{42,43} However, in the fourth phase of the transition, the age of delayed degenerative disease⁴⁴, health care is one of the prime forces of change. The effects of health care on disability are mixed: health care increases disability by delaying (cardiovascular) death, decreases severity by palliating disease symptoms and prevents disability by treating risk factors such as diabetes mellitus, hypertension and cholesterol. So it is no surprise that disability free life expectancy is increasing and age adjusted disability trends in elderly are still decreasing in the fourth phase of the transition.⁴⁵ However, if life expectancy increases further, senescent diseases, timed by the same evolutionary forces as senescent death, are uncovered. Health care needs will increase to reduce the ensuing disability. If and how disability will compress, will depend mainly on the effectivity of that health care, but there is little doubt that health care needs will increase.

References

1. Bennet N, Olshansky S. Forecasting US age structure and the future of social security: the impact of adjustments to official mortality schedules. *Population and Development Review* 1996;22:703-27.
2. Jorm AF, Korten AE, Jacomb PA. Projected increases in the number of dementia cases for 29 developed countries: application of a new method for making projections. *Acta Psychiatr Scand* 1988;78:493-500.
3. Schneider EL, Guralnik JM. The aging of America. Impact on health care costs. *JAMA* 1990;263:2335-40.
4. Bonneux L, Barendregt JJ, Meeter K, Bonsel GJ, van der Maas PJ. Estimating clinical morbidity due to ischemic heart disease and congestive heart failure: the future rise of heart failure. *Am J Public Health* 1994;84:20-8.
5. Carnes BA, Olshansky SJ, Grahn D. Continuing the search for a law of mortality. *Population and Development review* 1996;22:231-64.
6. Medawar PB. *An unsolved problem in biology*. London: Lewis, 1952.
7. Kirkwood TBL. Comparative and evolutionary aspects of longevity. In: Finch CE, Schneider EL, editors. *Handbook of the biology of aging*. New York: Van Nostrand Reinhold, 1985:27-44.
8. Kirkwood TB. Comparative life spans of species: why do species have the life spans they do? *Am J Clin Nutr* 1992;55(6 Suppl):1191S-1195S.
9. Kirkwood TB, Rose MR. Evolution of senescence: late survival sacrificed for reproduction. *Philos Trans R Soc Lond B Biol Sci* 1991;332:15-24.
10. Jazwinski SM. Longevity, genes and aging. *Science* 1996;273:54-9.
11. Dawkins R. *The selfish gene* (new edition). New York: Oxford University Press, 1989.
12. Darwin C. *On the origin of species*. London: John Murray, 1859.
13. Dawkins R. *The extended phenotype*. New York: Oxford University Press, 1989.
14. Dawkins R. *River out of Eden*. London: Phoenix, Orion books ltd, 1995.
15. Williams GC. Pleiotropy, natural selection and the evolution of senescence. *Evolution* 1957;22:398-411.
16. Kirkwood TB. Evolution of ageing. *Nature* 1977;270:301-4.
17. Parker GA, Maynard-Smith J. Optimality theory in evolutionary biology. *Nature* 1990;348:27-33.
18. Partridge L, Barton NH. Optimality, mutation and the evolution of ageing. *Nature* 1993;362:305-11.
19. Sacher GA. Mammalian life histories: their evolution and molecular-genetic mechanisms. *Adv Pathobiol* 1980;7:21-42.
20. Harvey PH, Clutton-Brock TH. Life history variation in primates. *Evolution* 1985;39:559-81.
21. Kirkwood TB, Franceschi C. Is aging as complex as it would appear? New perspectives in aging research. *Ann N Y Acad Sci* 1992;663:412-17.
22. Weindruch R. Caloric restriction and aging. *Sci Am* 1996;274:46-52.
23. Sohal RS, Weindruch R. Oxidative stress, caloric restriction, and aging. *Science* 1996;273:59-63.

24. Polder J, Meerding JW, Koopmanschap MA, Bonneux L. The cost of diseases in the Netherlands, 1994. Rotterdam: Department of Public Health, 1997.
25. Manton KG, Tolley HD. Rectangularization of the survival curve. Implications of an ill posed question. *J Aging Health* 1991;3:172-93.
26. Kirkwood T. How can we live forever? *BMJ* 1996;313:1571.
27. Peto R, Lopez AD, Boreham J, Thun M, Heath C, Jr., Doll R. Mortality from smoking worldwide. *Br Med Bull* 1996;52:12-21.
28. Gunning-Schepers L. The health benefits of prevention: a simulation approach. *J Health Policy*. 1989;12/1-2 (special issue):93-129.
29. Tsevat J, Weinstein M, Williams L, Tosteston A, Goldman L. Expected gains in life expectancy from various coronary heart disease risk factor modifications. *Circulation* 1991;83:1194-1201.
30. Manton KG, Stallard E. Chronic disease modelling. London, England: Charles Griffins & co., 1988.
31. Manton KG, Stallard E. Longevity in the united states: age and sex-specific evidence on life span limits from mortality patterns 1960-1990. *J Gerontol A Biol Sci Med Sci* 1996;51:B362-75.
32. Olshansky S, Carnes B, Cassel C. In search of Methuselah: estimating the upper limits to human longevity. *Science* 1990;250:634-640.
33. Bonneux L, Looman CW, Barendregt JJ, Van der Maas PJ. Regression analysis of recent changes in cardiovascular mortality and morbidity in the Netherlands. *Br Med J* 1997;314:789-92.
34. Hahn RA, Eberhardt S. Life expectancy in four U.S. racial/ethnic populations: 1990. *Epidemiology* 1995;6:350-5.
35. Rosenwaike I. A note on new estimates of the mortality of the extreme aged. *Demography* 1981;18:257-66.
36. Kannisto V, Lauritsen J, Thatcher A, Vaupel J. Reductions in mortality at advanced ages: Several decades of evidence from 27 countries. *Population and Development Review* 1994;20:793-810.
37. Omran AR. The epidemiologic transition: a theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly* 1971;49:509-38.
38. Van Hoorn WD. Mortality: trends, backgrounds and forecast. *Monthly population statistics: Statistics Netherlands*, 1997:10-17.
39. Pecesqui JS. A hypothesis for the origin and evolution of menopause. *Maturitas* 1995;21:83-9.
40. Fries JF. Aging, natural death, and the compression of morbidity. *N Engl J Med* 1980;303:130-5.
41. Bonita R. Epidemiology of stroke. *Lancet* 1992;339:342-4.
42. Murray CJL, Lopez AD. Regional patterns of disability free life expectancy and disability adjusted life expectancy. The global burden of disease study. *Lancet* 1997;349:1347-52.
43. Barendregt JJ, Nusselder WJ, Bonneux L. Global Burden of Disease [letter]. *Lancet* 1997;350:142-143.
44. Olshansky SJ, Ault AB. The fourth stage of the epidemiologic transition: the age of delayed degenerative disease. *Milbank Memorial Fund Quarterly* 1986;49:355-91.

45. Manton KG, Corder L, Stallard E. Chronic disability trends in elderly United States populations: 1982-1994. *Proc Natl Acad Sci U S A* 1997;94:2593-8.

PART VII

Discussion and summary

General discussion and conclusions

Introduction

In the general introduction of this thesis we presented its main theme, the future evolution of morbidity. We argued that ‘generic’ morbidity was the resultant of diverse developments in various diseases; to gain insight, we had to start from these specific developments. The burden of morbidity of a disease is determined by the prevalence and severity of its stages, ‘stocks’ caused by ‘flows’ of incidence, remission or death. Capturing changes asks for dynamic models, modelling both stocks and flows, describing the main epidemiological characteristics of a disease. In part II we described a number of diseases and their developments over time, using such dynamic epidemiological models. In part III we proposed a methodology to describe a large number of diseases simultaneously. In part IV we developed an indicator of public health that allows to summarize the disease specific results in a generic indicator capable of measuring compression and expansion. In part V we had a brief look at the relation between health care costs, life extension and primary prevention. In part VI, finally, we reflected on the ultimate, biological causes of morbidity and mortality.

This leads to three questions. The main research question of this thesis is: “Will there be compression or expansion of morbidity?”. The two main characteristics of the methods used are a disease specific, epidemiological approach, and a reliance on modelling to describe the complex interactions between morbidity and mortality and to integrate data from various sources. We therefore formulated the two additional questions: “What can a disease specific approach contribute to the understanding of compression and expansion at the level of overall population health status?”, and: “How can epidemiology and modelling techniques be integrated into public

health models fit to answer questions of compression and expansion on the disease specific level, and on the level of overall population health?" This discussion starts at the last question, and will end at the first.

Models of disease

The model paradox

The main research question is one about future developments, or prediction. According to Popper prediction and explanation are closely related because they consist of the same elements: initial conditions, a theoretical framework, and an outcome. It depends on which two out of three we take, provisionally, for granted to determine what is done. When we look for initial conditions, given an outcome and a theoretical framework, an explanation is given. Using initial conditions and theoretical framework allows us to make predictions. And of course, given initial conditions and outcome we can put the theoretical framework to the test (Popper, 1961).

Theoretical frameworks are necessary for explanation and prediction. The theoretical frameworks we use are in large part embodied in the models described: as such models are no different from theoretical frameworks, they are a detailed specification of a causal theory (McClelland, 1975). The necessary explicitness of models is an asset, that makes them, in principle, better transferable and testable. But the epistemological status of models is no different from any theoretical framework, and therefore we can concentrate on the usefulness of models in public health research.

The main advantage of models is that they allow the description of very complex relationships. Complexity ensues, among others, from relations of many to many variables, from non-linear relations between variables, and from dynamic variables that display time-dependent behavior.

Rather paradoxically the capacity to handle complexity often turns into the largest disadvantage of modelling. Many models, including some of our own earlier work, are too complex, defy explanation, and thus do not comply with basic methodological requirements of reproducibility and intelligibility. But despite the habit of over-complexity models tend at the same time to be too simple to answer research questions of interest. We will illustrate this dilemma with the problem of modelling multiple diseases and risk factors simultaneously, and derive from its discussion the general rule how to avoid the dilemma. This rule may not be exactly new, but the discussion may prove enlightening nevertheless.

Population heterogeneity

A basic methodological problem of a multiple disease and risk factor model is how to deal with population heterogeneity. Independence assumptions, as presented in chapter 12, facilitate a multi-disease model but rule out many interesting research questions where heterogeneity or disease dependencies are involved. People run different risks on disease incidence, and

consequently on prevalence and mortality. As shown in chapters 13 and 14 this results in mortality selection: the higher mortality risk lowers the prevalence of people at this higher risk.

Mortality selection defeats straightforward ways to describe risk factor exposure in the population with a continuous distribution, as would be preferable for most risk factors. Because persons at the extreme of the distribution are at the highest risk, they are selectively removed from the population, with the consequence that the continuous distribution is not a good description of the exposure any more. We are aware of just one multivariate risk factor model with continuous distributions that is able to deal with mortality selection, but it does so at the price of a host of restrictive assumptions (Manton & Stallard, 1988).

A possible solution to the problem, but one not pursued in this thesis, is to use a micro-simulation technique. When a population is described at the level of the individual, any number and type of relations between risk factors, diseases, and causes of death can, in principle, be described. However, while micro-simulation can solve the heterogeneity problem, there are also drawbacks to the technique.

The solution chosen in chapter 13 is to use a categorical risk factor distribution: diabetics and non-diabetics. Dividing the population into (more) homogeneous subpopulations allows to give each its own disease and mortality risks. But problems remain. It will be clear that with multiple risk factors (and in particular with several exposure categories per risk factor) the number of subdivisions grows exponentially. This becomes even more true when, like in the *Prevent* model, time lags between changes in risk factor exposure and health effects are taken into account (Gunning-Schepers, 1989; Gunning-Schepers, Barendregt, & van der Maas, 1989)

Yet another solution, and one that has been used in chapters 15 and 18 (as in the *Prevent* model), is to ignore population heterogeneity (and by implication mortality selection) altogether, and describe the effects of changes in risk factor exposure for a homogeneous population. This greatly simplifies the model, but whether it produces valid results depends on the research question.

As was explained in part III in a homogeneous population only substitution of diseases and causes of death exists, while in a heterogeneous population in addition to substitution also competition exists. When the research question explicitly addresses a consequence of changing competition, as chapter 13 does for the diabetes prevalence after heart disease mortality reduction, then of course the model should allow for heterogeneity. When, on the other hand, the consequences of eliminating a risk factor are assessed, as is done in chapters 20 and 24 with smoking, competition plays no role: it exists only when the excess mortality risk of one smoking related disease changes (in this case, disappears), while the excess risk for others does not. So, unless one assumes that current smokers would have excess risks for some diseases even when they would never have smoked, competition will not affect outcome.

A general rule for modelling

This leads to the general rule to avoid the dilemma of too complex models that are still too simple: for every research question the model used should be the simplest one capable of answering it. A custom-made model made to fit the research question, instead of a general, all-purpose, public health model (the original aim of the TAM-project, see general introduction). Not exactly a new rule indeed, because it goes straight back to Ockhams razor, and even further down to Aristoteles (*Encyclopaedia Britannica*, 1995).

Thus, where possible, static life table models should be used instead of dynamic ones, one-stage disease models instead of multi-stage ones, and a homogeneous risk factor model instead of a heterogeneous one. Given this parsimonious approach, models are very useful, because the causal structure and quantified nature allow to explain observations and make predictions (see the heart disease chapters), and allow to derive which way the balance of two opposing tendencies will go (as for example in the prevention and costs chapters, 23 and 24), and can help to estimate an incidence in cases where data are very weak or unavailable (chapters 4, 5 and 13).

A demographic epidemiology

For some time now epidemiologists are debating the return to public health of epidemiology as a science. Many epidemiologists feel that their craft has drifted away too far from its original goal of improving public health, and is increasingly occupied with measuring as precisely as possible attributes that bear little or no relevance for population health status (Pearce, 1996). While few epidemiologists will dispute that the application of their results should be the ultimate goal, it is less clear how this "return to public health" should be achieved. Some proposals are clearly not appreciated by everybody (Anonymous, 1997).

Most of the research reported in this thesis is based on the application of epidemiology in the field of public health, and from the outset it was clear that we might need to get our hands dirty (see chapter 2). We make assumptions, combine results from different studies and from different populations, and use data from non-epidemiological, administrative registries. In all cases a model serves as an organizing principle: specifying causal relations between the variables enormously reduces the set of possible values that the variables can take on.

The basic ingredient of all these models is age, the main determinant of health. Each prevalent case at some age necessarily must have become incident at some earlier age. When people start surviving previously fatal infarctions at some age they will necessarily be prevalent at later ages, however shortly. To model age and the "age dynamics" explicitly we employ methods from demography, in particular the (multi-state) life table and extensions that include time. A concise description of our modelling approach could be "embedding epidemiology in a demographic framework".

The demographic framework provides not just a model for age, but also the translation to outcomes on the level of overall population health status. By design life tables take care of substitution of causes of death: this averts overly optimistic estimates of the health gain of interventions that affect mortality. And the life table produces indicators of overall population health status: life expectancy and disability adjusted life expectancy (DALE), which allow to measure compression and expansion.

The marriage of epidemiology and, largely demographic, modelling techniques forms a powerful tool to translate epidemiological findings to outcomes on the level of population health.

Disease specific approach

Costs and benefits

The choice to describe population health status and the changes therein on the level of specific diseases and causes of death has two immediate consequences:

- There is a large number of diseases and causes of death, so a description of public health on the disease specific level will somehow have to deal with this large number. For example, Murray and colleagues described for the Global Burden of Disease project the epidemiology of 107 diseases and conditions (Murray & Lopez, 1996b).
- The disease specific descriptions have to be integrated to one description of population health status. This thesis presented to that end a multi-disease model and a mapping from disease specific to generic morbidity.

The disease specific approach therefore requires a substantial effort, as compared to one which directly measures morbidity on some generic level, like disability. Moreover, there will always some morbidity left unexplained: even one-hundred diseases and conditions will leave some morbidity not accounted for, and at least some morbidity is difficult or impossible to explain by specific diseases, particularly among the elderly.

Some research questions do not require a disease specific approach. For example, those who just want to measure morbidity and its changes over time, are better off measuring generic morbidity directly, and avoid the disease specific overhead.

Many policy oriented questions, on the other hand, stress the need for a disease specific approach. To determine the effect of policy interventions on population health status requires a link between exposure to risk factors, the occurrence of and survival with specific diseases, and some generic outcome measure like Disability Adjusted Life Expectancy (DALE).

This is a strong argument for a disease specific approach, but there is also at least one strong objection: lack of data. Data on risk factor and disease epidemiology are as a rule scarce, not consistent between studies, and time series are in most cases simply not available (Baan, Bonneux,

Ruwaard, & Feskens, 1997; Bonneux, van de Mheen, Gunning-Schepers, & van der Maas, 1996; van de Mheen, Bonneux, & Gunning-Schepers, 1995). Much of the effort of the disease specific chapters of part II was therefore devoted to the assimilation of available data to consistent sets of incidence, prevalence, and mortality.

In summary, there are strong arguments for a disease specific approach, but that also carries a considerable cost. Whether these costs are justified, depends on the benefits.

The benefits are largely determined by the disease specific differences. We have already pointed out that the original proponents of both the expansion and compression hypotheses had different diseases in mind: Kramer and Gruenberg primarily congenital mental disorders, and Fries primarily cardiovascular diseases (Fries, 1980; Gruenberg, 1977; Kramer, 1980). And indeed we also find differences between diseases. A broad distinction can be made between 'fatal' and 'disabling' diseases.

Fatal diseases

Much of our effort has been directed towards the main causes of death: cardiovascular disease and cancer. The two groups of cardiovascular diseases modelled, coronary heart disease including heart failure, and stroke, showed complex interactions and paradoxical results (see chapters 4-7). In coronary heart disease, expansion of morbidity is distinct, driven by decreasing mortality (see chapters 4,5 and 7). Expansion is likely both in numbers of disabled and in severity of disability. Acute coronary heart deaths are exchanged for severe morbidity, caused by chronic congestive heart failure. Because 'outflow' (deaths) decreased in all disease stages, the milder stages acted as risk factor for the more severe stages, causing a cumulation of cases in the severe heart failure stage (chapter 4).

In stroke, compression is likely to have occurred until 1987 (chapter 6 and 7). The main difference between stroke and coronary heart disease is the lack of major progress in survival after a first event. Therefore, the trade off of disease for death was less pronounced. Maybe another difference with coronary heart disease was a more pronounced incidence decline of stroke, caused probably by treatment of hypertension (Bonita, 1992; Bonita & Beaglehole, 1986; Higgins & Thom, 1993; Klag & Whelton, 1993; McGovern, et al., 1992).

Recently, compression of stroke seems to have come to a stop. Since 1987 (and the massive introduction of thrombolysis in the years before), the prevalence of vascular disease is expanding rapidly, increasing the risk for cerebro-vascular disease (chapter 7, (Higgins & Thom, 1993; Howard, 1993)). But for both cardiovascular diseases some delay of disease until higher ages could be observed: prevalence decreased at younger ages, but increased among the oldest old.

Chronic vascular disease is an important risk factor for other senescent diseases. Chronic vascular disease is a direct risk for the vascular dementia (Kase, 1991), but also for Alzheimers' disease (Breteler, Claus,

Grobbee, & Hofman, 1994; Breteler, et al., 1994; Edelman & Wei, 1996; Hofman, et al., 1997; Sliemers, et al., 1997). A certain pathway is declining mortality selection of unfavorable ApoE subtypes, increasing the risks for both atherosclerosis and Alzheimers disease, (obviously not excluding other possible pathways such as endothelial damage by β -amyloid peptides) (Thomas, Thomas, McLendon, Sutton, & Mullan, 1996). A similar effect of declining mortality selection will be causing an increase of diabetes mellitus. Diabetes is a risk factor for heart disease, which implies that diabetes prevalence is being kept down by excess mortality from heart disease. As this excess mortality declines, the prevalence of diabetes will consequently go up (chapter 14). The risk of hip fractures increases also, to some extent because of increased tendency of falling in cardiovascular impaired patients (Grisso, et al., 1991). Consequently, this expansion of chronic vascular diseases will weigh heavily in the morbidity balance.

In cancer, the overall picture is different. First, smoking related cancers follow the smoking epidemic with an exceedingly long lag time (Gunning-Schepers, 1989; Gunning-Schepers, Barendregt, & van der Maas, 1989). Only in the eighties, society has started to reap the benefits of the declining proportion of male smokers since the sixties. Because most smoking related cancers are highly fatal, the ensuing compression of morbidity is modest (see chapter 19). In the (solid) cancers with better prognosis, the history is dominated by increasing case detection, both increasing incidence and decreasing mortality (chapter 8) (Black & Welch, 1993; Doll & Peto, 1981; Feinstein, Sosin, & Wells, 1985). This means that morbidity, caused by diagnosis and treatment is increasing, but the severe terminal disease is decreasing (de Koning, van Oortmarssen, van Ineveld, & van der Maas, 1990). This is another example of exchange. Early detection in cancer exchanges death and terminal disease for increasing numbers of cancer cases. Incidence increases, prevalence increases, but the severe disability of terminal disease is diminishing.

Disabling diseases

Two important 'disabling' diseases were considered: hip fractures and dementia. The traditional epitheton 'non-fatal' is less appropriate, as both are associated with important excess mortality.

There is more to hip fractures than osteoporosis (see chapter 9). A hip fracture is a senescent disease, a common endpoint of multiple diseases, as dementia is and heart failure. Risk for hip fractures is dominated by age and co-morbidity, more than by brittle bones. As such, hip fractures are bound to resist prevention, particularly of interventions targeting a single factor of senescence, such as postmenopausal bone loss among women. Indeed, we hypothesize that at least part of the hip fracture incidence increase in developed countries is caused by augmenting numbers of frail survivors with multiple diseases.

Dementia is the most health care demanding chronic neurodegenerative disease in the Netherlands (Koopmanschap, et al., 1994; Polder,

Meerding, Koopmanschap, Bonneux, & Van der Maas, 1997). We estimated the stage related excess mortality in the Rotterdam study, which turned out to be considerable in mild stages as well (see chapter 10). This has important consequences for strategies aimed to slow progression of disease. If these progression is slowed, stage related death is postponed too, causing disability from dementia to decrease less than expected. There is little hope that the burden of morbidity will decrease substantially, as long as the prospects for prevention remain extremely limited.

Other main causes of disability have not been covered in this thesis. Restricted in time and resources, we had to limit ourselves to only a few examples. In the Global Burden of Disease project the major burden of disability is determined by mental disability, which is caused by congenital handicap, neurodegenerative disease and chronic psychiatric conditions (Murray & Lopez, 1996a; Murray & Lopez, 1996b). Estimates of causes of health care costs support this finding for the Dutch circumstances (Koopmanschap, et al., 1994; Polder, Meerding, Koopmanschap, Bonneux, & Van der Maas, 1997). The second major cause of disability are the musculo-skeletal conditions, which are predominantly all types of joint disease, including dorsopathy (Koopmanschap, et al., 1994; Murray & Lopez, 1996a; Murray & Lopez, 1996b; Polder, Meerding, Koopmanschap, Bonneux, & Van der Maas, 1997).

Compression or expansion of morbidity?

From disease specific to overall morbidity and mortality

Because of the large holes in our coverage of total disability we can provide only partial answers to the question of future expansion or compression. But the partial answers serve very well to illustrate that the issue is more complex than the simple question suggests.

Future disability in low mortality countries is governed by senescent diseases and mental disability. Senescent diseases are characterized by the exponential increase of the occurrence of disease by age (Carnes, Olshansky, & Grahn, 1996; Medawar, 1952), mental disability by various causes of mental disorders. The main mental disorders are caused by neurodegenerative diseases (and are fundamentally senescent diseases), by congenital mental handicap or by chronic psychiatric conditions. This thesis does not consider congenital mental handicap and chronic psychiatric conditions, and is focused on senescent diseases.

The future of senescence is determined by changes in incidence, disease duration (mortality) and disease severity. Three types of epidemiological change may cause compression:

- The first, and by far the most easy to accomplish, is increasing mortality. As this is contrary to the primary aim of public health, we do not pursue the possibility.

- The second is to shorten disease duration by delaying incidence without delaying death. Fries took the example of stroke (see chapter 6; (Fries, 1980)). Indeed, as long as the force of mortality among patients does not decline, mortality from other causes will cut short disease duration at older ages. But compression of stroke stopped since 1987 (see chapter 7), and not only in the Netherlands (Howard, 1993; McGovern, et al., 1992); other examples of delayed incidence are scarce.
- The third way to achieve compression is by decreasing disease severity. Changes in disease severity have largely been ignored in the compression or expansion debate, but most health care is devoted to palliating symptoms of disease: fighting pain, alleviating discomfort, rehabilitating the disabled.

Three types of epidemiological change may cause expansion: increasing incidence, lengthening disease duration or intensifying disease severity.

- Increased incidence may be caused by increased risk or stepped up case detection. Cancer screening is an example of a health care intervention which causes disease prevalence to rise because of increased case finding and increased survival. The disability balance may nevertheless be positive as severe terminal disease may be foregone.
- The second cause of expansion is delaying death from degenerative disease by successful interventions that prevent patients to die. The saved patients will progress in the disease process, perhaps more slowly (see chapter 4).
- The same process may cause patients to accumulate in the more severe disease stages, thus intensifying average disease severity, the third cause of expansion. As health care is more and more successful to avert death from acute events, the patients will ultimately fade away in terminal chronic disease. The most fundamental reason of senescent disability is evolutionary. Natural selection has not prepared homo sapiens for old age, as 'premature' death was natural (Kirkwood & Holliday, 1979; Medawar, 1952; Parker & Maynard-Smith, 1990). As we delay mortality more and more, we will meet more and more bugs in our genetic software, which either have never been selected for eradication, or which confer advantage in juvenile survival and reproduction (Kirkwood & Holliday, 1979; Parker & Maynard-Smith, 1990).

In general there are more indications for future expansion than for compression. Cardiovascular disease, as a large cause of morbidity and mortality, shows an expansion by averting deaths from acute events, without stopping disease progression. The increase in life expectancy that follows gives other senescent diseases the opportunity to emerge. With cancer the drive for screening tends to cause expansion: while in some cases, like in breast cancer, the disability balance is most likely positive, in others, for example prostate cancer, it is almost surely negative (Adami, Baron, & Rothman, 1994).

Three major openings for the opposite trend of compression exist. The most striking one is still interventions to further reduce smoking prevalences. Our calculations show that, despite the higher life expectancy of non-smokers, the abolition of smoking would compress disability (chapter 19). These results, based on life table calculations with relative risks, are in agreement with more empirical studies (Rogers, Nam, & Hummer, 1994). In some countries, notably the USA, smoking prevalences have been reduced much further than in the Netherlands, where the decline has stalled, and may have reversed among the young (Ruwaard & Kramers, 1997). There is a pressing cause for an active policy to reduce smoking, and in particular to prevent the young from starting.

The second opportunity for compression of morbidity is by reducing the disability consequences of chronic disease. Much of the effort in health care is devoted to this cause, and progress has been made, for example in the rehabilitation of stroke patients and better medication for heart failure (SOLVD investigators, 1991). Further advances in this field can be expected.

The third opportunity to achieve compression is more research attention for large disabling, in particular mental and musculo-skeletal conditions. The relative neglect of these conditions has resulted in a large knowledge deficit, both epidemiological and etiological (van der Maas, et al., 1996). A research effort to reduce this deficit is called for.

The balance of expansion and compression

Changes in the overall population health status thus result from the balance of opposing tendencies towards compression and expansion in the various causes of morbidity and mortality. Moreover, currently observed changes may have to be traced back to developments many years ago. For example, the decline in male smoking prevalences between about 1970 and 1985 will have an impact on morbidity for some time to come, given the long time lags involved (see chapter 24).

Another complicating factor, and one we have not looked at, are cohort-related changes in socio-economic composition. Manton et al report declines in age-standardized prevalence of disability and a number of diseases, based on a longitudinal study (Manton, Corder, & Stallard, 1997). They hypothesize that part of the change may be explained by the higher educational and income status of the successive cohorts. If a higher socio-economic status is indeed a direct cause of better health, the same hypothesis may be held for the Netherlands (Stronks, 1997).

Murray and Lopez report from the cross-sectional Global Burden of Disease study that people in countries which have progressed further in the epidemiological transition spend a smaller part of their life expectancy in disability than those still in the earlier stages (Murray & Lopez, 1997). They argue that this supports the compression of morbidity hypothesis. While we think the finding very plausible, we disagree with the conclusion (Barendregt, Nusselder, & Bonneux, 1997). The debate on compression

and expansion, perhaps rather parochially, so far concentrated on the low mortality countries in the last phase of the epidemiological transition. An expansion after having reached this last phase is entirely consistent with a compression during the transition through the earlier phases

To the question "Will morbidity expand or compress?" is but one valid answer: "It depends". With increases in life expectancy going on there seems a general tendency of expansion, but there exist opportunities for active policies to achieve compression. When these policies are pursued, on balance a compression of disability may well prevail. But of course such policies would have consequences for health care needs and costs.

Disability, health care needs, and costs

At first glance it might seem that expansion will coincide with greater health care needs and costs, while compression would generate the opposite. In particular Fries has repeatedly stated that delayed incidence will result in the double benefit of compressed morbidity and cost savings (Fries, et al., 1993; Fries, 1980). However, the relation between morbidity and disability on the one hand, and health care needs and costs on the other is not as simple as that.

A good counter example is provided by the results from the global Burden of Disease study. Comparing countries in the early stages of the epidemiological transition (developing countries) with those in later stages (established market economies) reveals that during the transition a compression of morbidity occurred, but also an enormous increase in health care expenditures (Murray & Lopez, 1996b). We think this is no coincidence: the ability of health care to compress disability expands the need for health care. Compression of disability may well go hand in hand with expansion of health care needs and costs.

Cancer screening is an example of a health care intervention which causes disease prevalence to rise. The disability balance may be positive as severe terminal disease may be foregone, but health care needs expand because of increased needs for (further) diagnosis and treatment. Even one of the classic examples of Fries of compression of disease, declining stroke occurrence, has been bought at least in part by health care: treatment of hypertension (Bonita & Beaglehole, 1986; Higgins & Thom, 1993; Klag & Whelton, 1993; McGovern, et al., 1992). For every stroke prevented, many had to take life long anti-hypertensive treatment.

Of the three policy opportunities for compression only successful interventions on mental and musculo-skeletal disorders show the potential of combining compression of morbidity with lower health care costs, provided that these interventions would not be too expensive (chapter 23). The second policy opening, greater efforts to reduce the disability consequences of chronic disease, will almost surely require additional health care money.

Even a free, successful intervention to abolish smoking will, while compressing morbidity (chapter 19), increase health care costs (chapter 24).

This result is not due to some specific characteristic of smoking and smoking-related diseases, but has a more general explanation. Both disability and costs rise steeply with age, driven by the age-related increase of senescent diseases. But since some diseases can be treated more or less successfully they do generate costs but no (or very little) disability, while diseases that generate disability and cannot be treated still generate costs for care. Therefore the age gradient of the cost curve is more steep than that of the disability curve. Interventions that extend life expectancy will consequently tend to add more costs than disability. In the case of non-smoking the net gain or loss of the intervention depends on the balance of the additions in the added life-years, and savings in earlier life. As the cost curve is steeper it is entirely possible that the costs on balance increase, while disability on balance decreases.

On the whole, and with the potential exception of the mental and musculo-skeletal disorders, there seems to be little scope for public health policies that both generate desirable changes in population health status and save health care costs. The allocation and containment of the health care budget is bound to remain a painful and exasperating exercise. This may be disappointing for some health researchers and policy makers, but, as economists are fond of saying, there is no such thing as a free lunch.

Priorities for further research

Given the size of the health care budget and the prominence people give to their health, it is surprising how little is known about the occurrence of the large causes of morbidity in the Netherlands. There exist efforts in various diseases, but a systematic overview of incidence and prevalence of the main causes of morbidity is wanting. Consequently policy makers have to decide on important policies in a haze of ignorance about repercussions for population health status and the health care budget. A good example is the current issue whether to offer expensive but effective cholesterol lowering therapy, and if so, to whom.

It would be a major advance for public health research and policy making in the Netherlands when in a systematic effort the epidemiology of major diseases (or disease categories) would be described by internally consistent estimates of age-specific incidence, prevalence and mortality. The Global Burden of Disease study may serve as an example (Murray & Lopez, 1996b). A Burden of Disease project for the Netherlands has started tentatively by repeating the assessment of disability weights for a large number of diseases. What is still lacking are consistent estimates of disease incidence and prevalence. Such estimates should now have the highest priority. To allow the assessment of temporal trends such an effort should preferably be made on a repeated basis.

Three other research themes follow directly from the work reported on in this thesis. The first is demographic epidemiology. Several examples in this thesis have shown that embedding epidemiology in a demographic

framework offers possibilities to apply epidemiological results to research questions in public health. A systematic investigation of possibilities and requirements of this interdisciplinary method is desirable.

The second theme is the relation between (changes in) population health status and health care costs. Combining the tool of demographic epidemiology with costs of illness data allows to explore this relation, as two examples in this thesis have shown (see part V). Further work in this direction should help to identify where health care money is spend less efficiently, or where gains in population health status could most economically be achieved.

The third theme is to further explore the epidemiological and population health status consequences of the evolutionary basis of senescent disease. Most of the work in the field of "biodemography" has so far been done by biologists and demographers. We are convinced that epidemiologists and public health researchers can benefit from the work they have done, but also can contribute to it.

Public health is a research field that, by its very nature, requires the input from various disciplines. Creating a truly interdisciplinary approach to the questions in the field is the challenge researchers face.

References

- Adami, H., Baron, J., & Rothman, K. (1994). Ethics of a prostate cancer screening trial. *Lancet*, 343, 958-60.
- Anonymous (1997). Putting public health back into epidemiology. *Lancet*, 350(9073).
- Baan, C., Bonneux, L., Ruwaard, D., & Feskens, E. (1997). Prevalences of diabetes in the Netherlands in the Netherlands. A quantitative review. *EJPH*.
- Barendregt, J. J., Nusselder, W. J., & Bonneux, L. (1997). Global Burden of Disease [letter]. *Lancet*, 350, 142-143.
- Black, W. C., & Welch, H. G. (1993). Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *N Engl J Med*, 328(17), 1237-43.
- Bonita, R. (1992). Epidemiology of stroke [see comments]. *Lancet*, 339(8789), 342-4.
- Bonita, R., & Beaglehole, R. (1986). Does treatment of hypertension explain the decline in mortality from stroke? *Br Med J*, 292, 191-2.
- Bonneux, L., van de Mheen, P., Gunning-Schepers, L. J., & van der Maas, P. J. (1996). Daalde hypertensie in Nederland tussen 1974 en 1993? *Ned Tijdschr Geneeskde*, 140(52), 2603-2606.
- Breteler, M. M., Claus, J. J., Grobbee, D. E., & Hofman, A. (1994). Cardiovascular disease and distribution of cognitive function in elderly people: the Rotterdam Study [see comments]. *BMJ*, 308(6944), 1604-8.
- Breteler, M. M., van Swieten, J. C., Bots, M. L., Grobbee, D. E., Claus, J. J., van den Hout, J. H., van Harskamp, F., Tanghe, H. L., de Jong, P. T., van Gijn, J.,

- & et al. (1994). Cerebral white matter lesions, vascular risk factors, and cognitive function in a population-based study: the Rotterdam Study. *Neurology*, 44(7), 1246-52.
- Carnes, B. A., Olshansky, S. J., & Grahn, D. (1996). Continuing the search for a law of mortality. *Population and Development review*, 22, 231-64.
- de Koning, H. J., van Oortmarssen, G. J., van Ineveld, B. J., & van der Maas, P. J. (1990). Breast cancer screening: its impact on clinical medicine. *Br J Cancer*, 61, 292-7.
- Doll, R., & Peto, R. (1981). The causes of cancer. Sources of bias in estimating trends in cancer mortality, incidence and curability (appendix C). *J Natl Cancer Inst*, 66(6), 1270-81.
- Edelberg, H. K., & Wei, J. Y. (1996). The biology of Alzheimer's disease. *Mech Ageing Dev*, 91(2), 95-114.
- Encyclopaedia Britannica (1995).
- Feinstein, A. R., Sosin, D. M., & Wells, C. K. (1985). The Will Rogers phenomenon: Stage migration and new diagnostic techniques as a source of misleading statistics for survival in cancer. *N Engl J Med*, 312, 1604-8.
- Fries, J., Koop, C., Beadle, C., Cooper, P., England, M., Greaves, R., Sokolov, J., Wright, D., & Consortium, t. H. P. (1993). Reducing health care costs by reducing the need and demand for medical services. *N Engl J Med*, 329, 321-5.
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. *N Engl J Med*, 303(3), 130-5.
- Grisso, J. A., Kelsey, J. L., Strom, B. L., Chiu, G. Y., Maislin, G., O'Brien, L. A., Hoffman, S., & Kaplan, F. (1991). Risk factors for falls as a cause of hip fracture in women. The Northeast Hip Fracture Study Group. *N Engl J Med*, 324(19), 1326-31.
- Gruenberg, E. M. (1977). The failure of success. *Milbank Q/Health Society*, 55, 3-24.
- Gunning-Schepers, L. (1989). The health benefits of prevention: a simulation approach. *J Health Policy*, 12, 1-256
- Gunning-Schepers, L., Barendregt, J., & van der Maas, P. (1989). Population interventions reassessed. *Lancet*, 1, 479-81.
- Higgins, M., & Thom, T. (1993). Trends in stroke risk factors in the United States. *Ann Epidemiol*, 3(5), 550-4.
- Hofman, A., Ott, A., Breteler, M. M., Bots, M. L., Slooter, A. J., van Harskamp, F., van Duijn, C. N., Van Broeckhoven, C., & Grobbee, D. E. (1997). Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. *Lancet*, 349(9046), 151-4.
- Howard, G. (1993). Decline in stroke mortality in North Carolina. Description, predictions, and a possible underlying cause. *Ann Epidemiol*, 3(5), 488-92.
- Kase, C. S. (1991). Epidemiology of multi-infarct dementia. [Review]. *Alzheimer Dis Assoc Disord*, 5(2), 71-6.
- Kirkwood, T. B., & Holliday, R. (1979). The evolution of ageing and longevity. *Proc R Soc Lond B Biol Sci*, 205(1161), 531-46.
- Klag, M. J., & Whelton, P. K. (1993). The decline in stroke mortality. An epidemiologic perspective. *Ann Epidemiol*, 3(5), 571-5.

- Koopmanschap, M. A., van Roijen, L., Bonneux, L., Bonsel, G. J., Rutten, F. F. H., & van der Maas, P. J. (1994). Costs of diseases in an international perspective. *Eur J Public Health*, 4, 258-64.
- Kramer, M. (1980). The rising pandemic of mental disorders and associated chronic conditions and disabilities. *Acta Psychiatr Scand*, 62 (suppl 285), 282-97.
- Manton, K. G., Corder, L., & Stallard, E. (1997). Chronic disability trends in elderly United States populations: 1982-1994. *Proc Natl Acad Sci U S A*, 94(6), 2593-8.
- Manton, K. G., & Stallard, E. (1988). *Chronic disease modelling*. London, England: Charles Griffins & co.
- McClelland, P. D. (1975). *Causal explanation and model building in history, economics and the New Economic History*. New York.
- McGovern, P. G., Burke, G. L., Sprafka, J. M., Xue, S., Folsom, A. R., & Blackburn, H. (1992). Trends in mortality, morbidity, and risk factor levels for stroke from 1960 through 1990. The Minnesota Heart Survey. *Jama*, 268(6), 753-9.
- Medawar, P. B. (1952). *An unsolved problem in biology*. London: Lewis.
- Murray, C. J. L., & Lopez, A. D. (1996a). Global and regional descriptive epidemiology of disability: incidence, prevalence, health expectancies and years lived with disability. In C. J. L. Murray & A. D. Lopez (Ed.), *The global burden of disease* (Vol. Chapter 4, pp. 201-246). Harvard: Harvard University Press.
- Murray, C. J. L., & Lopez, A. D. (Ed.). (1996b). *The Global Burden of disease*. Harvard: Harvard University Press.
- Murray, C. J. L., & Lopez, A. D. (1997). Regional patterns of disability free life expectancy and disability adjusted life expectancy. The global burden of disease study. *Lancet*, 349, 1347-52.
- Parker, G. A., & Maynard-Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348, 27-33.
- Pearce, N. (1996). Traditional epidemiology, modern epidemiology, and public health [see comments]. *Am J Public Health*, 86(5), 678-83.
- Polder, J. J., Meerding, J. W., Koopmanschap, M. A., Bonneux, L., & Van der Maas, P. J. (1997). *Costs of Diseases in 1994 (in Dutch)*. Department of Public Health.
- Popper, K. R. (1961). *The poverty of historicism*. London: Routledge & Kegan.
- Rogers, R. G., Nam, C. B., & Hummer, R. A. (1994). *Activity limitation and cigarette smoking in the United States: implications for health expectancies*. Paper presented at the Advances in health expectancies: proceedings of the 7th meeting of the International Network on Health Expectancy (REVES), Canberra, February 1994, Canberra.
- Ruwaard, D., & Kramers, P. G. N. (Ed.). (1997). *Volksgezondheid Toekomst Verkenning 1997. De som der delen* (Vol. p 91). The Hague: SDU.
- Slooter, A. J., Tang, M. X., van Duijn, C. M., Stern, Y., Ott, A., Bell, K., Breteleur, M. M., Van Broeckhoven, C., Tatemichi, T. K., Tycko, B., Hofman, A., & Mayeux, R. (1997). Apolipoprotein E epsilon4 and the risk of dementia with stroke. A population-based investigation. *Jama*, 277(10), 818-21.

- SOLVD investigators (1991). Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. *N Engl J Med*, 325, 293-302.
- Stronks, K. (1997). *Socio-economic inequalities in health: individual choice or social circumstances?* Thesis, Erasmus Universiteit Rotterdam.
- Thomas, T., Thomas, G., McLendon, C., Sutton, T., & Mullan, M. (1996). Beta-amyloid-mediated vasoactivity and vascular endothelial damage. *Nature*, 380, 168-71.
- van de Mheen, P. J., Bonneux, L., & Gunning-Schepers, L. J. (1995). Variation in reported prevalences of hypertension in The Netherlands: the impact of methodological variables. [Review]. *J Epidemiol Community Health*, 49(3), 277-80.
- van der Maas, P. J., Schaapveld, K., van der Heide, A., Davidse, W., Bonneux, L., & Barendregt, J. J. (1996). *Prioriteiten in Preventie Onderzoek*. Raad voor Gezondheidsonderzoek.

Summary

Degenerative disease in an aging population

Compression or expansion of disease?

The main research question of this thesis is how morbidity will evolve in an aging population. "Morbidity in a population" is the sum of diseases and the disability they cause. It incorporates all discomfort, brought about by disease: pain, physical disability hampering mobility, as well as depression or cognitive decline. "Population aging" has two causes: one is that the proportion of elderly in the Dutch population will rise in the first half of the next century due to the large size of the post-war birth cohorts, the second is that an increasing proportion of any birth cohort reaches old age. We mostly focus on the latter cause, while not completely ignoring the former.

The main research question is closely related to the 'compression or expansion of morbidity' question, that has been debated for quite some time now. Kramer and Gruenberg started off the debate by positing that morbidity will expand, because medical technology successfully lowers mortality. As a consequence, frail and diseased people survive longer, increasing health care needs. Fries put forward the antithesis that morbidity would compress. He states that the human life span is fixed by biological limits, and that in industrialized societies populations are near those fixed limits. Secondly, he observes the presence of a considerable burden of preventable chronic disease, caused by unhealthy lifestyles. If the incidence of chronic disease can be delayed, by health promotion and other means of primary prevention, but the lifespan is fixed by 'natural death', morbidity will be compressed at the end of life. A third position is held by Manton who argues that medical interventions will slow the progression of degen-

erative disease, resulting in postponed death, and more but less severe disability.

This compression versus expansion debate so far is mostly held in terms of generic morbidity, and only partly on a quantitative level. A more detailed and quantitative approach might be helpful, and therefore we formulate two additional questions. Morbidity and mortality are phenomena determined by incidence and duration of diseases (albeit that among the oldest old specific conditions are hard to distinguish from senescent decay). So the first additional question is what a disease specific approach can contribute to the understanding of compression and expansion of morbidity.

The second additional question originates from the observation that prevalence of morbidity is the resultant of incidence, cure, and mortality, and that such dynamic variables are best described using dynamic modelling. The second additional question, therefore, is how epidemiology and quantitative modelling techniques can be integrated into public health models fit to answer questions of compression and expansion on the disease specific level, and on the level of overall population health.

These two additional questions largely determine the outline of this thesis. We first describe recent and possible future developments in the epidemiology of a number of important diseases, using quantitative modeling techniques (Part II). Next we propose a methodology to combine the disease specific models in a comprehensive model of public health (Part III), and examine the possibilities to derive an indicator of population health status that allows to define and describe compression and expansion of morbidity (Part IV). In addition we look at the impact of prevention (a potential way to reach compression) on health care costs (Part V), reflect on the causes of senescent morbidity and mortality (Part VI), and draw conclusions (Part VII).

An epidemiological point of view

This thesis is based on epidemiology, the science of disease occurrence, and focuses in particular on degenerative diseases. Degenerative diseases are caused by 'age' and 'damage'. Damage is partly caused by the environment, and particularly by unhealthy life styles. Not all damage is preventable, but some is.

Primary prevention is then the avoidance of damage (such as not smoking), secondary prevention is early detection of a disease with the aim of intervening early in the natural history of the disease. Effective treatment may avert death and prevent or delay complications, rehabilitation may prevent permanent disability. Changes in morbidity are the sum of all the effects of all interventions in all diseases, in addition to explained and unexplained autonomous trends. Therefore it is difficult to say what will happen on the general level, unless the disease specific level is described first.

Part II focuses on the disease specific level. Chapters 4 to 7 show changes in cardiovascular morbidity. Because effectivity of treatment of

acute myocardial infarctions has improved dramatically, the increasing numbers of survivors boost numbers of patients with heart disease, at risk for further cardiovascular disease, predominantly heart failure (chapters 4 & 5). In stroke, control of hypertension is in all likelihood one of the main reasons of the incidence decline, but therapeutic progress in the management of the cerebrovascular accident itself remains limited (chapter 6). The consequence is that heart disease is expanding, and stroke morbidity was compressing. "Was" compressing, because the expansion of heart disease patients increases vascular risk, and therefore tends to expand stroke morbidity again. In the same disease group, determined by the same type of degenerative process (atherosclerosis), morbidity trends, which are heavily determined by health care interventions, are opposite. Chapter 8 shows opposite trends in the same disease: in colorectal cancer, incidence is increasing but mortality is declining. These changes are probably determined by earlier diagnosis. While numbers of new and old cancer patients are 'expanding', the prevalence of terminal disease is decreasing and the severity of disease is probably 'declining'.

In chapter 9, we show the effect of an isolated monofactorial risk factor intervention on a typical senescent disease such as hip fractures. As hip fractures are caused by many interacting age related factors, targeting of a single determinant such as bone mineral density is bound to be little efficient, even more so if the intervention (slowing of the fast postmenopausal bone loss) has to precede the presumed effect for many decades: hip fractures occur predominantly after age 80.

Multiple diseases

To describe changes in population health, more is needed than a single disease and a single risk factor. Part III describes how multiple diseases and risk factors can be considered together in a single model. At the heart of this part is "competing risks" theory. Causes of diseases and causes of death can be considered as competing hazards, dependent or independent. If they are considered independent, observed crude risks composed by several disease specific components can be split up in the various specific disease hazards. After elimination of one cause of death, another cause of death will 'substitute' the eliminated cause of death, why we call this competition under conditions of independence also 'substitution of mortality'. A hallmark of 'substitution of causes of death' is that the mortality of other causes of death is increasing, because the age distribution of the population is changing, but that at any age the death rates remain constant.

However, risks are not distributed independently over the population; some have a higher disease risk than others. This heterogeneous disease risk causes 'mortality selection': those with the higher risks die at younger ages, and the survivors at older ages are a selection of the 'strong'. After elimination of a cause of death, mortality selection will ease, and the frail survivors will consequently make the population on average less healthy. Now mor-

tality of other causes will go up, not only because of the changing age distribution, but also because of the increasing risk for disease in the frailer survivors. After age standardization, the death rates still increase. Because of the competition between the dependent risks, we call this 'competition of mortality'.

An example is in chapter 14, that shows the changes in prevalence of diabetes as a consequence of lower cardiovascular mortality. Diabetes is an important risk for cardiovascular death, and lower cardiovascular mortality will increase the prevalence. However, as shown in the same chapter, age standardized prevalence of diabetes does increase, but only weakly so. We therefore conclude that, unless the dependent risk and its consequences are the subject of the research question, in most cases little is lost by making the simplifying independence assumption.

A comprehensive measure of population health status.

In parts II and III we have shown how changing morbidity can be described by changing prevalences of diseases, calculated by single and multi-disease models. But now the problem arises how these changing prevalences can be summed and combined with mortality to an indicator of changes in overall population health status. Population health status is determined by morbidity and mortality from various diseases, and therefore an indicator should provide a common denominator for these various manifestations of ill health.

In the introduction of part IV, we present "time" as such a common denominator. Mortality causes loss of years of life, and similarly years lived with disease can be considered lost. The severity of the disease can be incorporated by the proportion of a year lived with disease that is considered lost: nearly all for very severe disease, only a little for mild disease.

Various methods to construct comprehensive measures of population health status, all based on time as a common denominator, are discussed. The most used measure is 'healthy life expectancy (HLE)', which combines prevalence of ill health and mortality in a life table. The HLE is then the number of years a life table cohort lives 'without disease'. Most implementations of the HLE do not account for disease severity, but for the threshold value used to discriminate health from disease. Typically the years lived with disease are considered completely lost (have value 0), which implies that interventions that are life saving but do not cure have strictly no value, which is contrary to societal preference. Another problem of many HLE implementations is the combination of cross-sectional disease prevalences and mortality: this combination of "stock" and "flow" variables causes biased results in case of swift changes in flows.

There is nevertheless a sound reason why HLE is the most used measure, despite its disadvantages. It is by far the most simple method, requiring

the least data. When 'change' is not an issue, as in a simple description of health status, the HLE is an excellent method. However, as the subject of this thesis is 'change', the HLE is insufficient.

Disability adjusted life years (DALYs) avoid most of the problems of HLE. The aim of DALYs is to evaluate policy decisions. Therefore DALYs are on a disease specific basis, use weights for disease severity and are able to gauge the effects of change by converting prevalences to incidence and duration. But they also use real populations, making DALYs dependent on the age distribution of the population at hand. Moreover, DALYs include discounting of future benefits and costs, and an age-weighting schedule that values years lived by economic productive adults more than years lived by the very young or the very old. This makes the DALY a complex indicator, which may, as chapter 18 shows, yield unexpected results.

It is possible to combine various attractive elements of HLE and DALYs: life table, disease specific approach, and disability weights. The result is a "disability adjusted life expectancy (DALE)". The DALE can be interpreted as the sum of healthy life years and diseased life years (weighted for disease severity) a member of the synthetic life table cohort is expected to live. This implies that the difference between the life expectancy (LE) and the DALE is a cross-sectional measure of the burden of morbidity in a population at a certain period. If that period (LE - DALE) increases, morbidity is expanding, if that period decreases, morbidity is compressing. While not without problems (chapter 20) its properties make the DALE a powerful indicator of population health status.

Prevention and health care costs

An important reason for the interest in compression and expansion of morbidity is the relation with health care costs. Part V presents estimates of the effects of preventive measures on health care costs using public health models. In the health economics literature, we argue in the introduction, estimation is often confused with evaluation, and age with time. Estimation is simply concerned with measuring the amounts of money involved with health care interventions. Evaluation, on the other hand, is concerned with deciding which of these interventions is economically most attractive.

Economic evaluation is complicated by the existence of 'time preference': we prefer to pay tomorrow instead of today, and to receive today instead of tomorrow. When the effects of interventions are spread out over time, time preference makes the alternative interventions as such incomparable. The standard solution to make comparison possible is to calculate the present value of these streams of future costs and benefits by applying discounting: amounts of money get a lower weight in the present value as they are further away in time.

In chapter 23 population health care costs are estimated in the absence of selected diseases as cause of death *and* cause of costs. Elimination of a disease will on the one hand save costs, but on the other generate costs be-

cause of added life years. It turns out that the elimination in only a few non-fatal cases will cause net savings: in most cases the disease specific saving is outweighed by the costs of added life years.

We did not apply discounting, for two reasons. The first is that we are only estimating, and not evaluating an intervention. The second is that in the analysis the life table is interpreted as a stationary population at one point in time. This means that there is no time dimension in the model, only age.

In chapter 24 the distinction between time and age becomes even more clear. We show that smokers are more expensive than non-smokers at all ages, that in the absence of smoking health care costs will nevertheless be higher, but that stopping with smoking is desirable from a purely economic point of view, given a suitable rate of time preference and length of evaluation period. Stopping with smoking increases health care budgets in the long run, because the costs of added life years are higher than the savings from prevented diseases. But initially costs become lower, making the intervention economically more attractive with higher discount rates and shorter evaluation periods.

Senescence from a perspective of theoretical biology

As degenerative diseases can be expected to cause much of future morbidity, a theory of degenerative disease is helpful in predicting future morbidity. Why does the risk for degenerative disease increase at middle age (even as soon as after sexual maturation)? Why is aging nearly synonymous with senescence? In Part VI we discuss the fundamental nature of senescence.

At the heart of evolutionary theory explaining senescence is the observation that old age is rarely reached in natural conditions, due to predation, starvation, and the like. Recent insights from evolutionary theory state that investing in old age is not very useful: there is a trade off between investing energy in offspring versus investing in body maintenance. The evolutionary optimal point will be one where body maintenance falls short to prevent senescence, because an organism investing in offspring will be the more successful. The *disposable soma* theory postulates that body maintenance is geared to attain a natural life expectancy, reached normally in the wild, but not more.

This theory permits predictions about the nature of senescence. First, senescence is predominantly a consequence of neglect. There is no fixed life span in the sense of a genetically programmed end. The natural end of life is probabilistic, caused by stochastic damage and deficient maintenance of a life sustaining system, leading to failure and death. If failure in a life sustaining component can be delayed, senescence rates in other components will go on. If these components cause disability instead of death, disability will replace death. The future of morbidity from degenerative diseases will

be determined by this equilibrium between delay of failure in life sustaining systems and delay of failure in systems responsible for disabling, non-fatal diseases.

Conclusion

We draw conclusions by discussing our three research questions: first the use of modeling, then the usefulness of a disease specific approach. Next the main research question is discussed, and finally we propose priorities for further research.

The use of public health models

Modelers are easily tempted to try to imitate a complex reality as closely as possible, and to expand their models ever further. But complex models are a mixed blessing: as data needs, assumptions, and uncertainties multiply, the intelligibility of models declines, the results can not be validated, let alone be reproduced independently. And, ultimately, even the most complex model is simplistic, compared to reality.

Modern epistemology stresses the central role of theory in the acquirement of knowledge: even an observation cannot be made without a theory. A computer model is nothing but a quantified theory, therefore the rules applicable to them are those of any other theory. A basic rule, known as Ockham's razor, states that all elements which are not strictly needed for an explanation have to be eliminated. The best model is the simplest model capable of answering the research question.

Given this rule, models are useful in public health research for two reasons. The first is that they allow to reconstruct a consistent set of age and time specific incidence, prevalence and mortality for a disease, based on the available figures. The second use is to calculate changes in morbidity. Dynamically determined by flows, changing prevalence of morbidity over time and age is hard to predict without formal modeling.

Why a disease specific approach?

The usefulness of a disease specific approach is well illustrated by the examples in the original articles arguing for either expansion or compression of morbidity. Kramer, defending expansion, takes Down's syndrome, dementia, and other types of chronic mental handicap as examples. Fries looks at cerebrovascular disease and the health consequences of smoking. Both are right.

In this thesis we show how one cardiovascular disease (coronary heart disease) has expanded, while the other (stroke) compressed. In colorectal cancer, numbers of patients are expanding, but severity of disease becomes less. A single scenario applicable to all diseases seems very unlikely, as incidence and prognosis of conditions are dependent on many processes.

Policy makers want to change the evolution of morbidity in a desirable direction. Therefore they need causal models, linking risk factors and diseases with mortality and morbidity. We present such models and show how morbidity compresses if risk factors for a disease decline, and how morbidity expands if the prognosis improves. But we could do so only by taking disease specific characteristics into account.

Compression or expansion?

How will morbidity evolve? The right answer is: it depends. Compression of morbidity is caused by increasing mortality, decreasing incidence and decreasing severity. Expansion of morbidity is caused by the opposite. Increasing mortality is by far the most effective way of decreasing morbidity, illustrating that compression of morbidity cannot be an aim in itself.

We find more indications for expanding than for compressing morbidity. Decreasing mortality of cardiovascular diseases and increasing case detection of cancers are powerful generators of morbidity. Any delay of mortality will lead to increased disability of non-fatal senescent conditions, if they are independent of the delayed cause of death.

However, there are also processes compressing morbidity. Stopping with smoking is the most striking opportunity. And while life saving interventions in health care generally lead to expansion of morbidity, health care interventions may also cause compression. The Burden of Disease project showed that the epidemiological transition coincides with compression of disability, in particular of infectious diseases and maternal causes of disability. And finally, it is no Law of Nature that no progress can be made in the therapy and prevention of disabling degenerative diseases: research directed at preventable determinants of chronic psychiatric disease and joint disease should have the highest priority.

How morbidity will change is uncertain, and will depend to some extent on policy choices in public health and health research. If relatively more resources are spent at prevention and treatment of disability to improve the quality of life, and less to life extension, morbidity might compress.

Compression of morbidity is often thought to cause compression of health care costs. The relation is not so simple, however: while stopping with smoking compresses morbidity (chapter 19), it expands health care costs (chapter 24). There exists a poorly understood paradox between needs and costs. The existing morbidity is the result of not prevented disease and prevented death. If health care can successfully treat, it compresses morbidity, but it also uses resources. This leads to the paradox that morbidity compresses, while health care costs expand. Indeed, one of Fries' prime examples, stroke, is an instance of that paradox. At least part of the decline in stroke is caused by hypertension control, which prevents stroke disability, but consumes considerable resources. There is little doubt that needs and demands for health care will go on increasing, even as morbidity

compresses. The only exception might be the prevention or cure of disabling but non-fatal conditions that currently require a lot of care.

Priorities for further research

Our priorities for further research consist of one general recommendation and three specific research fields. The general recommendation is that Dutch public health policy needs a national "Burden of Disease" project. Consistent age specific estimates of incidence, prevalence and mortality of major diseases are badly needed. Together with the project describing "Costs of Diseases in the Netherlands", policy makers would possess information about all three dimensions needed to determine efficiency: mortality, morbidity and costs. To allow assessment of temporal trends, such national studies should be repeated on a regular basis.

The first research field of interest to be explored is demographic epidemiology. Our thesis shows several examples how the embedding of epidemiology in a demographic framework allows to explore the consequences of epidemiological change for population health. The second field is the relation between changes in population health status and health care costs. In part V are two examples, that show the consequences of a changing epidemiology for health costs. Further work in this direction should allow to identify where money is spent less efficiently and where gains in population health status could be achieved economically. The third field relates the consequences of the evolutionary basis of senescence to demographic epidemiology. "Biodemography", which studies human mortality using concepts of theoretical biology, has been founded by demographers and biologists. Their insights certainly can enlighten epidemiologists, but just as likely the research field of biodemography would profit from epidemiology, particularly the epidemiology of senescent diseases.

Samenvatting

Degeneratieve ziekte in een verouderende bevolking

“Compressie of expansie van ziekte?”

De kernvraag van dit proefschrift is de vraag hoe ziekte verandert in een verouderende bevolking. Ziekte in de bevolking is de optelsom van alle vormen van ongemak, die hun oorzaak vinden in een aandoening. Dit geldt zowel voor fysieke beperkingen die bewegingsvrijheid of activiteiten van het dagelijks leven hinderen als voor pijn, depressie of verstandelijke achteruitgang.

De vraag over ‘verandering in ziekte in de bevolking’ komt overeen met de door Kramer, Gruenberg en Fries ingeleide discussie over ‘expansie of compressie van morbiditeit’. Kramer en Gruenberg verdedigden de expansiegedachte. De kern van hun stelling was dat de levensverwachting toenam door de toenemende successen van de medische technologie, en dat daardoor meer mensen chronische aandoeningen langer overleefden. Als antwoord op deze eerder pessimistische visie formuleerde Fries zijn befaamde omgekeerde stelling. Fries betoogde dat de levensverwachting in ontwikkelde landen niet veel meer kon toenemen. Daarentegen bleef er in die geïndustrialiseerde wereld een groot potentieel aan voorkombare degeneratieve aandoeningen. Als de incidentie van degeneratieve ziekte kon uitgesteld worden, maar de sterfte niet, dan volgde daaruit natuurlijkerwijze ‘compressie’ van ziekte aan het einde van het leven. Fries veronderstelde dus, in tegenstelling tot Kramer en Gruenberg, dat de incidentie van degeneratieve ziekten kon afnemen, maar de sterfte niet. Manton zal later opmerken dat de voorstanders van zowel de expansie- als de compressie-theorie ziekte en sterfte als onafhankelijke processen behandelen. In de expansietheorie worden de mensen ziek zoals vroeger, maar sterven ze op

hogere leeftijd, in de compressietheorie blijven mensen langer gezond, maar leven ze toch niet langer.

Ter beantwoording van de vraag van expansie of compressie stellen ons daarom twee bijkomende vragen.

De eerste vraag is wat de bijdrage is van een ziektespecifieke aanpak tot het vraagstuk van de toekomstige morbiditeit. Ziekte en sterfte in de bevolking wordt uiteindelijk bepaald door de incidentie, duur en letaliteit van specifieke aandoeningen (al kunnen op hogere leeftijd de bijdragen van specifieke aandoeningen en aftakeling slecht gescheiden worden).

De tweede vraag is hoe wiskundige modellen en epidemiologische gegevens kunnen geïntegreerd worden om uitspraken te kunnen doen over expansie of compressie van ziekten.

Een epidemiologische invalshoek

Door zich te baseren op aantallen en aard van de aandoening, baseert dit proefschrift zich op de epidemiologie: de leer van het voorkomen der aandoeningen. Voor een pure beschrijving van de staat van de volksgezondheid zou een beschrijving van de gezondheidstoestand van de betreffende bevolking voldoende zijn. Maar als we deze veranderingen willen begrijpen, en indien mogelijk beïnvloeden, hebben we een verklarend model nodig dat een verband legt tussen oorzaak en gevolg. Een medisch-epidemiologisch model is dan het meest voor de hand liggend.

In een hoog ontwikkelde gemeenschap na de demografische transitie, gekenmerkt door een lage sterfte en een hoge levensverwachting, zijn de belangrijkste oorzaken van morbiditeit psychische stoornissen en degeneratieve aandoeningen. Dit proefschrift spitst zich toe op de degeneratieve aandoeningen. Degeneratieve aandoeningen worden veroorzaakt door (leeftijd en schade. We worden geboren met een bepaald genetisch gezondheidspotentieel en worden gedurende ons leven aangetast door beschadiging. Deze schade is deels onvermijdelijk, deels veroorzaakt door vermijdbare ongezonde omstandigheden.

Primaire preventie bestaat vooral uit het voorkomen of verminderen van de milieuschade (bv. het voorkomen van roken), secundaire preventie uit het vroegtijdig opsporen en behandelen van degeneratieve aandoeningen (bijv. kankerscreening). Behandeling kan de gevolgen van een aandoening voorkomen of uitstellen (bijvoorbeeld thrombolysen na een hartinfarct). Revalidatie kan invaliditeit voorkomen of verminderen (zoals na een beroerte) en palliatieve behandeling kan de nog overblijvende levenskwaliteit bewaren. "Verandering in morbiditeit" is de som van de effecten van al deze interventies.

In deel II van dit proefschrift staan meerdere voorbeelden die het nut van een epidemiologische, ziektespecifieke invalshoek illustreren. Hoofdstukken 4 tot 7 bespreken veranderingen in cardiovasculaire morbiditeit. Er zijn steeds meer patiënten met een chronische hartziekte, omdat behande-

ling en prognose van een acute hartziekte zo sterk verbeterd is. De patiënten overleven, maar zijn daarom niet 'genezen' van hun chronische aandoening. Daarentegen was er tot 1987 een sterke afname van patiënten met een beroerte, mede door behandeling van hoge bloeddruk. Bij verschillende types van hart- en vaatziekten, gedeeltelijk gedetermineerd door het hetzelfde degeneratieve proces, atherosclerose, constateren we een tegengestelde evolutie: 'expansie' van chronische hartziekte en 'compressie' van beroerte. Sinds 1987 lijkt de expansie van hartziekte uiteindelijk aanleiding te geven tot een afbreken van de compressie in beroerte.

Hoofdstuk 8, over dikke-darmkanker, toont hoe de toenemende incidentie en de afnemende sterfte alleen maar goed verklaard kunnen worden door toenemende vroegdiagnose. Daardoor verandert wel het karakter van de morbiditeit van dikke darmkanker: er komen steeds meer nieuwe en oude patiënten met kanker, maar minder terminaal zieken. Bij dezelfde aandoening is er dus een expansie van aantallen, maar een compressie van ernst. Hoofdstuk 9 toont het belang van 'leeftijd' voor heupfracturen. Waar het probleem van heupfracturen nogal eens herleid wordt tot osteoporose, blijkt veroudering als oorzaak van vallen een veel belangrijker determinant. Opsporen en/of preventief behandelen van osteoporose lijkt hierdoor weinig zoden aan de dijk te zetten.

Het nut van volksgezondheidsmodellen

"Volksgezondheidsmodel" betekent in dit proefschrift een model dat een wel bepaald gezondheidsprobleem beschrijft in termen van vóórkomen van ziekte en sterfte in de algemene bevolking. Omdat dit complex is, is het formuleren van een volksgezondheidsmodel een evenwichtsoefening tussen eenvoud en complexiteit, tussen wetenschappelijk herleiden en inhoudelijke relevantie. In dit proefschrift dienen modellen een dubbel doel: ze brengen gegevens uit verschillende bronnen met elkaar in verband in een consistent geheel, en maken dynamische veranderingen van morbiditeit over tijd en leeftijd zichtbaar

'Metasynthese'

Veranderingen in de epidemiologie van degeneratieve aandoeningen worden aangedreven door veranderingen in 'stromen'. Nieuwe patiënten stromen binnen door een incidentie, en oude patiënten stromen buiten door sterfte of (zelden relevant in degeneratieve aandoeningen) door genezing. Wie veranderingen in de epidemiologie wil volgen heeft bijgevolg gegevens nodig over de 'stromen' van incidentie en sterfte. Terwijl sterfte aan belangrijke aandoeningen deels kan afgeleid worden uit het register van de doodsoorzaken, zijn gegevens over veranderingen in incidenties bijzonder lastig te vinden. Het Nederlandse Ziekenhuisregister registreert echter sinds de jaren '70 diagnoses bij ontslag. Voor sommige aandoeningen, zoals acute hartinfarcten of opnames voor kanker, is dit een vrij getrouwe

weergave van de totale aantallen klinische 'gebeurtenissen', zeker op niet te oude leeftijd. Indien we een 'model' van de ziekte hebben dat de ziektegeschiedenis kwantitatief beschrijft door incidenties, recidieven en sterfte, kunnen we een incidentie 'terugrekenen' op basis van deze gebeurtenissen. Uit de combinatie van incidentie en sterfte kan dan weer de prevalentie berekend worden. De tijdstromen uit het ziekenhuisregisters gekoppeld aan de tijdstromen uit het sterfteregister leveren dan veranderingen in prevalenties op. Voorbeelden van deze terugrekenen-modellen kunt u eveneens vinden in deel II (met name de hoofdstukken 4, 5, 6 en 8). Een andere benadering van hetzelfde probleem staat in hoofdstuk 13, waar een incidentie van diabetes berekend wordt op basis van schattingen van prevalentie en sterfte. Dit combineren van gegevens uit allerhande bronnen om met behulp van een rekenmodel de ziektegeschiedenis te reconstrueren wordt in analogie met 'meta-analyse' ook wel 'meta-synthese' genoemd.

Veranderingen over tijd en leeftijd.

Morbiditeit is een voorraad, afhankelijk van een in- en uitstroom. Zoals de waterstand van een meer afhankelijk is van het water dat er is ingestroomd en nog niet is uitgestroomd, zo is de huidige morbiditeit afhankelijk van voorbije stromen van incidentie, genezing en sterfte. Voor degeneratieve aandoeningen in een menselijke bevolking betekent dit ook dat de morbiditeit het gevolg is van incidentie, opgetreden op jongere leeftijd en sterfte, nog op te treden op latere leeftijd. De dynamische veranderingen over tijd en leeftijd, en de gevolgen daarvan voor een reële bevolking, zijn bijzonder moeilijk te schatten zonder een formeel model, zeker als we meerdere oorzaken van ziekte en sterfte tegelijkertijd beschouwen. Vooral de exponentiële toename van incidentie en sterfte van degeneratieve aandoeningen met de leeftijd spelen het menselijk verbeeldingsvermogen parten. Voor de niet-ingewijde blijft het verbazing wekken dat wanneer alle oorzaken van sterfte aan hart- en vaatziekten zouden voorkómen worden, samen de helft van alle doodsoorzaken, de levensverwachting toch maar met een vijftal jaren zal toenemen.

Méer ziekten

Om veranderingen in volksgezondheid te beschrijven is méér dan één ziekte en méér dan één determinant nodig. Hoe meerdere ziekten, meerdere risicofactoren en ziekten die tegelijkertijd een risicofactor voor andere ziekte vormen kunnen gemodelleerd worden staat formeel beschreven in deel III. De harde kern van deze methodensectie bestaat uit de formele beschrijving van ziekte en sterfte als concurrerende risico's. Oorzaken van ziekte en sterfte kunnen als onafhankelijk van elkaar beschouwd worden; via eenvoudige wiskundige technieken kunnen risico's dan uitgesplitst worden in hun samenstellende specifieke *hazards*. Bij uitschakelen van de ene doodsoorzaak, zal een andere ze dan 'vervangen' (waarom we dit ook

'vervangende sterfte' noemen). Anderzijds is het risico op ziekte en sterfte niet homogeen verdeeld over een bevolking: de een heeft een hoger risico dan de ander. Dit geeft aanleiding tot sterfte-selectie: mensen met een hoger risico zullen eerder overlijden. Bij uitschakelen van een doodsoorzaak, zal sterfte-selectie afnemen, en zal de overlevende bevolking een hoger risico hebben op afhankelijke doodsoorzaken. Daardoor zal sterfte aan andere oorzaken toenemen (ook na leeftijdsstandaardisatie). Wegens afhankelijkheid en de competitie tussen risico's noemen we dit 'concurrerende sterfte'.

Het expliciet modelleren van afhankelijkheden is ongetwijfeld realistischer, maar er dient een hoge prijs voor betaald te worden. De gegevensbehoefte en de modelcomplexiteit neemt toe met het exponent van het aantal risicofactor-toestanden dat men wil beschouwen. Op basis van Ockham's scheermes (hou verklarende modellen zo eenvoudig als mogelijk) is het modelleren van afhankelijkheden bijgevolg enkel aan de orde indien dit relevant is. Een voorbeeld is in hoofdstuk 14, dat veranderingen in sterfte aan hart- en vaataandoeningen gebruikt om veranderingen in prevalentie van Diabetes Mellitus te verklaren; hart- en vaataandoeningen zijn een belangrijke doodsoorzaak bij Diabetes.

Van prevalenties naar een integrale volksgezondheidsmaat

In delen II en III hebben we aangetoond hoe we de veranderende morbiditeit van specifieke aandoeningen kunnen beschrijven met behulp van prevalenties, verkregen door ziektemodellen, en hoe we meerdere aandoeningen en determinanten kunnen combineren in een algemeen volksgezondheidsmodel. Het optellen van prevalenties van aandoeningen is echter het befaamde probleem van het optellen van de appels en de peren; eczeem is niet direct vergelijkbaar met longkanker. Bovendien toonden we in hoofdstukken 4, 5 en 7 aan hoe de morbiditeit kan toenemen door de dalende sterfte. Het zou de grote successen van de cardiologie geen recht doen indien we die toenemende morbiditeit als louter negatief voor de volksgezondheid beschrijven. Het doel van een geïntegreerde maat is dus het combineren van ziekte (morbiditeit) en sterfte in één maat, zodat zinvolle vergelijkingen tussen aandoeningen mogelijk worden. We moeten een gemeenschappelijke noemer identificeren voor zowel de ernst van aandoeningen als voor sterfte, zodat prevalenties van verschillende aandoeningen en verloren levensjaren door sterfte kunnen opgeteld worden.

In de inleiding van deel IV (hoofdstuk 15) presenteren we 'tijd' als een gemeenschappelijke noemer. Wie overlijdt verliest levensjaren door te sterven, maar wie ziek is ook: de zieke is beperkt in zijn gewone activiteiten. Wie zich niet erg ziek voelt, beschouwt dit waarschijnlijk niet als een groot verlies. Maar wie niet in staat is om wat dan ook uit te voeren, en zich bovendien ellendig voelt, zal weinig 'waarde' hechten aan de tijd dat hij

ziek is. De levensverwachting kan dan verminderd worden met de tijd doorgebracht met ziekte, gewogen voor de gemiddelde ernst van die ziekte. Het resultaat is '*disability adjusted life expectancy (DALE)*' of voor morbiditeit gecorrigeerde levensverwachting, de som van de levensjaren doorgebracht in goede gezondheid en doorgebracht met ziekte, gewogen voor de gemiddelde ernst van die ziekte. Het verschil met de ongecorrigeerde levensverwachting is dan een maat voor de gemiddelde duur en de gemiddelde ernst van ziekte in de gemeenschap. De vraag over expansie of inkrimping van ziekte kan bijgevolg beantwoord worden door de verandering in verloren levensverwachting door ziekte te meten: de DALE minus de LE (*life expectancy*, de ongecorrigeerde levensverwachting). Neemt deze toe, dan neemt morbiditeit in de gemeenschap toe, neemt deze af dan neemt morbiditeit af.

Methoden en problemen

In deel IV worden de verschillende methoden en problemen beschreven van geïntegreerde gezondheidsmaten, zoals de gezonde levensverwachting of de *Disability Adjusted Life Years (DALY)*; niet te verwarren met DALE). De gezonde levensverwachting combineert prevalenties van 'ziekte' met sterfte in een sterftetafel. De prevalenties zijn niet gewogen voor de ernst van de aandoening, wat betekent dat er een drempelwaarde moet gekozen worden tussen 'gezond' en 'ziek'. Daardoor komt gezonde levensverwachting in families, afhankelijk van de onvermijdelijk arbitraire keuze van een drempelwaarde. Bovendien is het gestandaardiseerd opmeten van dergelijke drempels in verschillende bevolkingen en over verschillende tijdstippen nagenoeg onmogelijk, wat de vergelijkbaarheid ernstig in gedrang brengt. Omdat de jaren doorgebracht met ziekte geen waarde hebben, wordt er evenmin waarde toegekend aan levensverlenging bij zieken, ook al is hun levenskwaliteit niet noodzakelijk heel laag.

Tijd en leeftijd zijn de essentiële dimensies van iedere geïntegreerde maat, en zoals steeds bij de determinanten tijd, leeftijd en geboortecohorte dienen er keuzen gemaakt te worden tussen perspectieven. Indien men de huidige prevalenties combineert met de huidige sterfte, zoals bij het berekenen van een gezonde levensverwachting volgens de Sullivan methode, vergelijkt men een 'voorraad', opgebouwd door instroom uit het verleden, met sterfte, een uitstroom uit het heden (hoofdstuk 16). Indien er grote veranderingen hebben plaats gehad in deze stromen in het recente verleden, kan dit tot foutieve schattingen leiden.

Door zijn relatieve eenvoud levert de gezonde levensverwachting een bruikbare beschrijving van de huidige staat van de volksgezondheid, maar de gezonde levensverwachting is niet geschikt om veranderingen in die staat betrouwbaar weer te geven. Omdat de methode niet dynamisch is, is een historische beschrijving van trends over de tijd onbetrouwbaar, en omdat er geen ziektespecifiek causaal model aan ten grondslag ligt zijn effecten van een veranderend gezondheidsbeleid niet voorspelbaar.

De DALY is ontwikkeld door Murray en Lopez om de effectiviteit van bundels interventies weer te geven. De DALY voorkomt de methodologische problemen van de gezonde levensverwachting door ziektespecifieke incidentie-duur modellen te gebruiken, en te wegen voor de ernst van de aandoening. De DALY gebruikt vergelijkbare methoden met de DALE, maar door zijn ander doel zijn er ook grote verschillen. De DALY is afhankelijk van de leeftijdsstructuur van de bevolking waarop de maat wordt toegepast, en om redenen van vergelijkbaarheid wordt een ideële standaard-sterftetafel gebruikt. De DALY incorporeert zowel een leeftijdsvoorkeur als een tijdsvoorkeur: gewonnen levensjaren op kinderleeftijd of op oude leeftijd worden minder zwaar gewogen, en levensjaren gewonnen in de toekomst wegen minder zwaar naarmate die toekomst verder verwijderd is (discontering). Er bestaan hier solide economische redenen voor, al blijkt de combinatie van expliciete leeftijdsgewichten in een sterftetafel (die eveneens weegt voor leeftijd door meer verloren levensjaren toe te kennen naarmate de patiënt jonger overlijdt) tot andere resultaten te komen dan verwacht (hoofdstuk 18).

Tot slot is er het steeds weerkerend probleem van afhankelijkheden. Co-morbiditeit is een vorm van afhankelijkheid, waar de ernst van de gezondheidstoestand afhangt van meer dan één aandoening. Zoals voordien beschreven betekent rekening houden met afhankelijkheden dat complexiteit en gegevensbehoefte toeneemt met het exponent van het aantal ziekte-toestanden dat men wil beschrijven. Hoofdstuk 20 toont dat het expliciet modelleren van co-morbiditeit weinig effect heeft op de DALE; rekening houden met co-morbiditeit is dus enkel noodzakelijk indien co-morbiditeit het onderwerp van studie uitmaakt.

Volksgezondheidsmodellen en kosten van de gezondheidszorg

Deel V bestaat uit economische analyses ingebed in volksgezondheidsmodellen. Nu dient ook economische theorie geïntegreerd te worden in de epidemiologische en demografische technieken. Economische evaluaties beschrijven keuzen in het gebruik van schaarse middelen. Indien kosten en baten van verschillende investeringen optreden op verschillende momenten in de toekomst moeten we rekening houden met tijdsvoorkeur. *Ceteris paribus* verkiezen we volgend jaar te betalen in plaats van vandaag, en morgen te genieten van een betere gezondheid in plaats van volgend jaar.

Indien het doel echter is kosten te schatten, zoals in hoofdstuk 23 dat de kosten van levensverlenging schat, speelt tijdsvoorkeur geen rol. Kosten zijn sterk leeftijdsafhankelijk, zo aanzienlijk dat bij het elimineren van fatale aandoeningen, zoals hart- en vaatziekten en kanker, de gezondheidszorgkosten paradoxaal toenemen. De kosten van de geëlimineerde aandoening vallen weg, maar de bevolking verouderd door de afgenomen sterfte.

Schatten en evalueren komen beiden aan bod in hoofdstuk 24 over de kosten van roken. Rokers kosten meer aan gezondheidszorg dan niet-rokers zolang ze leven, maar niet-rokers kosten meer omdat ze langer leven. Wie verwacht dat niet roken de kosten voor de gezondheidszorg zal verlagen, zal dus bedrogen uitkomen, máár enkel op lange termijn. Het duurt bijna twee decennia vooraleer de kosten van levensverlenging opwegen tegen de besparingen door minder ziekte. Omdat de besparingen vóórlopen op de kosten, is niet-roken economisch voordelig bij een redelijke tijdsvoorkeur. Het kost meer, maar we waarderen die late kosten minder dan die eerdere besparingen.

Leeftijd vanuit een fundamenteel biologisch perspectief

Toekomstverwachtingen over verouderingsgebonden aandoeningen moeten uiteindelijk gebaseerd zijn op een theorie over veroudering. Deel 6 bespreekt de meest fundamentele theorie over de oorzaak van veroudering, de evolutieleer. Veroudering is een kenmerk van alle hogere diersoorten, er moet een reden voor zijn. De kernreden is dat de levensloop in natuurlijke omstandigheden beperkt is door hongersnood, ongevallen, infectieziekten, of andere jagers (inclusief onze soortgenoten). De prehistorische mens bereikte zelden een hoge leeftijd. Daardoor stond het genoom, verantwoordelijk voor het fenotype op hogere leeftijd, niet onder selectiedruk en konden fouten in de genetische software, die hun effect slechts hebben op hoge(re) leeftijd, zich ongestraft opstapelen.

Naast deze accumulatie van mutaties, schadelijk op hoge leeftijd, had het geringe evolutionaire belang van een hoge ouderdom nog een tweede effect. Energie is schaars. Investeren in hoge leeftijd is verloren vanuit een evolutionair standpunt, als die hoge leeftijd niet omgezet wordt in meer leefbaar nageslacht. Daarom worden energievragende onderhoudsprocessen verwaarloosd, als die energie beter kan aangewend worden voor de voortplanting.

Uit deze theorie volgen een aantal voorspellingen over de aard van het verouderingsproces. Het einde van het leven is niet bepaald door een 'hard' genetisch programma, maar door een 'zacht' probabilistisch einde, waar stochastische schade en gebrekkig onderhoud uiteindelijk leidt tot falen van een systeem, noodzakelijk om te overleven. Gezien velerlei systemen op velerlei niveaus noodzakelijk zijn, zijn het aantal betrokken genen en gen-interacties astronomisch hoog, en lijken genetische ingrepen niet echt een optie. Door de aard van het evolutionaire optimaliseringsproces is er een grote variatie in de snelheid van verouderen, zowel in bevolkingen als in systemen binnen één individu. Als het falen van één levensnoodzakelijke component kan uitgesteld worden, zal veroudering blijven optreden in de andere componenten. De toekomst van de morbiditeit wordt dus bepaald door het evenwicht tussen uitstel van falen in systemen verantwoordelijk

voor degeneratieve, niet fatale aandoeningen en uitstel van falen van levensnoodzakelijke componenten, leidend tot de dood.

Besluit

In het begin van dit proefschrift hebben we ons de vraag gesteld hoe morbiditeit zou evolueren in de toekomst. We kozen voor een ziektespecifieke aanpak omdat dit ons inzicht kon bieden in de onderliggende redenen van verandering. Om een dergelijke ziektespecifieke aanpak mogelijk te maken gebruikten we rekentechnieken om gegevens uit allerhande bronnen te synthetiseren in ziektemodellen, pasten we die modellen toe om veranderingen over tijd en leeftijd te berekenen en ontwierpen we een methode om die veranderingen in morbiditeit en mortaliteit te aggregeren in een unieke maat voor de staat van de volksgezondheid, de Disability Adjusted Life Expectancy (DALE). We zullen eerst het gebruik van rekenmodellen bespreken, en vervolgens het nut van een ziektespecifieke aanpak. Dan komen de belangrijkste resultaten aan de orde. Tot slot zullen we ons afvragen welke de prioriteiten zijn voor toekomstig onderzoek.

Het nut van volksgezondheidsmodellen

Zonder de PC en de moderne software zou dit proefschrift waarschijnlijk niet mogelijk geweest zijn, maar de enorme uitbreiding van mogelijkheden is ook een giftig geschenk. De verleiding is groot om een complexe realiteit steeds getrouwer te trachten weer te geven, en modellen steeds verder uit te breiden. De honger naar gegevens blijkt gauw niet te stillen. Naarmate de veronderstellingen en de onzekerheden zich vermenigvuldigen, wordt het model minder doorgrondelijk, minder valideerbaar en minder reproduceerbaar door onafhankelijke onderzoekers. En uiteindelijk blijkt ook het meest complexe model te simplistisch in vergelijking met de realiteit. Het eenvoudige basisprincipe van verklarende modellen werd het meest radicaal geformuleerd door de 13 de eeuwse monnik William van Ockham: verwijder alle elementen uit de verklaring die niet strikt noodzakelijk zijn.

De tegenstelling met het naïeve geloof in de kracht van theoretische modellen wordt geleverd door het naïeve geloof in het geobserveerde gegeven. In de moderne wetenschapsfilosofie is het echter een vaststaand feit dat er zelfs geen observatie gemaakt kan worden zonder theoretische context. Computermodellen leveren een dergelijke theoretische gekwantificeerde context. De kwaliteit van een model wordt geleverd door de mogelijkheid zijn voorspellingen te verwerpen, wat de filosofische grondreden is om modellen doorzichtig te houden.

Volksgezondheidsmodellen zijn dus rekenmodellen die voorspellingen leveren over ziekte en sterfte in een bevolking op basis van een gekwantificeerde theorie. Dit proefschrift identificeerde een dubbel doel. Een eerste doel is het opbouwen van een consistente set gegevens over de des-

criptieve epidemiologie van een aandoening. Het rekenmodel dwingt tot consistentie, door incidentie, prevalentie en sterfte met elkaar te verbinden. Een tweede doel is het in beeld brengen en berekenen van lastig te doorzien dynamieken. Vooral verschuivingen over tijd en leeftijd van tijds- en leeftijdsafhankelijke parameters zijn niet gemakkelijk intuïtief te bevatten.

Waarom een ziektespecifieke aanpak?

Het nut van de ziektespecifieke aanpak wordt goed geïllustreerd door de artikelen te bekijken die de basis vormden van de discussie over expansie en inkrimping van ziekte. Kramer, die de thesis van expansie van ziekte verdedigde, legde de nadruk op Down, dementie en andere vormen van verstandelijke handicap. Fries nam beroertes als voorbeeld, die sterk zijn gedaald. Beiden hadden op basis van het door hun gekozen voorbeeld 'gelijk'. In dit proefschrift toonden we hoe de ene hart- en vaataandoening toenam (hartziekte), terwijl de andere afnam (beroerte). In dikke-darmkanker neemt het aantal patiënten toe, maar het aantal terminaal zieken neemt af. Een uniek scenario is onwaarschijnlijk omdat incidentie, beloop en ernst van aandoeningen afhankelijk zijn van velerlei beïnvloedbare processen.

Een beleidsmaker wil enerzijds een idee hebben over de verwachte evolutie, anderzijds wil hij die evolutie waar mogelijk beïnvloeden naar een gewenste richting. Dit kan niet zonder een idee te hebben van de causale verbanden, die bepaald worden door leeftijd, geslacht, risicofactor en uiteindelijk aandoening. Zowel sterfte als morbiditeit wordt uiteindelijk veroorzaakt door ziekte of aftakeling, en kunnen dus niet onafhankelijk van elkander behandeld worden, zoals dit bij Fries of Kramer het geval is.

In dit proefschrift staan meerdere specifieke voorbeelden hoe morbiditeit 'expandeert' door verbeterende prognose verbetert en dalende sterfte, en hoe morbiditeit 'comprimeert' als risicofactoren voor die aandoening afneemt.

Compressie of expansie?

Hoe zal morbiditeit nu evolueren? Het juiste antwoord is, dat dat er van af hangt. De grootste bronnen van morbiditeit in Nederland zijn aangeboren zwakzinnigheid, chronische psychiatrische aandoeningen en veroudering. De nadruk lag in dit proefschrift op verouderingsgebonden aandoeningen. We identificeerden meerdere epidemiologische processen die kunnen leiden tot expansie of tot compressie. Compressie van morbiditeit wordt veroorzaakt door toenemende sterfte, afnemende incidentie, of afnemende ziekte-ernst. Het feit dat toenemende sterfte de gemakkelijkste manier is om ziekte te comprimeren toont dat compressie van ziekte geen doel op zich is. Toenemende morbiditeit wordt veroorzaakt door toenemende incidentie (ook door toegenomen diagnostische activiteit), afgenomen sterfte en toenemende ernst (dit laatste staat meestal in verband met uitgestelde sterfte).

Over het algemeen zijn er meer aanduidingen voor expansie dan voor compressie. De dalende sterfte aan hart- en vaataandoeningen en de toenemende vroegdiagnose bij kanker is een krachtige motor voor expansie van chronisch zieken. Ieder uitstel van sterfte zal steeds aanleiding geven tot toename van die verouderingsgebonden aandoeningen die onafhankelijk zijn van de dalende sterfte-oorzaken. Er zijn echter ook mogelijkheden tot compressie. Stoppen met roken biedt de grootste kansen.

Terwijl de gezondheidszorg deels aanleiding geeft tot expansie, door verbetering van de prognose, is gezondheidszorg ook verantwoordelijk voor compressie, door genezing (vooral van invaliderende infectieziekten) of het verminderen van de ernst van de aandoening. Murray en Lopez tonen in het *Burden of Disease* project dat de epidemiologische transitie gepaard gaat met compressie van morbiditeit, vooral van infectieziekten en moederlijke oorzaken. Het aandeel van de gezondheidszorg hierin is niet gering. En tot slot is het niet een natuurwet dat er geen voorkoombare determinanten van niet-fatale aandoeningen zijn. Onderzoek naar voorkoombare determinanten van chronische psychische aandoeningen en ziekten van het bewegingsstelsel hoort de hoogste prioriteit te genieten op de onderzoeksagenda.

Morbiditeit en kosten van de gezondheidszorg

Het debat over compressie en expansie van ziekte had mede de kosten van de gezondheidszorg als achtergrond. Er bestaat een tot nog toe slecht herkende paradox tussen zorgbehoefte en gezondheidszorgkosten. De bestaande morbiditeit is het resultaat van voorkomen sterfte en niet voorkomen ziekte. Het is deels de taak van de gezondheidszorg morbiditeit (en mortaliteit) te comprimeren. Waar dit mogelijk is, ontstaat vraag naar gezondheidszorg, wat leidt tot de paradox dat compressie van ziekte gekocht wordt met expansie van gezondheidszorg. Het type voorbeeld van compressie, beroerte, is gedeeltelijk verworven door medicamenteuze behandeling van hoge bloeddruk. Een ander voorbeeld van deze paradox uit ons proefschrift is stoppen met roken. Terwijl stoppen met roken de morbiditeit doet afnemen, doet het de kosten van de gezondheidszorg toenemen. Sommige aandoeningen kunnen behandeld worden. Deze geven dan geen aanleiding meer tot morbiditeit, maar wel nog tot behandelkosten.

Hoe morbiditeit in de toekomst zal evolueren, is onzeker, en zal deels van keuzen in de gezondheidszorg en het gezondheidszorgonderzoek afhangen. Naarmate meer middelen zullen besteed worden aan het verbeteren van de levenskwaliteit (*adding life to years*) en minder aan levensverlenging (*adding years to life*) zal de morbiditeit afnemen. Maar dat de behoefte aan en de vraag naar gezondheidszorg daarbij zal toenemen, staat buiten kijf: er is geen lineair verband tussen morbiditeit en kosten van gezondheidszorg.

Hoe nu verder?

Gegeven de hoeveelheid middelen besteed aan gezondheidszorg en het belang dat de mensen hechten aan gezondheid, is het merkwaardig dat de kennis over het vóórkomen van de grote oorzaken van ziekte zo beperkt en zo versnipperd is. Terwijl verklarende epidemiologie grote successen heeft geboekt (en daar terecht trots op is), was er weinig aandacht voor goede beschrijvende studies. Het Nederlandse volksgezondheidsbeleid heeft daarvoor behoefte aan een nationaal "Burden of Disease" project. De eerste Nederlandse studie die gewichten voor de ernst van de belangrijke ziekten bepaalt is gepubliceerd, nu zijn er nog consistente schattingen nodig van incidentie, prevalentie en sterfte van deze ziekten. Er is grote behoefte aan een dergelijke nationale studie, die met een zeker ritme zou moeten herhaald worden, samen met de studie over "Kosten van Ziekten in Nederland". Voor het eerst zal dan gelijktijdig informatie bestaan over de drie voor volksgezondheid relevante assen: sterfte, morbiditeit en kosten.

Terwijl er in onderzoek de trend bestaat om steeds verder te specialiseren, en steeds meer deelgebieden af te splitsen, is er evenzeer behoefte aan onderzoek dat de grote lijnen weer opneemt en de verschillende vakgebieden weer integreert. Dit onderzoek toonde het belang van het leggen van verbanden tussen epidemiologie, economie en demografie, en uiteindelijk ook met theoretische biologie. Het inbedden van epidemiologie in demografische technieken leverde krachtige modellen om vraagstukken over de volksgezondheid te beantwoorden. Het combineren van de demografische epidemiologie met de gegevens over kosten van ziekten legde verbanden tussen de staat van de Nederlandse volksgezondheid en gezondheidszorgkosten, en zou in de toekomst richtingen moeten kunnen aanwijzen waar maximale gezondheidswinst het meest economisch kan bereikt worden. Fundamentele biologische theorie over de natuurlijke geschiedenis van veroudering maakte voorspellingen mogelijk over de aard van ziekte en sterfte (en dus epidemiologie) op hoge leeftijd.

Volksgezondheid bestrijkt vele vakgebieden. De grootste wetenschappelijke uitdaging voor de nabije toekomst ligt in de verdere integratie van epidemiologie, demografie, economie en biologie.

Dankwoord

Gezamenlijk

Als wij allen die ons hebben bijgestaan in de lange geschiedenis van dit project en onze eigen lange wetenschappelijke voorgeschiedenis met naam en toenaam zouden moeten vermelden, dan zou de omvang van dit dankwoord beginnen concurreren met de omvang van dit proefschrift. We willen ons daarom bij voorbaat verontschuldigen bij al diegenen die we hier ten onrechte niet vermelden. We zullen eerst onze gezamenlijke dank betuigen, met daarna van ieder van ons tweeën een meer persoonlijke noot.

Het lijkt maar eerlijk om te beginnen met diegenen die aan de wieg van dit proefschrift hebben gestaan. De wieg heette TAM, maar de ambities die in die wieg lagen waren ver van tam. TAM stond voor "Technology Assessment Methods", en had als hoog gegrepen doel 'referentie-waarden', een soort waardebepalende peilstokken, te ontwikkelen voor een grote reeks medische interventies met behulp van een "volksgezondheidsmodel", dat NIMPH gekerstend werd: Netherlands Integrated Model of Public Health. Dit was niet mogelijk geweest zonder de lange termijn visie en de financiële steun van het toenmalige ministerie van Welzijn, Volksgezondheid en Cultuur en het huidige Ministerie voor Volksgezondheid, Welzijn en Sport, met name vooral Jannes Mulder en Josee Hulshof, die TAM en zijn bedenkers steeds gesteund hebben.

Bij TAM hoorden vroeger een heleboel namen, die al dan niet van dichtbij ons reilen en zeilen volgden. Bij diegenen die ons een vliegende start gaven horen zeker Gerrit van Oortmarssen, Ben van Hout, Gouke Bonsel en Mark Koopmanschap speciaal te worden vermeld. Gerrit was de strenge wiskundige, die ons wild geraas enigszins wanhopig binnen formele grenzen trachtte te houden, en zonder Ben hadden we nooit dat belachelijk ingewikkelde maar toch ook belangrijke boek van Manton en Stallard gelezen en begrepen. Gouke's encyclopedische medische kennis en Mark's economische inzichten bleken steeds beschikbaar. Terwijl de vorigen ons verlieten, is één gebleven, altijd GLIMmend van bekwame behulpzaamheid, de aardigste biostatisticus van het westelijk halfmond (en omstreken), Caspar Looman. Welwillende peters aan de TAM wieg waren ook nog Dik Habbema, Johan Mackenbach, Frans Rutten, en Louise Gunning-Schepers; over Louise lees je verder nog meer.

Na deze TAMmers van het eerste uur zijn anderen gekomen en weer gegaan. Leona van Roijen versterkte de economische poot, en Jacqueline van den Bosch kwam helpen om de almaar groeiende berg, hoe zullen we het zeggen, beknopt gedocumenteerde computercode in bedwang te houden, maar leverde daarnaast ook onschatbare wiskundige en modelleer expertise.

Ook drie medici op stage kruisten ons pad: Louis Niessen, Babs Reichgelt en Elke Witthaus, ieder zijn of haar sporen nalatend in dit

proefschrift. Hieronder wijdt Luc nog enige warme 'dokters onder elkaar' woorden aan deze ontmoetingen.

Als jonge gezondheidswetenschappers en deels aan onze zorg toevertrouwde AIO's hadden Perla van de Mheen en Caroline Baan ongetwijfeld, en terecht, op wat meer vastigheid gehoopt. Perla en Caroline, ben je soms in wanhoop over ons (geweest), het moge een magere troost zijn dat we veel van jullie hebben geleerd. En jullie wraak is zoet: Perla passeerde ons moeiteloos op het promotietraject, en wij hebben nog haast moeten maken om te voorkomen dat Caroline hetzelfde zou doen.

Tot slot mankeert hier nog iemand in deze lange TAM-lijst. We hadden gaarne samen opgetreden, met ons drietjes op dezelfde dag, maar uiteindelijk ging het dan toch niet door. Wie het onderwerp van dit proefschrift leuk vindt, mag zeker de toekomstige promotie van Wilma Nusselder, demografe en zeer verwante ziel, niet missen.

Wij konden onze eigenzinnige muzikjes maken omdat we zo goed ingebed zijn in een soliede, harmonieus instituut, het instituut Maatschappelijke Gezondheidszorg. Hier hoort een eresaluut voor de vele aardige en goede onderzoekers die mee dit instituut maken. Een dubbel saluut is er voor het secretariaat, de financiële administratie en de computer ondersteuning, alles onder leiding van Koos Lubbe. Het is een vaststaand feit dat het secretariaat steeds behulpzaam is, op alle vragen een antwoord weet, en zich steeds uit de naad wil werken voor weer een veel te laat binnengekomen opdracht. En tot ons onnoemelijke genoegen worden alle administratieve en financiële bekommernissen ver van ons gehouden, en werken onze computers.

Speciale dank gaat uit naar Anna Bosselaar, die bereid was in vliegende haast dit proefschrift op te maken, naar Karin Mulder, die tekende voor het omslag, en naar Karen Gribbling, die hier en daar ons kreuple Engels heeft trachten te behandelen.

En tenslotte, dit hele onderzoek was natuurlijk totaal uitgesloten zonder de man die diep inzicht paart aan een brede kijk: Paul van der Maas. De samenwerking tussen Paul en ons zal voor iedere buitenstaander altijd wel een onbegrijpelijk en vooral chaotisch gebeuren blijven. Vergelijk het met die puzzels, waar bizar kronkelende stalen staafjes in elkaar moeten gepast worden. Dat lukt zelden, en als het lukt krijg je die dingen niet meer uit elkander. Zo pasten de wel erg rare kronkels van onze drie stellen hersenen perfect in elkaar, en werd dat geheel veel meer dan de som der delen. Beste Paul, het spijt ons dat je op de voorkaft van dit boek als auteur moet ontbreken, maar we zijn wel allemachtig trots je als promotor te kunnen opvoeren.

Jan

Velen hebben mij de afgelopen jaren gevraagd hoe ik, als historicus, nu in vredesnaam ben beland in het volksgezondheidsonderzoek. In de loop van de tijd is daar een standaard antwoord op ontstaan: 'probeer maar eens een onderzoeksbaan te vinden als historicus, dan begrijp je het wel'. Ondanks

deze handicap denk ik toch met veel plezier terug aan de opleiding 'maatschappijgeschiedenis': het multi-disciplinaire karakter en de sterke methodologische vorming bewijzen nog altijd hun nut. De opleiding werd gedragen door een groep enthousiaste docenten, waarvan ik de meesten onrecht aandoe door er hier slechts twee te noemen: Antoon van den Braembussche (nog zo'n Belg op mijn pad) en Peter Klein.

Met de volgende fase in mijn opleiding is de naam verbonden van Joel Mokyr, van het department of economics van Northwestern University. Op basis van een scriptie waarin zijn werk weliswaar instemmend werd aangehaald maar op sommige punten ook stevig bekritiseerd, bezorgde hij mij een 'fellowship'. Daardoor kon ik kennis maken met de hogedrukpan sfeer van een Amerikaanse Graduate School, een ervaring met een hoog 'Pietje Bell in Amerika' gehalte, maar ook een uitstekende inwijding in de 'dismal science'.

Het getuigde van enige moed dat iMGZ en in het bijzonder Louise vervolgens met mij in zee durfden te gaan: op dat moment was ik, om mezelf te citeren, tenslotte niet meer dan een gesjeesde biologiestudent, gemankeerd historicus en werkloos econoom. Beste Louise, het was in de samenwerking met jou dat ik er achter kwam wat een interessant onderzoeksterrein volksgezondheid is. Dat dit proefschrift er nu ligt, is dus zeker voor een deel aan jou toe te schrijven.

Ten slotte wil ik dank zeggen aan mijn familie, al was het maar voor de getoonde kiesheid om niet bij iedere gelegenheid te informeren wanneer het proefschrift nu eindelijk eens klaar ging zijn. Mijn dank geldt in het bijzonder mijn ouders, die door alle eigenzinnige wendingen heen mij steeds zijn blijven steunen in het volste vertrouwen dat ik uiteindelijk wel op mijn pootjes terecht zou komen. Het is dan ook aan jullie dat ik mijn deel van dit proefschrift opdraag.

Luc

Als 'medicus' hou ik de beste herinneringen over aan mijn medische 'stagiars', en co-auteurs in de ware zin van het woord: Louis Niessen, Babs Reichgelt en Elke Witthaus. We zijn arts en generatie-genooot, en dat levert toch die extra herkenbaarheid op die de vonk nog sneller doet overspringen. Verder denk ik dankbaar terug aan de talrijke en soms verhitte discussies met Karin Meeter en Huib Pols, en via Louis Niessen met Peter Koudstaal. Als goede klinici hielden ze ons tegen de harde grond van echte, niet gesimuleerde, mensen en het moeilijke klinische werk. Vooral Karin, die onze eerste stapjes bijna moederlijk begeleidde, wil ik graag een schriftelijke natte kus op de wang drukken.

Als ik verder terugblik dan TAM en iMGZ, komen weer zoveel gezichten boven, dat ik vele namen onrecht aan moet doen. Wijlen Geoffrey Rose van de London School of Hygiene and Tropical Medicine, epidemioloog, arts en goeroe, leerde me dat epidemiologie een menselijke wetenschap is. De centrale vraag die hij stelde was of mensen er uiteindelijk

beter van werden. Alleen al het idee dat hij goedkeurend zou hebben geknikt bij lezing van dit proefschrift doet mij veel plezier.

Ik ben en blijf uiteindelijk een product van de afdeling Volksgezondheid van het Antwerpse Instituut voor Tropische Geneeskunde. Professor Van Baelen (hier geen Harrie) en Professor Mercenier (laat staan Pierre) leerden me eigennuttig en kritisch denken, ook en vooral over geneeskunde en volksgezondheid. En soms mis ik Pangu, Kangombe, de familie Criel, en al die anderen waarmee ik heb mogen werken in Kasongo. De vele praktische problemen en Kongolees-Zairese frustraties vervagen door het patina van de tijd, wat overblijft zijn de mooiste jaren van mijn beroepsbestaan. En tot slot zou het wel heel onrechtvaardig zijn Peter Piot te vergeten, die me aan mijn eerste 'echte' wetenschappelijke artikel hielp.

Waar het Instituut voor Tropische Geneeskunde zowel letterlijk als figuurlijk te klein bleek, is te schaarse middelen een terecht excuus. Afwezige ondersteuning door het beleid van wetenschappelijk onderzoek leidt tot afwezige ondersteuning van het beleid door wetenschappelijke onderzoekers. De recente Belgische geschiedenis leert dat dat toch niet straffeloos gaat. Als ik uiteindelijk de kans kreeg om uit te groeien tot wetenschapper, heb ik die van de Nederlandse samenleving gekregen. Van deze dankbare Belg hoor je zelden domme grappen over de gierige Hollander.

En tot slot ik ben niet, nergens, zonder mijn thuis, mijn veilige warme nest. Er is het gezin waar ik groot werd, deel van een grote Limburgse familie, met zijn bijna Afrikaanse gevoelens voor afkomst en geschiedenis. Ik weet dat jullie trots bent op mij, ik ben even trots op jullie. Ik draag mijn deeltje van dit proefschrift op aan mam en papa, mijn voorouders, die trotser dan trots zouden geweest zijn op dit kleine schakeltje in hun ketting.

En er is mijn kleine thuis, mijn altijd roerige gezin, waar het goed is om weer tot rust te komen, afstand te nemen van het razen van de tijd, en samen met Rita te leven met onze vier schakeltjes die we voegden aan de menselijke ketting. Eeuwig student, eeuwig zoekend naar de zin van het bestaan, heb ik die tegelijkertijd al lang gevonden in je armen en in onze vier fantastische kinderen waar deze keer ik trotser dan trots op ben.

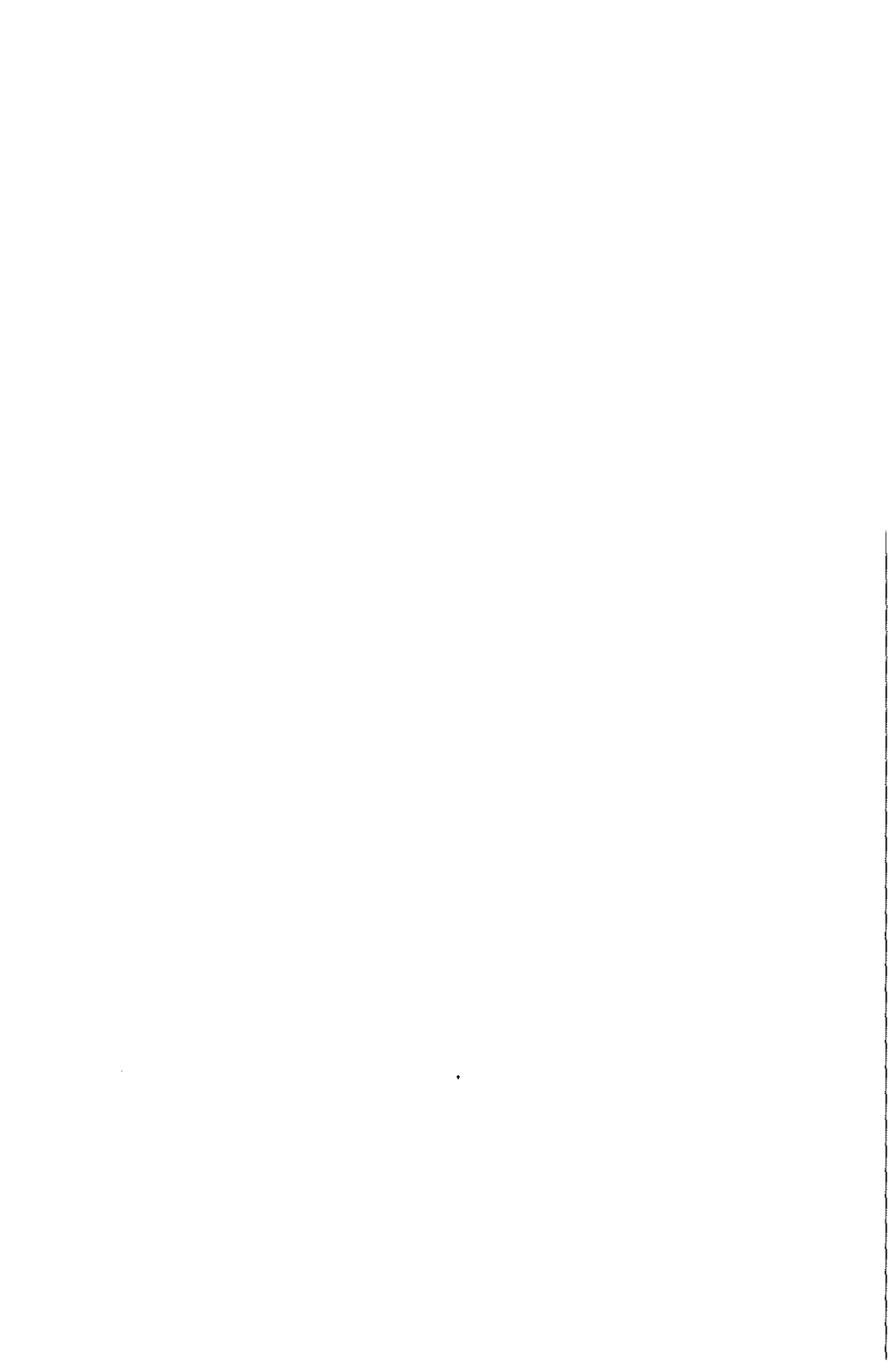
Curriculum vitae

Jan Barendregt is geboren op 5 augustus 1953. Hij deed Staatsexamen HBS-B in 1971, en ging in datzelfde jaar biologie studeren te Utrecht, maar maakte deze studie niet af. In 1978 begon hij met de studie Maatschappij-geschiedenis aan de Erasmus Universiteit, die hij in 1982 afrondde (cum laude). Daarna deed hij nog een Tweede Fase onderzoekopleiding bij hetzelfde instituut. In 1984-85 bracht hij ruim een jaar door aan de Graduate School van Northwestern University (Evanston, Illinois), waar hij een masters degree in economie behaalde. Sinds september 1986 is hij verbonden aan het instituut Maatschappelijke Gezondheidszorg van de Erasmus Universiteit. Hij is daar betrokken geweest bij een groot aantal verschillende onderzoeksprojecten: het Prevent-project, TAM (Technology Assessment Methoden), scenario-projecten voor ongevallen en voortplanting, en het Prioriteiten voor Preventie Onderzoek project. Op dit moment werkt hij aan projecten over de veranderingen in de epidemiologie van hartfalen, de relaties tussen dementie en hart/vaatziekten, en die tussen diabetes en hart/vaatziekten.

Luc Bonneux werd geboren in Werl, toenmalig West-Duitsland, op 16 november 1954 als zoon van Louis Bonneux en Rosalie Hechtermans.

Hij beëindigde de studie Latijn-Wetenschappen aan het Xaveriuscollege van Borgerhout (bij Antwerpen) in 1972. In 1979 behaalde hij de titel van dokter in de Geneeskunde aan de Universitaire Instelling Antwerpen. In 1980 behaalde hij het diploma Tropische Geneeskunde aan het Instituut voor Tropische Geneeskunde, Antwerpen. In 1980 en 1981 was hij eveneens actief als huisarts. Van 1981 tot 1986 was hij werkzaam in het Basisgezondheidsproject van de Eenheid voor Navorsing en Onderwijs in de Volksgezondheid te Kasongo in Zaire, het huidige Congo. In 1986 en 1987 was hij als arts verbonden aan de Leopold II kliniek voor Tropische Ziekten in Antwerpen en als wetenschappelijk onderzoeker aan het departement Microbiologie van het Instituut voor Tropische Geneeskunde. In 1988 behaalde hij de titel van Master of Science in Epidemiology aan de London School of Hygiene and Tropical Medicine. Sinds 1 mei 1989 is hij werkzaam aan het instituut voor Maatschappelijke Gezondheidszorg op de Erasmus Universiteit Rotterdam als wetenschappelijk onderzoeker.

Hij huwde in 1980 met Rita Saenen, en heeft 4 kinderen: Tom, Pieter, Sarah en Liesbet.



Stellingen

Stellingen behorende bij het proefschrift

*Degenerative disease in an aging population. Models and conjectures
van Jan Barendregt en Luc Bonneux.*

1. De functie van het disconteren van toekomstig geld of goederen is om de keuze tussen twee of meer wegens tijdsvoorkeur onvergelijkbare alternatieven mogelijk maken. Disconteren dient dan ook tot deze toepassing beperkt te blijven.
2. Beperking van de groei van de uitgaven voor de gezondheidszorg tot een percentage lager dan de groei van de Nederlandse economie betekent een relatieve bezuiniging. Dat is strijdig met de status van gezondheidszorg als luxe goed.
3. Incidentie-prevalentie-sterfte modellen zouden tot de standaarduitrusting van de epidemioloog moeten behoren.
4. Algemene uitspraken over de te verwachten compressie dan wel expansie van de totale morbiditeit krijgen slechts betekenis wanneer ze de grote verschillen in veranderingen van de morbiditeit van belangrijke ziekten expliciet in de verwachting betrekken.
5. Op populatieniveau wordt mortaliteitselectie een minder belangrijk fenomeen naarmate de overlevingscurve rechthoekiger is.
6. Het onderscheid tussen 'vervangende' (substitutie) en 'concurrerende' (competitie) ziekte en sterfte is niet louter academisch, maar cruciaal voor goed begrip van sommige epidemiologische trends.
7. Van de 'gezonde levensverwachting' zijn zoveel verschillende varianten, dat een schatting zonder nadere toelichting in feite geen informatie verschaft.
8. DALYs zijn net QALYs, maar dan anders.
9. De stelling van sigarettenfabrikanten dat hun reclame-inspanningen uitsluitend zijn gericht op bestaande rokers van andere merken is ongeloofwaardig en bovendien in tegenspraak met de constatering dat een groot deel expliciet is gericht op veelal nog niet verslaafde jongeren.
10. It ain't necessarily so. (I. Gershwin)
11. Een opleiding tot historicus vormt in principe geen beletsel voor een loopbaan als onderzoeker op het terrein van de volksgezondheid.
12. 'Sterven in schoonheid' is het vermoedelijke lot van de perfectionist in een 'publish or perish' omgeving.
13. Het begrip 'autoalarm' is bijzonder goed getroffen, gegeven de grote neiging van deze installaties om zonder enig merkbare aanleiding af te gaan.

*Rotterdam, 14 januari 1998
Jan Barendregt*

Stellingen

Stellingen behorende bij het proefschrift

Degenerative disease in an aging population. Models and conjectures
van Jan Barendregt en Luc Bonneux.

1. De morbiditeit in een moderne samenleving is de paradoxale optelsom van niet voorkomen ziekte en wel voorkomen sterfte. (*dit proefschrift*)
2. De dalende sterfte aan coronaire hartziekte wordt 'bekocht' met toenemende aantallen patiënten met hartfalen en beroerte. (*dit proefschrift*)
3. Preventie van heupfracturen door interventies die het snelle postmenopausale botverlies bij vrouwen moeten tegengaan is gedoemd tot lage doelmatigheid, omdat heupfracturen gemiddeld pas decennia later optreden. (*dit proefschrift*)
4. Verouderingsgebonden morbiditeit en mortaliteit zijn fenomenen van hetzelfde proces. De klassieke 'expansie' of 'compressie' theorieën, die beiden als onafhankelijk van elkaar behandelen, zijn daardoor onhoudbaar. (*dit proefschrift*)
5. Half de jaren 80 toonden eenvoudige modellen aan dat een AIDS epidemie in de geïndustrialiseerde wereld buiten groepen met hoge aantallen seksuele (of spuitenuilende) partners onwaarschijnlijk was. Het verwaarlozen van deze kennis heeft wereldwijd geleid tot verspilling van middelen en overbodige angst-neurosen. (Bonneux & Houweling, Ned Tijdschr Geneeskd 1989;133:1922)
6. *Non sunt multiplicanda entia praeter necessitatem* (William of Ockham, 1285-1347/49; vrij vertaald: vermijd alle overbodige complexiteit). Eenvoudige modellen doen ons de realiteit begrijpen, ingewikkelde modellen geven de realiteit weer zoals de onderzoeker vindt dat ze moet zijn.
7. Een eenvoudig demografisch model op basis van gegevens van Rwanda toont dat interventies die de kindersterfte verlagen weinig invloed op de bevolkingsexplosie hebben. Er is bijgevolg geen contradictie tussen volksgezondheid en duurzame ontwikkeling. (Bonneux, Lancet 1994;344:1689)
8. Egoïstische genen geven onvermijdelijk aanleiding tot altruïstische samenlevingen, omdat samenwerking een evolutionair superieure strategie is (uit Axelrod: *The evolution of cooperation* en Matt Ridley: *The origins of virtue*).
9. De wetgeving op de vrijheid van pers is gebaseerd op het recht op vrije meningsuiting en het bestraffen van aangetoond misbruik. De wetgeving op de bescherming van de persoonlijke levenssfeer ('*privacy*') zou zich moeten baseren op het recht van vrije informatiegaring en het bestraffen van aangetoond misbruik.
10. Het afschermen van wetenschappelijke gegevens voor onafhankelijke onderzoekers leidt tot overbodig dupliceren van onderzoek en is een invitatie tot fraude. Na een redelijke termijn dient het intellectuele eigendom van de individuele onderzoeker te vervallen en moeten gegevens publiek bezit worden.

11. Om vooruitgang te boeken in onze kennis over behandeling en eventuele preventie van chronische psychiatrische aandoeningen is deskundig (klinisch) epidemiologisch onderzoek een noodzakelijke vereiste.
12. De juiste vraag is niet 'of' een interventie kosten-effectief is, maar 'bij wie'. De effectiviteit van een interventie wordt immers bepaald door de prognose van de patiënt. Wie de doelmatigheid van de gezondheidszorg wil verbeteren, zal dus indicatiestellingen moeten kunnen sturen.
13. De grote belangstelling voor voedingsonderzoek in Nederland staat in schril contrast met de geringe belangstelling voor de kwaliteit van het voedsel geserveerd in een universiteitsrestaurant.

Rotterdam, 14 januari 1998
Luc Bonneux