Complementary analyses in economic evaluation of health care

Colophon

Complementary analyses in economic evaluation of health care
Marc Koopmanschap
Key words: economic evaluation, cost of illness, indirect costs

Photography: V. Nykl, AVC, Erasmus University Rotterdam
Printer: Pasmans offsetdrukkerij B.V., Den Haag
Layout: Angélique Visser

Complementary analyses in economic evaluation of health care

Complementaire analyses in economische
evaluatie van de gezondheidszorg

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. Lit.
en volgens besluit van het College van Dekanen.

De openbare verdediging zal plaatsvinden op
woensdag 15 juni 1994 om 13.45 uur

door

Marcus Anthonius Koopmanschap

geboren te Amsterdam

Promotiecommissie

Promotores	:	Prof.dr. F.F.H. Rutten
		Prof.dr. P.J. van der Maas
Overige leden	:	Prof.dr. L.J. Gunning-Schepers
		Prof.dr. R. Leidl

'Doubt is the beginning of wisdom, and complacency the major obstacle to change, so it is essential to keep exposing the inadequacies of the present situation'

(Alan Williams. Priorities and research strategy in health economics for the 1990s. Health Economics 1993;2:295-302.)

Contents

1	**Introduction**	1
	Economic evaluation	1
	Level of aggregation	1
	Costs to be included	4
	Key questions	5
	Reading guidance	7

Part I Detailed analysis

2	**Economic aspects of cervical cancer screening**	11
	Summary	11
	Introduction	11
	Material and methods	12
	Results	17
	Sensitivity analysis	22
	Discussion	23
	Conclusions	24
3	**Cervical cancer screening: attendance and cost-effectiveness**	27
	Summary	27
	Introduction	27
	Methods	28
	Results	29
	Discussion	40
	Conclusions	42
	Supplementary comment	44

Part II Aggregate analysis

4	**Cost of disease in international perspective**	47
	Summary	47
	Introduction	47
	Methods	48
	Results	53
	Discussion	62

5	**Current and future costs of cancer**	65
	Summary	65
	Introduction	65
	Material and methods	66
	Results	70
	Discussion	77

Part III Indirect costs of disease

6	**Indirect costs in economic studies: confronting the confusion**	81
	Summary	81
	Introduction	81
	A literature review	83
	Some controversies	84
	Why and how to include indirect costs	85
	Indirect costs and health policy	90
	The agenda for future research	90
7	**Towards a new approach for estimating indirect costs of disease**	93
	Summary	93
	Introduction	94
	The concept of indirect costs	94
	Operationalisation	97
	Results	101
	Discussion	102
	Conclusion	103
8	**The friction cost method for measuring indirect costs of disease**	105
	Summary	105
	Introduction	105
	The friction cost method	106
	Operationalisation of concepts	110
	Data and methods for the Netherlands	113
	Results	116
	Discussion	121

9	**The impact of indirect costs on outcomes of health care programs**	125
	Summary	125
	Introduction	125
	Methods	126
	Health care programs	128
	Results	131
	Analysis of impact of indirect costs	134
	Discussion	136
10	**Conclusions**	139
	Detailed analysis	139
	Aggregate analysis	140
	Indirect costs	141
Literature		145
Summary		157
Samenvatting		161
Appendices		165
Acknowledgements		
Curriculum vitae		

Publications

Chapters 2-9 are based on the following articles:

Ch. 2 Economic aspects of cervical cancer screening. Koopmanschap MA, Lubbe, JThN, Oortmarssen van GJ, Agt van HME, Ballegooijen van M, Habbema JDF. Soc Sci Med 1990; 30: 1081-1087.[1]

Ch. 3 Cervical cancer screening: attendance and cost-effectiveness. Koopmanschap MA, Oortmarssen van GJ, Agt van HME, Ballegooijen van M, Habbema JDF, Lubbe JThN. Int J Cancer 1990; 45: 410-415.[2]

Ch. 4 Cost of diseases in international perspective. Koopmanschap MA, Roijen L van, Bonneux L, Bonsel GJ, Rutten FFH, Maas PJ van der. EJPH (in press).[3]

Ch. 5 Current and future costs of cancer. Koopmanschap MA, Roijen L van, Bonneux L, Barendregt JJ. Eur J Cancer 1994; 30A(1): 60-65.[1]

Ch. 6 Indirect costs in economic studies: Confronting the confusion. Koopmanschap MA and Rutten FFH. PharmacoEcon 1993; 4(6): 446-454.[4]

Ch. 7 Towards a new approach for estimating indirect costs of disease. Koopmanschap MA and van Ineveld BM. Soc Sci Med 1992; 34: 1005-1010.[1]

Ch. 8 The friction cost method for measuring indirect costs of disease. Koopmanschap MA, Rutten FFH, Ineveld BM van and Roijen L van. J Health Economics (in press).[5]

Ch. 9 The impact of indirect costs on outcomes of health care programs. Koopmanschap MA and Rutten FFH (submitted).

1 Introduction

Economic evaluation

The steady increase in health care costs and the continuous emergence of new medical technologies have forced policy makers in health care to reconsider the current resource allocation and to become more selective with investing in new health care programs. Economic evaluations can support the decision making process, by providing systematic information on the costs and the consequences for health of investing in alternative health care programs.

Needless to say, economic evaluations of health care should be methodologically sound, the outcomes should be relevant for health policy and comparable to results of studies concerning other health care programs. With respect to the policy relevance it is important that the aggregation level of the analysis matches the specificity of the policy question: a study of costs and health effects of for example air pollution control will not need to be as detailed as an analysis of the cost-effectiveness of cimetidine versus surgery in peptic ulcer.

Economic evaluations may never become entirely comparable, but incomparabilities due to different methodologies can be reduced considerably. This raises two questions:

- what is the appropriate level of aggregation in the economic evaluation of health care?
- which cost items are to be included in economic evaluations and how should these be measured and valued, in particular the indirect costs of disease?

Level of aggregation

In economic evaluation, health care is usually analyzed on one of the following levels of aggregation:

- a specific health care intervention or program: 'the detailed level';
- total health care, a specific sector of health care or a disease category: 'the

aggregate level'.

Either level of analysis has advantages and limitations. The detailed analysis can provide a thorough and complete overview of the relevant aspects of a specific program. It is the appropriate level of analysis, if the policy decision is to choose among specific programs for a specific patient group. However, focusing on cost-effectiveness of individual programs bears the risk of overlooking interactions with other health care programs (Birch and Gafni, 1992; Drummond et al., 1993). A specific program may have a favourable balance of costs and health effects, if added to an already existing program which itself is not cost-effective. In case of end stage renal disease, the introduction of kidney transplantation involves cost savings as well as gains in length and quality of life. Candidates for kidney transplantation can only stay alive due to expensive facilities for haemodialysis (Rutten and van Hout, 1991). The decision to adopt the transplantation program should be based on an economic appraisal of the combination of the two medical technologies as compared to the situation without both technologies.

A second type of interaction is caused by common risk factors for different diseases. If the treatment of lung cancer would improve, thereby reducing lung cancer mortality, more smokers would stay alive, causing a more than proportional increase in prevalence and costs of cardiovascular disease and chronic lung disease. Another example of interaction is hormone suppletion for perimenopausal women in order to prevent osteoporosis, which may increase the prevalence of breast cancer, which in turn may improve the cost-effectiveness of breast cancer screening (Tosteson et al., 1990). Hence, the neglect of possible interactions between health care programs may invoke non-optimal resource allocation.

Furthermore, health care programs may influence the distribution of health (care). These equity consequences should preferably be considered in the context of all interventions for a specific disease or patient group, instead of being analyzed separately. The detailed type of analysis has a second, more practical limitation. It is simply not feasible to conduct economic evaluations of all existing and new health care programs within a reasonable time-span and research budget. As a consequence, existing health care programs are seldomly analyzed, so we lack information on costs and health effects of a substantial part of health care.

It follows that somehow economic evaluation on a more aggregate level is needed to provide a consistent and complete overview of costs and health effects in health care. Such overview could serve as a starting point for more detailed analysis in specific parts of health care that need closer investigation of costs and health effects. This aggregate approach also has some drawbacks. An estimate of costs and health effects of all interventions related to a certain disease always represents the combination of cost-effective and cost-ineffective programs. Withdrawing resources from specific subsectors or broad disease categories, based on unfavourable aggregate cost-effectiveness, may result in abolishing or reducing specific cost-effective programs. In addition, the estimation of the effectiveness of a bundle of interventions directed to a disease is a formidable task since it requires knowledge about the health status of patients in the situation *without* the existing health care programs.

The arguments above suggest that economic evaluation should be performed on both levels of analysis, as either provides relevant information for health policy. Economic evaluation on the detailed level is common - practice in the form of cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis. Thus far, aggregate economic evaluation is (usually) restricted to estimates of the current and future cost and morbidity and mortality of diseases, as we lack evidence on the total effectiveness of the programs involved. Shiell et al. argued that the relevance of such "cost of illness (or burden of illness) estimates" for allocation decisions is rather limited, because the costs of a disease may only represent the 'benefits of the unattainable': the savings as a result of eradicating that disease (Shiell et al., 1987). Furthermore, these cost of illness estimates do not take into account the costs of possible remaining comorbidity or the costs of competing diseases that will emerge after eradication of the specific disease studied.

However, in the absence of reliable evidence on aggregate effectiveness, information on costs and epidemiology can be incorporated in a disease model, in order to estimate future costs, morbidity and mortality of diseases. The disease model should contain the main elements of costs by disease phase or -stage. The model-based predictions can indicate the possible consequences of expected demographic, epidemiologic or technological scenarios. This type of scenario analysis is a form of partial economic evaluation that will not allow answering efficiency questions (Drummond et al., 1987), but it may still serve several purposes. First, it may be relevant for planning the capacity of specific health care services. Second, it provides an estimate of the growth rate of health care costs if health policy remains unchanged, indicating the need for cost reduction or the room for extending health care. Third, modelling several diseases simultaneously, as described by Bonneux and Barendregt (1992), facilitates a correct estimation of the future medical consumption of patients whose life has been extended (or its quality improved), due to an intervention, taking into account the possible higher risk for other diseases.

In case of comorbidity, things become even more complicated, as cost of illness estimates generally (this thesis is no exception) assign all costs to the primary diagnosis, whereas part of these costs may be related to comorbidity. Further analysis of the magnitude of comorbidity costs, especially for elderly patients, is urgently needed. Detailed comparison of the costs for patient-groups (age- and sex-specific) with the same primary diagnosis, with and without comorbidity, could shed more light on this matter.

If estimates of effectiveness of bundles of interventions will become available, the disease model can also produce results on aggregate cost-effectiveness. If the latter are unfavourable, it seems appropriate to conduct a critical analysis of the costs and health effects of the constituent interventions.

As a matter of fact, application of cost-effectiveness as only criterium for evaluation of health care is not to be recommended, because equity and ethical considerations should also be taken into account. It is to be expected that for example care for mentally retarded, parts of nursing home care and palliative care for advanced cancer would not pass the pure cost-effectiveness test. However, a civilized society would not withhold its citizens such basic care, irrespective of its cost-effectiveness.

Costs to be included

The literature on economic evaluation distinguishes 4 types of costs with respect to health care programs:

- direct medical costs (usually referred to as health care costs);
- direct non-medical costs (e.g. patient's travel costs);
- indirect medical costs (the costs of medical consumption during life-years saved as a result of an intervention);
- indirect non-medical costs (the value of production loss and/or the extra costs to continue production, due to illness or its treatment).

Health economists generally agree that the first and second type of costs should be included in economic evaluation. The issue of indirect medical costs will not be considered in this thesis, but authors tend to agree that inclusion of indirect medical costs depends on the specific research question. They may be not relevant for the costs and effects of a specific intervention (Drummond et al., 1987), but they should definitely be included in scenario-analysis, as they may represent an important part of the future costs of disease.

At the moment, the relevance of the indirect non-medical costs (further referred to as indirect costs) is subject to considerable debate. The discussion relates to the issue whether economic consequences of disease beyond the health care sector should have implications for health policy. Another important topic is the possible influence of including indirect costs on equity with respect to employed and not employed citizens.

If indirect costs are included in economic evaluation, the next question relates to the appropriate methodology to estimate these costs. The basic idea as proposed in this thesis is that indirect costs should represent the expected level of economic consequences in the real world, instead of estimating the potential consequences in a theoretical state of full employment equilibrium, according to the human capital approach. In this respect, three issues are crucial: the short term impact of absence from work on production and costs, the time-span during which production and costs are affected directly and the medium term macro-economic consequences of absence from work and disability. These issues will be discussed extensively in this thesis.

Key questions

Summarizing the previous discussion, the following questions are to be answered in this thesis:

- What can be concluded for health policy from a detailed economic evaluation of an individual health care program, as illustrated by a cost-effectiveness analysis of cervical cancer screening?
- What is the role of more aggregate economic evaluation of health care? To what extent may cost estimates, incorporated in disease models, permit conclusions on future costs and cost-effectiveness of health care?
- What is the relevance of indirect non-medical costs of disease in economic evaluation?
- If the indirect costs are important for health policy, how can reliable and realistic estimates of indirect costs be derived and what is their impact on the results of economic evaluations of specific health care programs?

Not all issues raised in this introduction are studied in this thesis. For example, the problem of taking into account the costs of comorbidity and estimates of cost-effectiveness of bundles of interventions will remain unsolved here. Table 1.1 gives an overview of the subjects analyzed in each part of this thesis: the detailed analysis (part I), the aggregate analysis (part II) and the indirect costs of disease (part III).

Table 1.1 The consideration of issues in economic evaluation in this thesis.

Issue	Detailed analysis (cervical cancer screening)	Aggregate analysis	Indirect non-medical costs of disease
Total current costs	yes	yes	yes
Total costs per patient	yes	Cancer: yes Other dis.: no	some diseases: yes
Future direct medical costs, demographical scenario	yes	yes	n.a.
Future direct medical costs, epidemiological scenario	yes	Colorectal ca: yes Other dis.: no	n.a.
Direct non-medical costs	yes	no	n.a.
Costs of comorbidity	no	no	no
Indirect costs of unpaid work	n.a.	n.a.	no
Efficiency of health care program	yes	no	no
(Cost)-effectiveness of health care program	yes	no	yes (theoretical intervention)
Competing mortality	yes	Cancer: yes Other dis.: no	no
Competing morbidity	no	no	no

n.a.: not applicable

Reading guidance

In this thesis it will be suggested that both the detailed and the more aggregate level of analysis are useful in conducting economic evaluations of health care programs.

In part I a detailed cost-effectiveness analysis of cervical cancer screening is performed. Chapter 2 focuses on the costs of screening and its impact on the costs of diagnosis and treatment, the selection of efficient screening policies and the diminishing returns of extending the screening program, whereas chapter 3 shows the role of the screening attendance rate for the cost-effectiveness of cervical cancer screening.

Part II describes the aggregate point of view in economic evaluation. In chapter 4 total health care costs for the Netherlands in 1988 are estimated by disease category, age-group and sex and compared to results for Sweden and the United States of America. The estimated costs for 1988 are used to predict the cost consequences of ageing. For six types of cancer the aforementioned cost estimates are refined and assigned to disease phases in the cancer disease model, as described in chapter 5. The future costs of cancer are predicted for a demographical scenario and, in case of colorectal cancer, also for several epidemiological scenarios, involving expected trends in incidence and survival.

The indirect costs of disease are discussed in part III. Chapter 6 reviews the empirical literature on indirect costs of disease and discusses the main methodological issues at stake. The exact concept of indirect costs as used in this thesis is defined. It is argued that indirect costs should be included in economic evaluation, provided that the real amount and value of the production lost and/or extra costs to continue production is to be estimated. A new concept for estimating the indirect costs of disease, the friction cost method, is introduced in chapter 7. This method is developed further in chapter 8, paying attention to the relation between short term absence from work and labour productivity, the situation on the labour market and the macro-economic consequences of absence from work and disability.

Chapter 9 describes the impact of including indirect costs, according to the friction cost method, on the results of several health care programs, including cervical cancer screening. The variation in impact of indirect costs across programs will be explained in terms of the type of health care program, the disease involved and the population aimed at. The conclusions are presented in chapter 10.

Part I Detailed Analysis

2 Economic aspects of cervical cancer screening[1]

Summary

The results of a cost-effectiveness analysis of cervical cancer screening in the Netherlands are reported, emphasizing the analysis of the costs of screening and consequent diagnosis and treatment. Many organized screening policies are evaluated, differing in age-range and interval between screens. The cost estimates are based on organization charts, file studies and tariffs. The costs of screening itself are by far the most important cost component. Screening increases the costs of diagnosis. Costs for primary treatment only rise for large screening policies. Screening causes savings in costs of terminal treatment, but these are small compared with the costs of screening. The costs per life-year gained for the most efficient policies amount to Dfl 24,000 for the policy with 7 invitations per woman in a lifetime and rise considerably in case of more than 10 invitations. Cervical cancer screening appears to be less cost-effective than breast cancer screening, but compared with other services the results are comparatively good. Implementing one of the efficient organized screening policies and discouraging spontaneous screening beyond that schedule leads to considerable savings. Moreover, many organized policies which are not efficient are still superior to spontaneous screening.

Introduction

In the Netherlands 300 women per year die from cervical cancer, i.e. 0.5% of total female mortality. Symptoms of cervical cancer only occur when the cancer is already invasive. The pre-invasive stages of cervical cancer last on average about 15 years and can be detected by taking a Pap-smear. Almost 100% of the women with early detected pre-invasive stages can be cured.

---

[1] Chapter based on Koopmanschap et al., Soc Sci Med, 1990

A considerable part of the pre-invasive stages will not become invasive but will disappear (regress) spontaneously, especially in younger ages. Neither the smear nor further diagnosis can discern regressive from progressive pre-invasive stages, so any screening policy inevitably causes diagnosis and treatment and costs for women who would never have developed invasive cervical cancer.

Taking Pap-smears has proved to be effective from the medical point of view: it can reduce cervical cancer mortality (Day, 1984). Discussion remains about the best screening policy: At what ages should women be screened? What is the appropriate interval between successive screens? Should women be invited? Who must take the smears, specially trained women or general practitioners and gynaecologists? To answer these questions it is necessary to estimate both health effects and costs of all possible courses of action. Moreover, costs and health effects of screening for cervical cancer should be evaluated in relation with other health services, given limited resources.

The study on which this article has been based aims at determining which policies of cervical cancer screening in the Netherlands generate the best results in terms of health effects and costs. (Habbema et al., 1988). To this end a prospective cost-effectiveness analysis has been carried out in which epidemiological, medical and economic aspects were studied. This article describes the economic part and focuses on the following items:

- the costs of screening itself for various policies;
- the influence of screening on costs of diagnosis and treatment;
- the cost-effectiveness of the most favourable policies, as compared with other health care services.

Material and methods

In general

During the last decade in The Netherlands, as in many other countries, early detection of cervical cancer was achieved in two ways. A screening program was in operation for all women aged 35-54, who were invited at 3-year intervals. At the same time gynaecologists and general practitioners have taken many smears, especially from younger women, without fixed intervals. As a result many women were screened very often. The efficiency of this combination of organized and spontaneous screening has been questioned.

In order to make recommendations about age-groups and intervals it is necessary to estimate costs and health effects of strategies in which only "organized" screening is performed. This means that women are invited regularly. The screening is performed exclusively by trained female personnel. No further smears are taken. To estimate the impact of "organization", costs and effects were also calculated for cases in which only spontaneous screening was performed. Early detection was carried out by general practitioners and gynaecologists only, according to the distribution and frequency of smears as observed in The Netherlands during 1985-1987.

With respect to organized screening, cost and health effects were calculated for about 100 policies. The number of invitations per woman varied between 3 and 25, the interval between two invitations ranged from 15 to 2 years. The age-group invited varied, between the ages of 20 and 75. The results were compared with the estimated costs and effects for the (hypothetical) situation where no early detection takes place: the zero-option.

A model for cervical cancer and the impact of screening on morbidity and mortality was developed. The micro-simulation program MISCAN was used to calculate screening results and the effects of screening (Habbema et al., 1984 and 1987). First, the model was fitted to data from the three Dutch pilot regions where organized screening took place during 1976-1984. Model assumptions about sensitivity of the Pap-smear and regression and duration of the pre-invasive disease stages were derived which offered a satisfactory fit with data concerning screen detected pre-invasive and invasive cases, interval cancers and mortality.

Next, the model was used to predict the health effects and costs of the policies. Data for the assumptions concerning organization and costs were collected by numerous interviews and analysis of cost accounts. They were integrated into the model by combining the predicted numbers of screens and the impact on diagnosis and treatment with the relevant cost-estimates.

The most important health effect of cervical cancer screening is the reduction in mortality, expressed in the number of life-years gained compared with the zero-option. From the total set of policies those with a maximum of life-years gained for a given level of costs were identified as efficient policies. Figure 2.1 shows the incremental costs and the life-years gained for all simulated policies. The marked points indicate efficient policies. The upper-left quadrant of these points is empty with neither policies producing more life-years gained for the same costs nor policies saving more life-years less expensively.

In the rest of this chapter results will only be presented for a limited number of efficient organized screening policies and one variant of spontaneous screening.

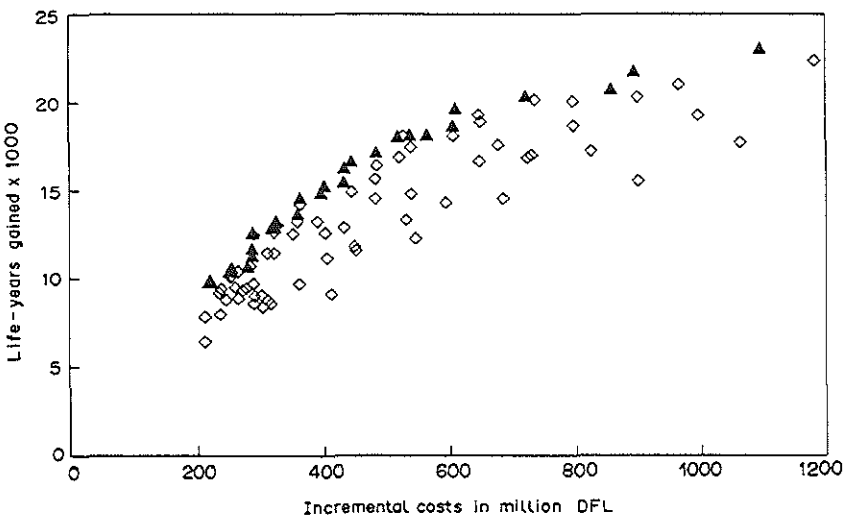

Figure 2.1 Incremental social costs and life-years gained for about 100 different organized cervical cancer screening policies. Screening during 1988-2015. Discount rate 5%. The marked points indicate the subset of efficient policies.

Cost calculations

All possible costs and savings that screening policies bring about were identified and measured, taking the concept of social costs as point of departure. This included non-financial costs and costs which do not fall upon the medical sector, but excluded value added tax. The costs of screening itself were derived by estimating the quantities and costs of each input of the screening program. This resulted in cost functions reflecting the influence of fixed and variable costs and economies of scale.

The costs of diagnosis and treatment have been approximated by the tariffs charged (1987) in The Netherlands. An analysis of the true resource costs was beyond the scope of this study. For each policy MISCAN calculated for each year the number of women invited, the number of women screened and the number of women needing diagnosis and/or therapy in various degrees. These outcomes combined with the cost functions and tariffs produced the annual costs.

Only after a considerable period the health effects of screening on cervical cancer become visible while the costs of screening are important from the start. To permit adequate evaluation, the screening program was assumed to be in operation during 1988-2015. After the year 2015, the terminated program will still cause health effects and changes in costs of diagnosis and treatment; these were calculated until 2088. Costs and effects were discounted to present values for 1988 with a discount rate of 5%, stipulated by the Dutch government (Klaassen, 1985) and several health economists (Russell, 1987). The estimated annual financial costs for government and insurers, including value added tax, will be presented too.

A fraction of the women screened has a suspicious smear, leading to diagnosis and, if necessary, treatment. Therefore three categories of costs are relevant: costs for screening, diagnosis and treatment.

Costs of screening

The organization of the cervical mass screening in the three Dutch pilot regions and expert opinions about future developments were used to draw the outline of an efficient nationwide organization. This outline is only a tool to assess all relevant costs. The activities are carried out at three levels: local, regional and national.

Local level

Each municipality sent a first invitation to the women in turn, by means of the up to date population registry. The women were screened in their municipality by a team of three trained smear-takers. The yearly capacity per team ranged from 19,000 (mobile unit) to 20,000 (static centre) smears. For a static centre the material costs are relatively low; use of a mobile unit increases the cost of taking smears by at least 25%. The number of static and mobile units depends on the population density and the number of women screened. The time and travel costs of the women attending were relevant social costs (see table 2.2). The average time for travelling and screening was estimated at 50 min and was valued by 25% of the average net labour income per labour hour per household, according to binding advices (COBA, 1975). Travel costs were estimated on the basis of public transport tariffs.

Regional level

With respect to cancer prevention The Netherlands was divided into 9 regions, each with on average about 800,000 female inhabitants. One coordination centre per region managed the screening teams, took care of local publicity and informed municipalities about the invitations. The centre sent test results to all women participating and re-invited non-participants. In cooperation with the cytological laboratory, quality control of the smears and training of the smear-takers was carried out. The centre was staffed by a manager, a medical consultant and administrative personnel. The workload and the costs varied according to the size of the program.

The cytological laboratory evaluated the smears and registered the test results in a local data-base connected to a national data-base. On average, full time analysts evaluated approx. 7,200 smears annually. Per 50,000 smears one senior analyst, one pathologist and two administrative employees are needed. Economies of scale existed for the costs of supervising and administrative personnel and part of the material costs.

National level

One centre took care of the national coordination and public relations. Using regional reports the policy was monitored and adapted if necessary. These costs were not related to the program size. The costs of registration consisted of costs for the national database, costs of training the laboratory personnel and costs of local computer facilities. The national data-base also registered results of other pathological examinations, so these costs are only partially included.

Spontaneous screening

This type of screening incurred no costs of coordination and invitation. To estimate the costs of smear-taking, we assumed that all smears were taken by general practitioners and calculated the labour costs of the general practitioner, the assistant and the material costs. Although this resulted in underestimating the costs since gynaecologists also took many smears which are more expensive, on the other hand, during a visit other medical activities may be performed as well, so only part of the costs should be attributed to the smear.

Costs of diagnosis and treatment

There are two situations where diagnosis and treatment take place: after an abnormal preventative smear or in case of possible symptoms of cervical cancer. In the zero-option, without any early detection, costs of diagnosis and treatment are only important for women with symptoms. In our study a detailed scheme was constructed to determine which diagnostic and therapeutic procedures took place for woman having a suspicious smear. This scheme is subdivided by stage of disease and enabled us to calculate changes in number and types of procedures and costs when the stage-distribution of diagnosed (pre)-cancers alters due to screening. For more details see Ballegooijen (1988).

Table 2.1 Financial costs of screening per year for organized screening policies with 250,000 and 1,000,000 smears per year and spontaneous screening, in millions Dfl. (including 4% repeated smears, attendance rate: 65%)

	Organized policies				Spontaneous	
Smears per year (x 1,000)	250	(%)	1,000	(%)	910	(%)
Coordination	3.4	(24)	4.3	(10)	-	
Invitations and results	1.1	(8)	4.4	(10)	1.2	(3)
Smear taking	3.5	(24)	13.3	(31)	14.3	(40)
Cytological evaluation	5.6	(39)	19.2	(45)	18.4	(52)
Registration	0.8	(6)	1.5	(4)	1.7	(5)
Total costs of screening	14.4	(100)	42.7	(100)	35.6	(100)
Per smear (in guilders)	58.0		43.0		39.0	

Results

Costs of screening

In table 2.1 the annual financial costs of screening are summarized for two different sizes of policies. Table 2.2 shows the present value of the social costs of policies, operational during 1988-2015, taking into account the demographic scenario. The results in tables 2.1 and 2.2 include the costs of 4% repeat smears taken because of insufficient quality of the first smear. The average attendance rate assumed is 65%, as observed in the pilot regions. The total financial costs per smear range from Dfl. 58 for 250,000 annual smears to Dfl. 43 for 1,000,000 smears per year. The economies of scale are considerable, chiefly because the costs for coordination and registration are almost fixed. The costs of coordination, registration and sending of invitations and test results form a substantial part of the total screening costs (15-30%), whatever the size of the policy. The "medical activities" smear-taking and cytological evaluation make up 60-75% of the total costs. Time and travel costs account for 10 % of the social costs of the screening (see table 2.2).

The last column in tables 2.1 and 2.2 indicates the costs of spontaneous screening on a scale of approx. 910,000 smears per year resembling the situation in the Netherlands during 1985-1987. The costs per smear are relatively low, the total costs of the spontaneous screening are considerable, due to the large number of smears.

Table 2.2 Present value for 1988 of social costs of screening for three efficient policies and spontaneous screening during 1988-2015 in millions Dfl. Discount rate: 5% (Including 4% repeat smears, attendance rate: 65%). Invitation schedule: number of invitations/age-range/interval.

	Efficient policies			Spontaneous screening
Invitation schedule	7/37-73/6	10/27-72/5	16/26-74/3	-
Average number of smears (x 1000)	390	610	975	910
Coordination	63	67	75	-
Invitation and results	23	35	56	18
Smear taking	77	118	187	204
Cytological evaluation	120	179	280	270
Registration	13	16	20	23
Time and travel costs	29	45	72	63
Total costs of screening	325	461	691	578

Costs of diagnosis and treatment
The costs of diagnosis and treatment were computed by combining the relevant number of women with the schemes for diagnosis and treatment and the tariffs charged. Figure 2.2 depicts the extra costs and savings for diagnosis and treatment induced by two different organized screening policies, compared with the zero-option. The data are not discounted (table 2.3 will show discounted costs).

Positive smears induce costs for diagnosis (DS). The more screening, the more positive test results, the higher the costs of diagnosis. Due to screening the number of women with symptoms diminishes, causing a small saving (DC). Screening increases the number of women with diagnosed pre-cancers, which incurs costs for primary treatment (PTS). These costs consist of two parts: costs for women detected by screening, who would develop invasive cancer without screening and costs for women with regressive lesions which never would become invasive. The prevention of invasive cancers permits substantial savings on primary treatment (PTC). With more intensive screening savings (PTC) rise, but the increase of the extra costs (PTS) is even sharper, since the proportion of regressive lesions increases. The reduction of invasive cancers causes a decrease in mortality. The savings on terminal treatment (TT) increase gradually with more intensive screening.

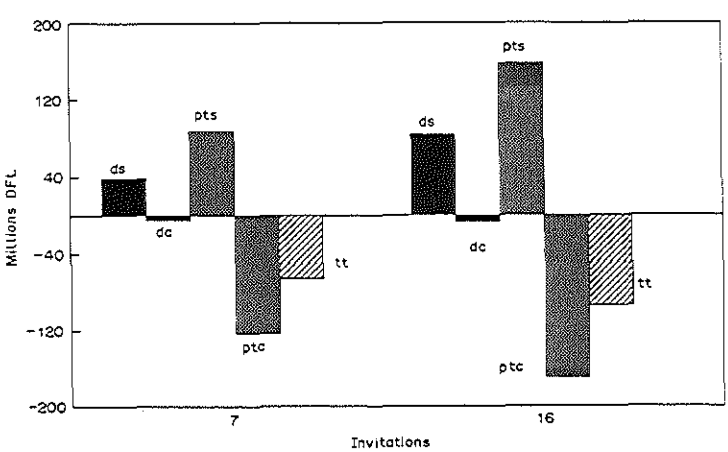

Figure 2.2 Extra social costs and savings for diagnosis and treatment compared with no early detection, for two efficient policies, in millions Dfl. No discounting.
ds = extra costs of diagnosis induced by screening; dc = savings in costs of diagnosis; pts = extra costs of primary treatment induced by screening
ptc = savings in costs of primary treatment; tt = savings in terminal treatment

The savings induced by screening lag behind the extra costs. Discounting gives the results shown in table 2.3. The discounted costs of diagnosis increase, particularly with more intensive screening. Screening with 7 invitations leaves the costs of primary treatment on balance unaltered. Intensifying screening means that the extra costs of primary treatment outweigh the savings at an increasing rate. Terminal treatment still produces savings, although discounting diminishes the absolute level. Spontaneous screening generates a high level of costs for primary treatment but this does not produce large savings on terminal treatment because the mortality reduction is relatively small.

Table 2.3 Present value for 1988 of social costs of screening, diagnosis, treatment and incremental costs for three efficient policies, spontaneous screening and no early detection in millions Dfl. Period 1988-2088. Discount rate: 5%.

	Efficient policies				
Invitations Costs	7	10	16	Spontaneous screening	No early detection
Screening	325	461	691	578	0
Diagnosis	27	39	54	47	7
Primary treatment	165	182	192	200	163
Terminal treatment	78	73	68	84	101
Total costs	595	754	1,005	909	271
Total incremental costs	324	483	735	638	-

Total costs

The costs of screening are by far the most important part of the total costs, especially for more intensive screening programmes. Costs of diagnosis and treatment are relatively small. Savings are to be expected on terminal treatment, but they are rather modest. The total incremental costs compared with the zero-option range from Dfl 324 million in case of 7 invitations per woman to Dfl 735 million if woman are invited 16 times. The incremental costs of spontaneous screening exceeds Dfl 600 million.

Cost-effectiveness

The cost-effectiveness of the policies is expressed as the incremental social costs per life-year gained. Table 2.4 shows the results for three efficient organized screening policies and one level of spontaneous screening. The additional health effects of intensifying a screening policy are diminishing rapidly. An increase from 7 to 10 invitations per woman results in an additional number of 3,900 life-years gained, as an increase from 10 to 16 invitations only saves 2,900 life-years extra.

Table 2.4 Present values for 1988 of incremental social costs, life years gained and cost-effectiveness of three efficient screening policies and spontaneous screening. All outcomes are compared with the zero-option. Discount rate 5%.

	Efficient policies			
Invitations	7	10	16	Spontaneous screening
Incremental social costs (millions Dfl)	324	483	735	638
Life years gained (x 1000)	13.3	17.2	20.1	13.4
Costs per life-year gained (x 1000 Dfl)	24.3	28.0	36.5	47.6
Marginal costs/life-year gained (x 1000 Dfl)	24.3	40.8	86.9	-

The incremental costs rise faster, more or less proportional to the number of invitations. Consequently the incremental costs per life-year gained increase if screening is intensified. The efficient policy with 7 invitations per woman costs about 24,000 guilders per life-year gained. For the policy with 10 invitations the costs amount to 28,000 guilders. For policies with more than 10 invitations the cost-effectiveness deteriorates rapidly.

Spontaneous screening on a scale reflecting recent Dutch practice generates a number of life-years gained equal to the efficient policy with 7 invitations, but the social costs are twice as high. When spontaneous screening takes place data indicate that part of the women, especially younger women, is screened very frequently (yearly) whereas the rest of the women only rarely has a screen and hence the health effects are limited. If no spontaneous screening is performed and the efficient policy with 7 invitations were to implemented 20 millions guilders per year (financial costs) could be saved.

For a comparison of the efficient policies as well as for comparison with other health interventions, the marginal cost-effectiveness is important. Table 2.4 shows that extra costs per extra life-year gained rise rapidly when switching from 7 to 10 invitations. Further intensifying the policy yields a considerable increase in marginal cost-effectiveness.

To ensure a rational health services budget, cost-effectiveness should be compared with other services. A recent Dutch study of breast cancer screening using the same methodological assumptions (Maas et al., 1989) reported that breast cancer screening with 10 invitations per woman cost Dfl. 9,700 per life-year gained. Further extension to 15 invitations costs Dfl. 25,800 per extra life-year gained. For the same number of invitations cervical cancer screening appears to be less attractive. Comparison with other health care services is difficult because of methodological differences between analyses. However, in Torrance's widely cited table of cost-effectiveness ratio's (Torrance, 1986) the efficient policies of cervical cancer screening with 7 or 10 invitations have an intermediate cost-effectiveness ratio.

Sensitivity analysis

Cytological evaluation

On the basis of our recently collected data a workload of 7,200 smears per analyst per year has been estimated. Assuming that the cytology is regionally concentrated so that expertise can attain a sufficient level, our preliminary view is that 7,200 is a reasonable workload.

However, Dutch pathologists recently agreed (Vooijs, 1987) that if 85% of the smears are preventive (15% of the smears are from women with symptoms) the optimal workload to aim for is 5,000 smears. With 100% preventative smears the workload will then be about 5,500 smears. The costs of cytological evaluation would increase 15%, total costs of screening and the costs per life-year gained would rise 6%, compared with the results presented. This workload forms the subject of a future study.

Discount rate

Results were also calculated using discount rates of 0 and 7%. Table 2.5 shows that discounting at a rate of 7% implies that the number of discounted life-years gained is only 12% of the amount without discounting, whereas the incremental costs are still 54% of the amount without discounting. This confirms the importance of using a uniform discount rate in cost-effectiveness analyses.

Table 2.5 Influence of the discount rate on social costs, health effects and cost-effectiveness for the efficient policy with 7 invitations. The percentages compared with no discounting are in parentheses.

	Discount rate		
	0%	5%	7%
Incremental costs (millions Dfl)	510 (100%)	324 (64%)	277 (54%)
Life-years gained	68,300 (100%)	13,300 (28%)	7,900 (12%)
Costs per life-year gained	7,500 (100%)	24,300 (342%)	35,000 (467%)

Discussion

Quality of life

Comparison of cervical cancer screening with other health care services should be based on costs per quality of life-year gained. Our study did not produce QALY-estimates, but two extreme variants were calculated to indicate the possible impact of quality adjustments. In the "weak variant" the quality of life reductions for invasive cancer or being treated by hysterectomy or radiotherapy are assumed to be modest. In the "strong variant" these reductions are greater and they are also applied to larger groups of women in less severe circumstances, like women with false-positive smears (Habbema et al., 1988).

For the efficient policy with 7 invitations the weak variant produces slightly more QALY's gained than without adjustment. The strong variant results in about 25 % less QALY's. We expect that empirical research will produce estimates between these two extremes.

Other studies

Comparison with other studies of cervical cancer screening is difficult for three reasons. First, the number of serious cost-effectiveness analyses on this subject is very limited. Second, some of them deal with the epidemiology in detail but use only superficial economic assumptions (Parkin and Moss, 1986). Third, the only detailed study of economic aspects (Luce, 1981) focused on screening in a clinical setting which is different from a mass screening program- with respect to organization and costs.

Organized and spontaneous screening
In the Netherlands, as in many other countries a combination of organized and spontaneous screening existed in the past and will probably continue in the future. The results of some of these combinations were calculated. They incur high costs for an intermediate level of health effects, because some women are screened in both contexts.

Screening in the future
In the Netherlands the organized screening program terminated in 1984 was resumed in 1989. Women are invited from age 35 up to age 53 with a 3-year interval, this in spite of our findings that such a schedule is not efficient (in the efficient policies screening is performed up to 70 years). The costs per life-year gained amount to Dfl 36,400, assuming the organization described in this paper. However, in the new setting the general practitioners take the smears; the rest of the organization is not yet entirely established. Our preliminary research indicates that if the coordination of the activities of 6,000 general practitioners and quality control is performed seriously the costs will not be much lower than reported here. The level of health effects is very uncertain, because the attendance rate and the quality of the smears in the new setting remain unknown.

Our findings did not result immediately in any policy change. The government department of health care has maintained its old policy for organizational reasons. Further study of the effectiveness of screening in women older than age 55 will be carried out, which may influence the policy in the near future.

Conclusions

A large part of the costs of cervical cancer screening programs consist of the costs of screening itself. Screening enlarges the costs of diagnosis. Costs of primary treatment remain unaltered for efficient policies with 10 or less invitations, but increase in case of more invitations. Costs of terminal treatment decline.

The incremental costs per life-year gained are about 24,000 guilders for the efficient policy with 7 invitations and increases to 36,000 in case of 16 invitations per woman. The substantial influence of the discount rate on the cost-effectiveness confirms the need of uniform methodology.

Cervical cancer screening is less cost-effective than breast cancer screening on the same scale. Compared with other health care services the efficient policies with 7 or 10 invitations are cost-effective. For the same amount of cost, organized screening programmes are always superior to spontaneous screening. Discouraging spontaneous screening in The Netherlands and replacing it by efficient policy with 7 invitations per woman can save over 20 million guilders (financial costs) per year, without any rise in cervical cancer mortality. The Dutch public health insurance funds are trying to restrict the spontaneous screening, except for situations in which women have possible cervical cancer-related symptoms, a small minority of the cases.

3 Cervical cancer screening: attendance and cost-effectiveness[1]

Summary

The influence of attendance on the health effects and cost-effectiveness of cervical-cancer screening is studied both for organized screening programmes and for spontaneous screening. The asymmetric distribution of smears among the female population and the higher risk incurred by women who never or only occasionally attend screening appear crucial in determining the health effects of screening. An increase in attendance rate induces a substantial rise in health effects and a less than proportionate rise in costs, thus improving cost-effectiveness. Wider screening coverage, in order to increase the number of life-years saved, is achieved more efficiently by encouraging a greater number of women to attend than by inviting the same number of women to attend more often, i.e. with a shorter interval between successive screens. Spontaneous screening is characterized by high coverage for younger women and low coverage for middle-aged and older women. This leads to a small amount of health effects and poor cost-effectiveness as compared with organized screening.

Introduction

Screening for cervical cancer has been performed in several countries with varying success (Hakama, 1986; Laara et al, 1987). This stresses the need for evaluation of screening policies with respect to health effects and costs. The level of attendance of women is the major determinant of the cost-effectiveness cervical-cancer screening policies (Habbema et al, 1988; Koopmanschap et al, 1989). Therefore, the following aspects are considered:

- determining the level of attendance. The impact of elements of the screening organisation on the level of attendance will be analyzed.

[1] Chapter based on Koopmanschap et al., Int J Cancer, 1990

- relating attendance, health effects and costs. Special attention will be paid to the distribution of smears among women and to the differences in risk between attenders and non-attenders.
- implications for health policy. Two different ways to increase the screening coverage, namely, increasing the number of women attending and increasing the frequency of screening, will be evaluated in relation to additional health effects and extra costs.

The argument will concentrate on organized screening programmes, under which women are invited for screening by specially trained personnel at certain ages and at predetermined intervals. The findings for such programmes will be compared with the coverage, health effects and costs for 'spontaneous' screening, which is performed only when requested by the woman concerned.

Methods

The term 'attendance rate' applies to organized screening. It is the percentage of women invited to undergo screening, who respond to the invitation. The term coverage is more generally applicable, denoting the percentage of women screened at least once during a certain period. In the case of organized screening, the coverage rate reflects the reach of screening over one or several invitation rounds. Coverage can be calculated whether spontaneous screening is performed exclusively or in combination with organized screening. The latter situation is common practice in many countries.

A study of health effects and costs

Health effects and costs were predicted for 3 screening settings in the Netherlands, each in operation for the period 1988-2015: organized screening only, spontaneous screening only and the combination of organized and spontaneous screening (Habbema et al, 1988; Koopmanschap et al, 1990). The health effects and changes in treatment costs resulting from screening and becoming manifest in part after termination of the screening have also been taken into account. The computer simulation program MISCAN was used for the calculations, which were based on a model for cervical cancer and the effects of screening (Habbema et al, 1984). The program gives the results of screening policies in terms of numbers of screen-detected pre-invasive and invasive cancers, interval cancers, incidence in non-attenders and reduction in mortality. Influence of age-specific incidence, length of pre-invasive stages, regression, sensitivity of the smear and the stage-specific survival are incorporated.

The model was constructed on the basis of data from British Columbia, Canada (spontaneous screening) and from the Netherlands, where organized screening was performed in a pilot study in 3 regions (Nijmegen, Utrecht and Rotterdam) between 1976 and 1984 (EVAC, 1987; Boyes et al, 1982). In the Netherlands, women aged 35 to 53 were invited at 3-year intervals; the average rate of attendance was 65% (EVAC, 1987).

Additional research was carried out to estimate the costs of screening and subsequent diagnosis and treatment. This made it possible to simulate the total health effects and costs of about 100 screening policies. These policies differ in the number of invitations per woman in a lifetime (3-25), the length of the interval (2-15 years) and the age-group invited (always within the range 20-75 years).

The cost-effectiveness of the simulated policies is expressed as the incremental cost per life-year gained, compared with the (hypothetical) situation in which there is no early detection of cervical cancer. Policies producing the greatest number of life-years gained for a given level of costs (efficient policies) have been identified (Habbema et al, 1988). The results for different levels of attendance for 2 efficient policies will be presented. The conclusions are valid for all efficient policies. With respect to the health effects, only the number of life-years gained by screening will be shown. Taking the other health effects into consideration does not substantially alter the conclusions, which are presented elsewhere (Habbema et al, 1988).

Results

The most important determinants of attendance are first discussed and, wherever possible, their influence is quantified. The link between attendance and both health effects and costs is then presented.

Determinants of attendance

Age-group screened

In the Dutch pilot projects of cervical-cancer screening, the attendance rate of women aged 34 to 36 was 10% higher than for women aged 52 to 53 (EVAC, 1987). All programmes of organized screening in other countries report highest attendance rates for women aged 30 to 49 and slightly lower rates for age 20 to 29 (Evans et al, 1980; Berrino et al, 1979: Magnus et al, 1987; Johannesson et al, 1982; MacGregor et al, 1986). With increasing age, attendance declines; at age 50 to 59 the rate is about 10-15% lower than at age 30 to 49 and at age 60 to 69 the attendance is about 30% lower. For breast-cancer screening, a similar relation between age and attendance has been reported (Verbeek, 1985).

Socio-economic status
In the Dutch pilot projects, a lower attendance rate was found for women with the lowest as well as the highest socio-economic status (Jansen and Kremers, 1985). The latter group compensated by having spontaneous screens more often. Data from other countries also indicate lower coverage for women of lower socio-economic status (Evans et al, 1980; Armstrong et al, 1986; Ebeling and Nischan, 1986; Kleinman and Kopstein, 1981). For the UK Evans reported 30% difference in attendance between the highest and lowest status group.

Marital status
The Dutch projects reported a 15% lower than average attendance rate for unmarried and divorced women (EVAC, 1987). Data from other European countries show that the attendance of married women is about 10% higher than that of unmarried women (Evans et al, 1980; Berrino et al, 1979; Ebeling and Nischan, 1986).

Publicity
In the Dutch pilot region Nijmegen the influence of publicity on attendance was investigated. In some municipalities, publicity in the form of descriptive brochures, newspaper articles and information meetings had a positive but statistically not significant influence. In other municipalities the influence was significant but the difference in publicity was confounded with other differences between municipalities (Emanuel-Vink and Tax, 1979). Women with lower socio-economic status respond more frequently to personal contact, whereas women with higher socio-economic status mostly acquire their information from letters and mass media (Wakefield, 1976).

Invitation system
In the Dutch pilot projects, an up-to-date population registry in each municipality enabled selection and invitation of the eligible women. The invitation included an appointment for screening. This first invitation resulted in 60% attendance. A reminder yielded about 5% extra attenders, total attendance thus being on average 65% (EVAC, 1987). In the UK, the lack of an accurate and complete population registry would help to explain the low coverage (Elkind et al, 1987; Draper, 1982).

Almost all examples of successful screening policies (successful with respect to participation and health effects) use invitation systems based on population registries (Magnus et al, 1987; Johannesson et al, 1982; Hakulinen and Hakama, 1985; Berget, 1979). The absence of a registry or invitations appears to be closely correlated with low coverage (Parkin et al, 1982; Olesen, 1986). One exception is the British Columbia project (Anderson et al, 1988).

Few studies analyze systematically the influence of the invitation system on attendance. An invitation system with appointment induces an absolute increase of 15% in the attendance rate as compared with invitations making women responsible for arranging time and place of screening (Wilson, 1987). Results from several countries and the Dutch breast cancer screening program indicate that a reminder can increase attendance by 5-10% (absolute increase) (Berrino et al, 1979; Verbeek, 1985; Wilson, 1987; Sansom et al, 1975).

Travel distance and contribution in costs
In the Dutch projects the average travel distance for the women was less than 5 km, due to the use of stationary and mobile screening units. In the UK and Italy, attendance appears to be higher if (semi-) mobile screening units are used to shorten the travel distance (Berrino et al, 1979; Brindle et al, 1976). A study of breast cancer screening with mobile units in Scotland revealed a significant relation between travel distance and attendance rate (Henderson et al, 1988).

In the Dutch projects until 1984 women paid a small part of the screening costs. This did not appear to influence attendance. Between 1985 and 1987 the contribution in costs increased substantially and attendance fell, indicating that charging substantial contributions may impede attendance. Data from abroad are scarce, but it is remarkable that screening in Sweden and Finland combines high attendance with the absence of cost contributions (Hakulinen and Hakama, 1985; Kjellgren, 1986).

Assumptions used for attendance
The calculations are based on the observed attendance patterns in the Dutch pilot projects. Facilities such as an invitation system based on population registries including reminder, mobile screening units and publicity were considered indispensable to attain 65% attendance. Therefore, the costs of these facilities were taken into account.

For the relation between age and attendance we assumed a 65% attendance rate up to age 50, gradually decreasing to 60% at age 60, 55% at age 70 and 40% at age 75. The influence of socio-economic status on attendance has not been explicitly modelled, but is taken into account by the relation between attendance and risk (see below). The influence of marital status has not been used.

Attendance and health effects
The health effects of screening will be presented for three situations. First, attendance is assumed to depend on age and on the screening organisation only. Subsequently, the pattern of individual attendance is brought in. Finally the relation between attendance and risk will be added.

Age
In figure 3.1, the upper, dotted lines represent the health effects for two screening policies when only the relation between age and attendance is taken into account. Next to 65% attendance, the health effects are depicted for 40% and 75% average attendance. These results are too optimistic however, because taking into account the individual attendance and risk will result in lower health effects.

Individual attendance pattern
If all women invited had the same probability of attending at each screening round, the number of smears would be equally distributed over the women. However, in the Dutch projects, women who had attended the previous screening had a 50% higher probability of attending the current screening than women who had not (Jansen and Kremers, 1985). Moreover, 10% of the women invited did not attend any screening. In Denmark, the same attendance pattern was observed: a difference in attendance of 61%, and again 10% of the women never attending (Berget, 1979). The same attendance pattern was found for breast-cancer screening in the USA, and in the Nijmegen region of the Netherlands (Verbeek, 1985; Habbema et al, 1983).

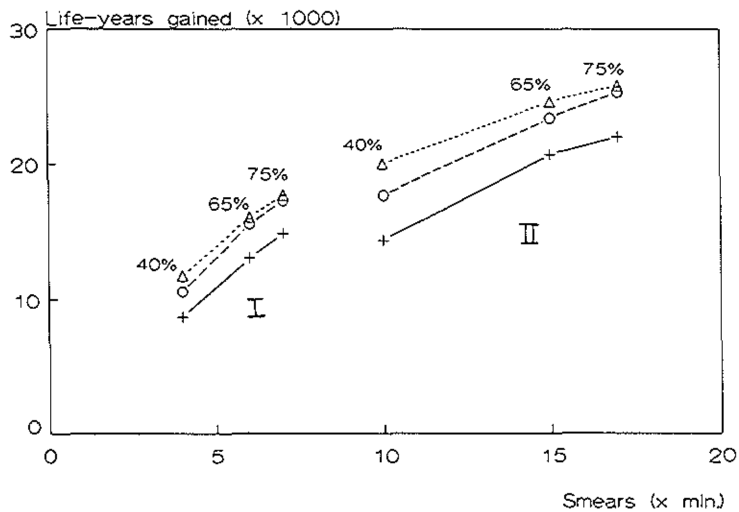

Figure 3.1 Number of smears and life-years gained by screening in the Netherlands during 1988-2015, for three attendance rates (40%, 65%, 75%) discount rate: 5%. Results for two organized policies: I: 7 invitations between age 37 and 73, interval 6 years. II: 16 invitations between age 28 and 72, interval 3 years.
Factors taken into account when estimating the health effects (see text):
△.....△ : age only
●-----● : age and attendance pattern
+———+ : age, attendance pattern and risk differences

The following assumption was used: irrespective of the average attendance, women who attended the previous screening have a 50% higher attendance probability than women who did not. Ten percent of the women never attend screening. Consequently, some of the women will be screened very regularly, some only occasionally, while others will never be screened. The more frequently a woman is screened, the lower the extra health effects of each additional screen.

Therefore, the more unequal the distribution of a certain number of smears over the women invited, the lower the aggregate health effects of screening. The middle, dashed lines in figure 3.1 depict the health effects resulting from adding this assumption. Compared with the outcomes for uniformly distributed attendance, health effects are between 2 and 10% lower.

Attendance and risk

Consensus exists that, controlling for age, risks of cervical cancer are unequally distributed. For example, on average, women with lower socio-economic status have a higher risk.

Data from British Columbia indicate that the incidence of cervical cancer in women who never attend screening is higher than in the total population (Boyes et al, 1982). Analysis of the Dutch data appears to confirm this hypothesis (Habbema et al, 1988).

In the three Dutch pilot regions, the cytological detection rate of CIN II+ was about 50% higher for women attending the screening after a reminder compared with women attending after the first invitation (EVAC, 1987). The difference is highly significant ($p<0.001$). Data from Denmark and Norway support the assumption that non-attenders have higher incidence and mortality risks (Magnus et al, 1987; Berget, 1979). However, in a case-control study of about 150 women in the Nijmegen region, women having first sexual intercourse at young age, a risk factor, were surprisingly found to be screened more often (Graaf, 1987).

Other risk factors should be positively correlated with non-attendance, to maintain compatibility with the evidence on the overall risk mentioned earlier. This evidence was reinforced in our modelling efforts to explain the incidence rates in attenders and non-attenders both from British Columbia and the Netherlands. For obtaining a good fit, it had to be assumed that *on the whole* lower attendance is related with higher risk. We assumed that the relative risk of attenders compared with that of the entire population (in the same age-group) increases from 0.69 in case of 40% attendance to 0.74 if attendance is 75%. By incorporating the relation between risk and attendance in this direct way, the underlying mechanism is ignored.

The resulting health effects are depicted by the lowest, solid lines in figure 3.1. The influence of risk turns out to be substantial, irrespective of the number of invitations. The health effects thus decrease another 13-17%. These more realistic estimates result in a 15-28% lower level of health effects compared with the situation in which attendance is assumed to be influenced by age only.

Cost-effectiveness

Figure 3.1 also shows that higher attendance leads to a substantial increase in health effects, especially if the screening policy is not very intensive, as in case of 7 invitations. On the other hand, the total costs increase less than proportionally with the level of attendance (Koopmanschap et al, 1990). The total costs are dominated by the costs of screening itself. The changes in costs of subsequent diagnosis and treatment induced by screening are relatively small. The costs of screening itself increase less than proportionally because the major share of costs for coordination and registration is fixed. This implicates that the incremental costs per life-year gained (as compared with the situation without any early detection) fall as attendance increases (table 3.1). Only in case of 16 invitations does a rise in attendance from 65 to 75% fail to improve cost-effectiveness further. This illustrates the decrease in extra health effects when an already intensive policy is extended.

Table 3.1 Life-years gained (x 1,000) and incremental costs US $ (million) and cost-effectiveness (x 1,000 US $) for efficient screening policies with 7 and 16 invitations in the Netherlands during 1988-2015 for several attendance rates. Discount rate: 5%.

Policy	Attendance	Costs	Life-years gained	Cost/life-year gained
		($ million)	($ x 1,000)	($ x 1,000)
7*	40%	126	8.7	14.5
7*	65%	163	13.1	12.5
7*	75%	176	14.9	11.8
16**	40%	267	14.4	18.5
16**	65%	366	20.7	17.7
16**	75%	407	22.0	18.5

* 7 invitations between age 37 and 73 with a 6-year interval.
** 16 invitations between age 28 and 72 with a 3-year interval.

Improving the health effects of screening

The health effects of cervical-cancer screening can be improved in two ways:
A) by trying to increase attendance, without changing the invitation schedule;
B) by intensifying the policy by sending more invitations per woman (in the same age-range, at shorter time intervals).

Two examples will illustrate the divergent results of these options. Suppose that an efficient policy with 7 invitations is in operation. Only a first invitation is sent, with no reminder, and the attendance is 60%. A comparison is made between the results of sending a reminder (A) and those of increasing the number of invitations per woman in her lifetime from 7 to 8 (B).

Both methods produce approximately the same amount of extra health effects (table 3.2 and figure 3.2). The costs of raising attendance to 65% by sending a reminder could be calculated on the basis of the Dutch data. This option gives rise to extra costs of 3 million US $ for producing the reminder and additional coordination, and another 9 million US $ because 700,000 additional smears are taken, i.e. 12 million US $ in all. The alternative, extending the invitation schedule (B) leads to a greater number of additional smears (2.2 mln), costing an extra 16 million US $, so sending a reminder is the more cost-effective option.

Table 3.2 Incremental costs (in US $ millions) and life-years gained (x 1,000) by (A): increasing attendance from 60% to 65% by sending a reminder; versus (B): intensifying the policy from 7 to 8 invitations. Screening in the Netherlands during 1988-2015. Discount rate: 5%. (rem=reminder).

Invitations	Attendance	Smears	Costs	Life-years	Cost/life-years gained	
		(mln)	($ mln)	($ x 1,000)	($ x 1,000)	
					Average	Marginal
7*	60% (no rem)	9.9	152	12.0	12.7	-
7* A	65% (rem)	10.6	163	13.1	12.4	10.0
8** B	60% (no rem)	12.1	167	13.2	12.7	15.0

* 7 invitations between age 37 and 73 with a 6-year interval.

** 8 invitations between age 37 and 72 with a 5-year interval.

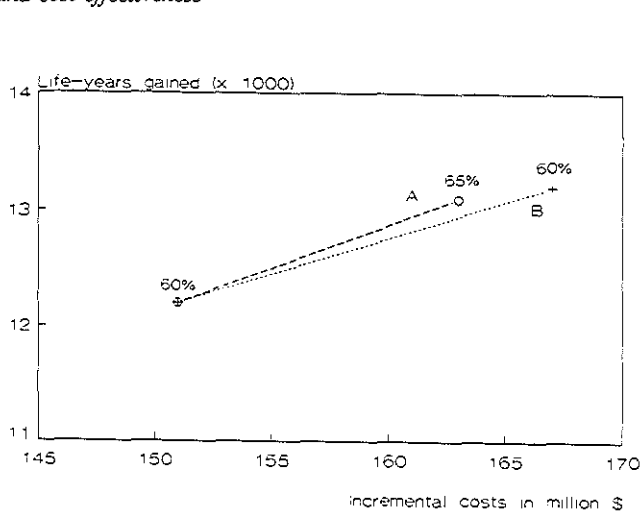

Figure 3.2 Example of incremental costs and life-years gained by (A): increasing attendance from 60% to 65% by sending a reminder; versus (B): intensifying the policy from 7 to 8 invitations. Screening in the Netherlands during 1988-2015. Discount rate 5%.

Intensifying the policy by sending more invitations implies that most of the extra smears will be taken from women who are already screened relatively often; consequently the additional health effects of each extra smear are limited. Increasing the attendance rate means that the (smaller number of) extra screens will be allotted to the women least screened, i.e., with higher risk than average. The extra health effects for each additional smear are larger than in case of inviting more often.

In the second example it will be assumed that attendance can be raised further, up to 78% ((A) in figure 3.3 and table 3.3). Seventeen million US $ extra costs are necessary for taking, evaluating and registering the additional smears. Attaining the same amount of extra health effects by extending the invitation schedule to 9 invitations (B) doubles the extra costs: 34 million US $. Raising attendance seems by far the most attractive option.

However, stimulating attendance, for instance by additional publicity, will definitely require extra costs, over and above the 17 million US $ already mentioned. The exact amount of extra costs is difficult to estimate: they will probably rise for each percentage of higher attendance. Figure 3.3 shows that the maximum amount of extra costs, allowed by rational budget spending, is the difference in additional costs between method A en B: 34-17= 17 million US $. In general, the more intensive the policy, the greater the amount of costs can be devoted to stimulating attendance. For example, raising attendance from 65% to 75% for the efficient policy with 16 invitations may cost 70 million US $.

Table 3.3 Incremental costs (US $ mln) and life-years gained (x 1,000) by (A): increasing attendance from 65% to 78%; versus (B): changing the policy from 7 to 9 invitations. Screening during 1988-2015. Discount rate: 5%

Invitations	Attendance	Smears	Costs	Life-years	Cost/life-years gained	
		(mln)	($ mln)	($ x 1,000)	($ x 1,000)	
					Average	Marginal
7*	65%	10.6	163	13.1	12.4	-
7* A	78%***	12.7	180	15.4	11.7	7.4
9** B	65%	13.5	197	15.4	12.8	14.8

* 7 invitations between age 37 and 73 with a 6-year interval.
** 9 invitations between age 35 and 75 with a 5-year interval.
*** 78% was chosen, it yields the same health effects as the policy with 9 invitations.

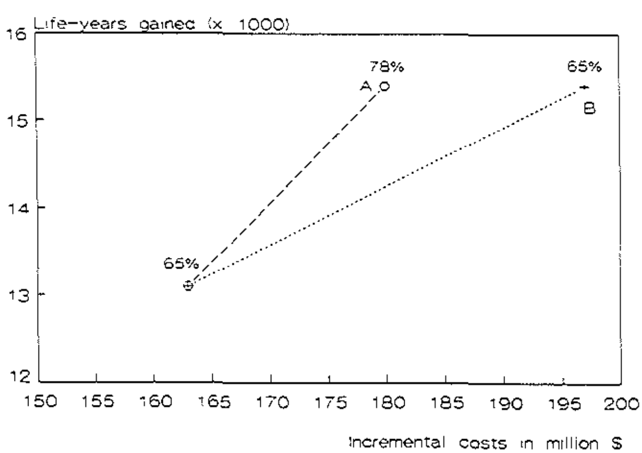

Figure 3.3 Example of incremental costs and life-years gained by (A): increasing attendance from 65% to 78%; (B): intensifying the policy from 7 to 9 invitations. Screening in the Netherlands during 1988-2015. Discount rate 5%.

Elements of the screening organisation, such as the invitation system, publicity and travel distance, could be used as instruments to increase attendance. This has also been suggested by several authors (Hakama et al, 1985; Eardley et al, 1985; Haran et al, 1986). Further research on the costs of these instruments and the resulting extra attendance should be stimulated. Decisions concerning these features should preferably be based on considerations of cost-effectiveness, along the lines presented in this chapter.

Spontaneous screening

Data for the Netherlands reveal that in the case of spontaneous screening, without a population-based invitation system, the distribution of smears among women is more uneven than in the case of organized screening. Some women are screened only occasionally and about 10% are never screened. There is no clear indication that women at higher risk will participate more frequently than in the case of organized screening (Habbema et al, 1988).

In the Netherlands, as in many other countries, the coverage is the highest in the age range 20 to 35 (Hakulinen and Hakama, 1985; Parkin et al, 1982; Anderson et al, 1988; Kjellgren, 1986; Choi and Nelson, 1986). In this age-group, screening is often performed every 2 years or more frequently. The coverage rate in older age groups is substantially lower. This implies that spontaneous screening does not reach those who would really benefit from screening: middle-aged and older women at high relative risk (Laara et al, 1987).

Table 3.4 compares the coverage rates for one efficiently organized screening policy and for spontaneous screening. The coverage rates differ substantially, for approximately the same number of smears and level of costs. Spontaneous screening shows a relatively high coverage below 40. However, organized screening (7 invitations between age 37 and 73) reveals a far better coverage of the middle-aged and older women. Since the incidence of invasive lesions is relatively low below age 35, better coverage in this age range yields only a small number of life-years gained. Furthermore, the high portion of regressive lesions in younger women results in a relatively large amount of unnecessary diagnosis and treatment. Organized screening, directed to the age groups with the highest incidence, is superior with respect to the number of life-years gained and consequently in terms of costs per life-year gained.

Table 3.4 Coverage rate of cervical cancer screening (% of women screened at least once in 6 years) by age, for organized screening only and spontaneous screening only in the Netherlands 1988-2015. Number of smears, life-years gained, incremental costs, incremental costs per life-year gained (discount rate 5%).

		Organized screening (7)*	Spontaneous screening
Coverage rate:			
Age	25	0%	28%
	30	0%	45%
	40	65%	49%
	50	65%	48%
	60	60%	30%
	70	55%	7%
Number of smears (mln)		10.6	12.7
Incremental costs ($ mln)		163.0	178.0
Life-years gained ($ x 1,000)		68.0	43.0
Cost/life-year gained ($ x 1,000)		12.4	26.9

* 7 invitations between age 37 and 73 with a 6-year interval.

Discussion

The model used remains a simplified approximation of a complex reality. It is not always possible to detect the very true nature of the underlying relations. For example, it remains unclear if the negative relationship between attendance and risk can be explained by socio-economic status. If so, future social developments influence the link between attendance and risk.

Other factors
Efforts may be directed to influence other factors besides attendance which can alter costs and/or health effects. One of the most important determinants is

the sensitivity of the smear, which was varied in our study between 60% and 80%. Such variation does not influence the cost-effectiveness to the same extent as attendance. The same conclusion holds for other factors.

Organized and spontaneous screening
A combination of organized and spontaneous screening, a common practice in many countries, is very inefficient, according to our calculations. Many women, especially young women, will be screened in both contexts, inducing only a marginal improvement in coverage. The additional health effects of the spontaneous screening are limited, whereas the large amount of smears results in high costs. Again, intensive screening of younger women leads to a relatively large amount of unnecessary diagnostic and treatment procedures.

Generalization
The conclusions for the Dutch situation concerning the importance of attendance and the superiority of organized screening will to a large extent be valid for other countries, since we found many striking similarities between our data and other published findings. For countries with substantially higher incidence, all other parameters being equal, the cost-effectiveness will be more favourable than in our calculations.

The future of screening in the Netherlands
From 1989 onwards the Dutch organized-screening programme (7 invitations between age 35 and 53, with a 3-year interval) will be resumed, but modified as follows:
- only a first invitation, no reminder, is to be sent to women in the target age range;
- women should make an appointment to visit their general practitioner for a smear.

The absence of a reminder, and rendering women responsible for making the appointments will probably lower the attendance substantially. It therefore seems unfavourable from the point of view of cost-effectiveness.

Inviting women with higher risks
The foregoing analysis may suggest that special efforts to invite women at high risk for cervical cancer will increase the health effects of screening at little extra costs. Three arguments may be advanced against this suggestion. First, it is impossible to identify high-risk women by means of a population

registry. Second, general measures, such as a reminder, directed towards increasing attendance, tend to raise the attendance of women at higher risk. Third, special campaigns might be justified if some of the women had a very high risk compared with the rest. This is not the case for cervical cancer (Hakama, 1986).

Equity
Women with low socio-economic status have a higher-than-average risk for cervical cancer and below-average attendance at screening. Increasing attendance will attract these women more than proportionally, and therefore contribute to a reduction in health inequalities.

Breast cancer screening
Our research group also carried out a cost-effectiveness analysis of breast-cancer screening (Maas et al, 1988, 1989). For breast cancer, no clear relationship exists between attendance and risk. The much shorter duration of screen-detectable breast cancer makes frequent screening more cost-effective than in the case of cervical cancer. Taking this point into account, increasing attendance does not bear a clear advantage over intensifying the policy, as in the case of cervical-cancer screening.

Conclusions

Attendance rate is a major determinant of the health effects and cost-effectiveness of cervical cancer screening.

For a given number of smears, the amount of life-years gained is maximal if as many women as possible are reached. This holds true particularly because, on average, less frequent attenders or non-attenders are those at greater risk. Stimulating attendance preferentially attracts the least-screened women at greater risk. This is more effective, therefore, than inviting women more often during their lifetime, which merely means that women screened regularly will be screened even more often. Moreover, stimulating attendance helps to reduce health inequalities resulting from socio-economic differences.

A reminder invitation (after the first invitation based on a population registry) is a cost-effective instrument for improving the attendance. Other aspects of the screening organization, such as specific publicity and mobile screening, appear to be useful tools for further increasing attendance. Spontaneous screening results in low coverage for middle-aged and older women. Hence the health effects and cost-effectiveness are relatively poor. The same applies, *a fortiori*, to combined organized and spontaneous screening, which is common practice in many countries. Spontaneous screening therefore should be discouraged and resources should be reallocated to well organized screening programs.

Supplementary comment

The previous review of costs and health effects of cervical cancer screening was based on the analysis of centrally organized screening in the eighties in three Dutch regions as well as cervical cancer screening in other countries.

Since 1992 the cervical cancer screening program in the Netherlands is less centrally organized, with nearly complete national coverage. Women are invited to visit their general practitioner to be screened. The analysis has been updated for both costs and health effects. However, the main conclusions as described above still hold (van Ballegooijen et al. 1993). The costs of the screening program itself remain dominant, the subset of efficient policies is still characterized by a broad age-range and a relatively long interval between successive screens.

The relation between costs and effects has become somewhat less favourable, mainly because of the lower attendance rate, underlining the crucial role of attendance for cost-effectiveness. Recently, the Dutch government decided to extend the interval between successive invitations from 3 to 5 years and to broaden the age-range from age 35-53 to age 30-60.

Part II Aggregate Analysis

4 Cost of diseases in international perspective[1]

Summary

This study compares the health care costs of the Netherlands with the United States and Sweden and estimates the impact of demographic change on costs. Total health care costs were allocated to disease, age, sex and specific subsectors. For the Netherlands 75% of the costs in 1988 were assigned to specific diseases. Costs of mental disorders and other chronic non-fatal diseases dominate, followed by cardiovascular diseases. The effect of age is strong from age 70 onwards. The effect of sex, adjusting for age, is small, except for elderly women, being more expensive. Both total and disease-specific costs are similar in the Netherlands and Sweden, but differ from those in the United States. The available data suggest that the differences in medical practice and health care system may explain a substantial part of the divergent results; demographic or epidemiologic aspects seem less important. Ageing induces, in the Dutch case, a modest 0.7%point annual increase in costs. The contribution of other forces in raising costs is probably more important. A structural upward pressure on costs also prevails in the Netherlands and Sweden, but it is more prominent in the U.S.A., due to a large amount of expensive surgery and high administration costs.

Introduction

The level of medical consumption is determined by epidemiology, medical practice, economic factors including the type of health insurance system and cultural circumstances. Having stated this, it is surprising that the recent debate on value for money in various health care systems focuses largely on aggregate expenditure levels, paying little attention to the important relation

[1] Chapter based on Koopmanschap et al., EJPH, 1994

between disease and medical consumption (Schieber and Poullier, 1991). In this study the costs of illness are estimated for the Netherlands, covering the entire range of diseases and providing data on disease specific costs, according to age, sex and health care sector.

Next the level of medical care costs and the distribution according to disease will be compared with the United States and Sweden. For some diseases we will indicate whether the remarkable differences in costs appear to be due to divergence in demography and epidemiology or whether supply and financing of health care are most relevant in explaining these differences. We can not provide a complete analysis for all diseases, as this requires a reliable international overview of (longitudinal) morbidity data, which is currently lacking. Still this analysis may contribute to the debate on the quality and costs of health care.

Another important policy item is estimating the consequences of ageing. It will be shown how rapidly medical costs may increase due to demography alone, compared with the impact of other expenditure increasing factors.

Methods

Cost of illness in the Netherlands

Total health care costs in the Netherlands, excluding care in old peoples homes and social work, amounted to 39.8 billion Dutch guilders (ECU 15.9 billion), according to the national Financial Overview of Health Care (WVC, 1989). These costs were allocated to age, sex, health care sector and disease categories (Koopmanschap et al., 1991). To this end health care was divided into 18 sectors. All diseases of the International Classification of Diseases 9-CM (WHO, 1975) were assigned to 48 disease categories, using as criteria comprehensive data on morbidity, mortality, costs and expected developments. They can be aggregated to the standard 17 ICD-chapters. The primary diagnosis of patients was used to assign costs to specific diseases.

Most data were extracted from national registers with nearly complete coverage. The Dutch hospital registration system was described by Casparie (Casparie and Hoogendoorn, 1991). For some other sectors we use data from large surveys such as the Dutch National Health Interview Survey (CBS, 1988), which is comparable to its US counterpart NHIS. For a few sectors insurance data were used (table 4.1). The available data on diagnosis were all registered by the responsible physician or nurse, conform the ICD 9-CM rules. Costs incurred by dental or maternity services were allocated to dental diseases and pregnancy and delivery respectively. Information on diagnosis was not complete or absent regarding care for physically handicapped, home help, pharmaceuticals, general preventive care, patient transport and administration. They represent 25 % of the health care costs.

Table 4.1 Overview of methods and data used for allocation of costs to disease categories per health care sector in the Netherlands (NL), Sweden and the United States.

Health care service	% health care costs NL	Data on diagnosis NL	Data type and source NL	Allocation Netherlands	Allocation Sweden	Allocation United States
Hospital: nursing and physician services	32%	+	National register	Hospital days	Hospital days & cost per day; Outpatient visits	Hospital days & expenses per day; Outpatient visits; Surgical procedures
Hospital: surgery	3%	+	National register	Surgical procedures and tariffs		
Nursing homes	10%	+	National register	Nursing home days	Nursing home days	Residents & monthly expenses
Psychiatric care in-patient	6%	+	National register	Residents	Expenditures	Residents
Psychiatric care out-patient	2%	+	National register	Clients	Expenditures	Psychiatric visits
Care for mentally retarded	6%	+	National register	Residents	Expenditures	Residents **
Care for physically handicapped	3%	+-	Registers and samples	Residents	Not allocated	Residents **

Health care service	% health care costs NL	Data on diagnosis NL	Data type and source NL	Allocation Netherlands	Allocation Sweden	Allocation United States
General practitioner	5%	+	Detailed sample GP and patients	Visits	Visits	Visits
District nursing	2%	+	Regional register	Visits	Visits	Hospital days *
Home help	4%	-	Sample	Visits	Not allocated#	Hospital days *
Maternity services	1%	+	Insurers data	Expenditures	Expenditures	Visits
Physical therapy	3%	+	Sample	Visits	Visits	Hospital days *
Dental care	4%	+	Dutch HIS	Visits	Not allocated	Visits
Drugs	9%	-	Dutch HIS	Prescriptions	Prescriptions and prices	Physician visits
Eyeglasses and appliances	2%	+	Insurers data	Expenditures	Not allocated	Persons glasses/ hearing aids
Governmental public health care	2%	+-	Government department	Equal costs per person	Not allocated	Not allocated
Patient transport	1%	-	Insurers data	Expenditures	Not allocated	Not allocated
Administration (insurers & government)	6%	-	Government and insurers	Equal costs per person	Not allocated	Not allocated

+ : information on diagnosis conform ICD 9-CM
+- : information on diagnosis partially conform ICD 9-CM
- : no information on diagnosis available
* : costs were allocated to disease, using the distribution of hospital days
** : not allocated separately, as they are included in the hospital sector
# : in Sweden home help is not included in health care but in social services

For each health care sector total costs were assigned to age, sex and diagnosis by using utilization of services, the number of patients or the expenditures per group of patients (table 4.1). For example, the number of visits to physical therapists determines the distribution of costs for physical therapy, assuming each visit has the same costs. Regarding hospitals and nursing homes, we used the number of hospital days; for psychiatric care and care for mentally or physically handicapped the number of patients was used to allocate costs. For most ambulatory services we used the number of visits to attribute costs. In case of maternity services, appliances and patient transport cost allocation was based on expenditures.

Data on age and sex were available for all health services, except for general preventive care and administration; considered as indivisible goods, they have been assigned by assuming equal costs per citizen.

International comparison

The results for the Netherlands will be compared with Sweden (Lindgren, 1990) and the United States (Rice et al., 1985). All three studies use the primary diagnosis to assign costs to disease categories. The data used are a mixture of national and regional registrations, samples and surveys on health care utilization and costs. In the US study non-personal health costs such as costs for government public health care and insurers and government administration were excluded. Costs of patient transport were not allocated to a specific disease (table 4.1). In Sweden, home help is a part of social services and was therefore excluded. The costs of dental care, appliances, patient transport, government public health care and administration were not assigned to disease.

The proportion of costs which could be assigned to specific diseases is similar in the three countries. For Sweden 78% was assigned, for the Netherlands 75%, and the United States 82%. (table 4.2). As detailed cost analyses of the entire health care system would have been too complicated each study used aggregated data on utilization and costs to assign costs to diagnosis, age and sex. The procedures used in distributing costs to diagnosis are fairly similar, with a few exceptions. The US study used estimates of costs per hospital day by diagnosis category (table 4.1). In the Swedish study, costs per hospital day were made disease specific, by using cost estimates for several hospital departments. For the Netherlands, we assumed equal cost per day for all diseases, as reliable disease specific data on the costs per hospital day are

Table 4.2 Distribution of direct costs by disease categories (17 chapters of the ICD 9-CM) in the U.S.A. 1980, Sweden 1983 and the Netherlands 1988 in percentages of total direct costs.

Disease category (ICD chapters)	U.S.A.*	Sweden	Netherlands
	%	%	%
Infective diseases	1.8	1.6	1.0
Neoplasms	5.5	5.1	4.6
Endocrine and metabolic dis.	3.1	2.3	1.7
Diseases of the blood	0.4	0.5	0.3
Mental disorders	8.3	21.1	19.9
Dis. of the nervous system	7.1	4.2	3.8
Cardiovascular diseases	13.6	12.3	8.7
Respiratory diseases	7.0	5.0	3.1
Dis. of the digestive system	12.9	3.8	7.6
Genito-urinary diseases	5.1	3.0	2.7
Pregnancy and delivery	*	2.7	3.5
Diseases of the skin	2.5	1.7	0.9
Locomotor diseases	5.5	3.9	7.0
Congenital anomalies	0.5	0.5	0.4
Perinatal diseases	**	0.5	0.9
Symptoms	1.6	5.1	4.3
Accidents, poisoning and violence	7.7	4.8	4.3
Other diseases	3.6**	0.0	0.0
Not allocated	14.2*	21.9	25.2
Total	100.0	100.0	100.0

* For comparability, the U.S.A. percentages derived by Rice were corrected for non personal health care costs (e.g. insurers- and government administration and government public health services). These costs now make up the lion's share of unallocated costs.
** For the U.S.A. the costs of pregnancy and congenital anomalies could not be discerned separately and are mentioned as costs of other diseases.

currently not available. The Dutch data on non-hospital care appear to be more complete and reliable. For example, in the US study the costs of several home health care services had to be assigned to diseases, using the distribution of hospital days.

Demography and costs
Using the estimated average costs per person by disease, age and sex for 1988, we calculated future health care costs, on the basis of expected demographic development (CBS, 1989). We used 2030 as the endpoint of our predictions, when ageing of the Dutch population is expected to reach its maximum. In order to relate the increase in costs to expected economic growth we used the "Europe-scenario" drawn up by the Netherlands Central Planning Bureau, predicting a 2.8% yearly increase in real GNP (van der Berg, 1991).

Results

Costs by disease and sex
In figure 4.1 we show the costs by 17 disease categories and sex. Mental disorders (including mental retardation) generate by far the highest costs, accounting for 20% of total costs, followed by cardiovascular diseases (9%), disease of the digestive system (8%, including dental diseases) and locomotor diseases (7%). The most expensive disease categories are chronic, non-fatal illnesses. Cardiovascular diseases and cancer, accounting for 71% of Dutch mortality, make up only 14% of the health care costs. Women account for 57% of total costs, including the costs of pregnancy. The distribution of costs among ICD-chapters is quite similar for both sexes. Costs of cardiovascular disease are relatively more important for men (10% vs 8%), locomotor diseases for women (8% vs 6%). Due to standard registration procedure, the small amount of pregnancy-related costs for men are attributable to healthy babies who stay a short period in hospital, after birth.

The results for the U.S.A. (Rice et al., 1985) are similar. Rice concludes that women account for 59% of health care costs and that the distribution of costs by disease is not very sex-specific.

Costs by disease and age
The pattern of costs of illness related to age is quite diverse. For the age-group 0-19 year, costs of mental disorders, gastro-intestinal diseases (in particular dental care) and perinatal diseases are the most expensive. In the

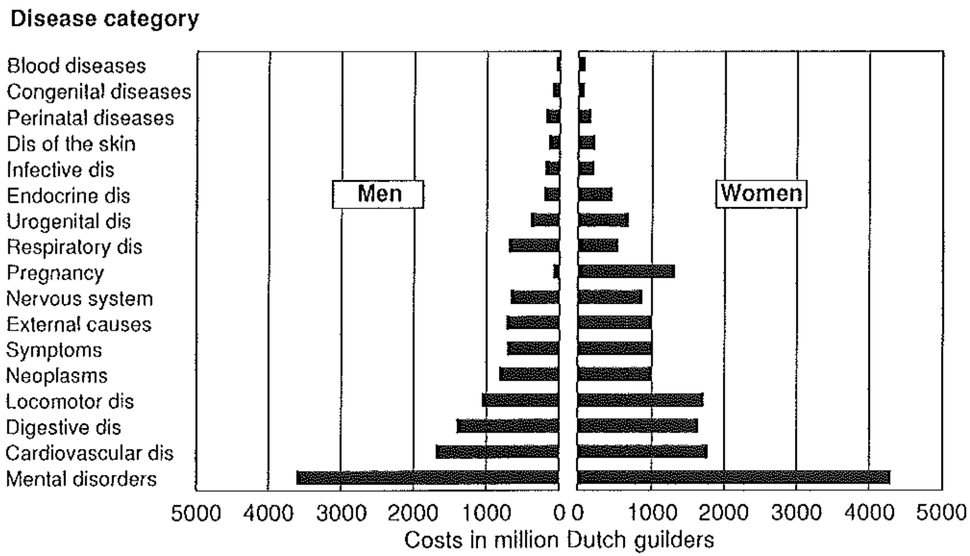

Figure 4.1 Costs of health care by disease category and sex for the Netherlands 1988, in million Dutch guilders.

age group 20-44, costs of mental disorders still dominate, followed by costs of pregnancy, dental diseases and locomotor diseases. From age 45 to 64 cardiovascular, locomotor disease and cancer represent the greatest costs, followed by mental disorders. Costs of cardiovascular disease are the most important in people aged 65-79. From age 80 onwards, the typical impairments of the most elderly are reflected in high costs for dementia, stroke, locomotor disease and non-traffic accidents.

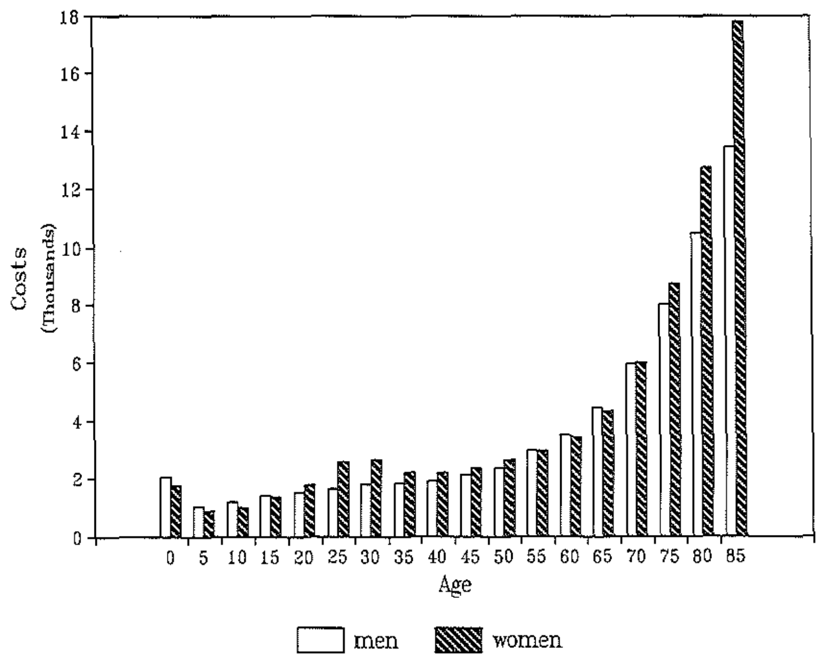

Figure 4.2 Average health care costs per year per person by age and sex for the Netherlands 1988, in thousand Dutch guilders.

Costs by age and sex

Figure 4.2 presents the average annual costs per person by age and sex. These costs rise gradually from dfl 1,000 for age 5-9 to 3,000 for age 55-59. From age 60 on costs accelerate, reaching a maximum of dfl 18,000 for age 85 and older. Given their age, differences in costs between men and women are small, except for the higher costs for women of age 80 and older. The causes of this exception are at least twofold: these women are on average older than men in the same age-group and more often single, which increases their probability of hospitalization (Huijsman, 1990).

The higher costs for women aged 20-44 are due to pregnancy and childbirth. Our Dutch cost estimates correspond quite closely to results for the U.S.A. In 1980 the average costs per U.S.A.-citizen older than 65 were 2.75 times the average for all citizens (Hodgson and Kopstein, 1984); the corresponding Dutch ratio is 2.86 in 1988. Waldo's estimates for health care costs by age for the U.S.A. in 1987 confirm this conclusion (Waldo et al., 1989).

International comparison

Table 4.2 presents the distribution of health care costs by ICD-chapters for the United States (1980), Sweden (1983) and the Netherlands (1988). A satisfactory explanation of all differences requires a systematic analysis of health care systems, demography and epidemiology, which is beyond the scope of this study. The main obstacle for a complete analysis is the lack of reliable, comparable data regarding disease specific morbidity. Nevertheless, our partial comparison reveals interesting outcomes.

The results for Sweden and the Netherlands are very similar; the low costs for diseases of the digestive tract in Sweden are due to the omission of dental care in Lindgren's study. The U.S.A. pattern of costs deviates on several points from the two European countries. The low costs for mental disorders in the U.S.A. probably do not reflect a lower prevalence. Two meta-analyses show no different prevalence rates of dementia for Europe (Jorm, 1987) and the U.S.A. (Ritchie, 1992). For mental retardation evidence is less firm, but prevalence estimates for children in the U.S and Scandinavia are in the same range (Yeargin-Allsopp et al., 1992; Dupont, 1989). Therefore, it appears that the low costs for mental illness in the U.S.A. are due to the small extent of inpatient psychiatric care: the number of beds per 1,000 citizens in psychiatric hospitals in the Netherlands and Sweden is much higher than in the U.S.A., both for 1980 and 1989 (table 4.3).

For cardiovascular disease comparative morbidity data are not available, but the high U.S.A. costs may be at least in part related to more surgery, as the relative frequency of coronary artery bypass grafting in the U.S.A. in 1988 is more than twice the Dutch and Swedish rate (table 4.3). In addition, the rate of increase in performing bypasses between 1985 and 1988 for the U.S.A. was much larger than for the Netherlands. The automatic reimbursement of costs in the U.S.A. clearly provides an economic incentive for increasing the number of CABG's, whereas in the Netherlands and Sweden the number of bypasses is highly regulated (Banta and Bos, 1991).

Regarding digestive diseases the story is analogous: the surgery rate in the U.S.A. is 16.6 per 1000 citizens as compared to 9.2 for the Netherlands (OECD, 1991; SIG, 1990), whereas in both countries surgery for digestive diseases makes up 16% of total surgery.

The substantial costs of accidents, poisoning and violence in the U.S.A. are undoubtedly due to their high incidence. The U.S.A. mortality rate for traffic accidents involving passenger cars is twice as high as for the Netherlands and Sweden. As this cannot be explained by differences in lethality, it points to higher incidence, caused by the larger mobility in the United States (United Nations, 1987 and Beeck et al., 1989 and WHO, 1991). With respect to homicide an average U.S.A. citizen has a 10 times higher risk of being killed than the average Dutchman, indicating a much larger incidence of accidents related to violence (WHO, 1991).

Turning to total costs, table 4.3 shows that both the share of costs in GDP and the average costs per capita are substantially higher in the U.S.A. Furthermore, the costs are still rising in the U.S.A. while they are stabilizing in the Netherlands and Sweden (OECD, 1991). Hence, one may ask: what makes U.S.A.-health care so expensive?

Table 4.3 Public health and health care in the Netherlands, Sweden and the U.S.A.

Item	Year/Gender	Netherlands	Sweden	U.S.A.
Epidemiology				
Life-expectancy at birth in years[2]	male female	73.7 80.5	74.2 80.4	71.6 78.6
Infant mortality rate per 100,000 live births[2]	male female	798 563	668 553	1,099 866
Mortality rate for traffic accidents with passenger cars per 100,000 persons[1]	1987	5.2	5.7	10.6
Mortality rate for homicide per 100,000 persons[2]	male female	1.1 0.7	1.5 0.9	12.8 4.0
Costs				
Health care costs as share of GDP in %	1988 1990	8.2% 8.1%	8.5% 8.7%	11.4% 12.4%
Costs p person US $ PPP	1990	1,182	1,421	2,566
Input				
Physicians per 1000 inhab.	1988	2.4	2.9	2.3
Psychiatric care beds per 1000 inhabitants	1980 1989	1.7 1.6	3.2 1.8	0.9 0.6
Hospital beds per 1000 inhab.	1988	4.5	4.1	3.7
Output				
Av. length hospital stay	1988	11.3	7.5	6.5
In hospital surgery per 1000 habitants	1988	56.8[3]	NA	105
CABG surgery per mln inhab.[4]	1985 1988	480 583	250 464[5]	780 1,265
Incomes				
Av. earnings per physician US $ PPP	1988	70,000[6]	48,000	144,700

* PPP= Purchasing Power Parities; Sources: if not indicated by superscript: OECD, 1991; [1] United Nations, 1987; [2] WHO, 1990 and 1991; [3] SIG, 1990; [4] Banta and Bos, 1991; [5] Jendteg, 1991; [6] Personal Communication, 1991.

If a less favourable age-composition of the U.S.A.-population were the cause, standardizing the population of the Netherlands to that of the U.S.A. regarding age and sex (US Bureau of the Census, 1989), combined with actual pro capita costs for the Netherlands, should result in higher projected costs compared to the actual Dutch costs. However, this calculation gives almost identical costs, so differences in demography can be ruled out as an explanation. Moreover, although the Swedish population is considerably older, Swedish health care costs are still quite moderate.

Undoubtedly, the high administration costs in the U.S.A., especially within hospitals, play an important role in expansion of costs. For the U.S.A., these were estimated as between 19 and 24% of the total health care costs in 1987 (Woolhandler and Himmelstein, 1991). Using the same definition with respect to administration costs for the Netherlands we estimate them to be 11-14% in 1988 (table 4.4).

Table 4.4 Administration costs (Woolhandler's definition) as % of total health care costs for the U.S.A. 1987 and the Netherlands 1988.

Expenditure category	Netherlands 1988	U.S.A. 1987
Program administration and insurance overhead[1]	4.4%	5.1%
Hospital administration[2]	3.5%	7.8%
Nursing home administration[2]	0.8%	1.3%
Physicians' overhead[3]	3.5%	5.1-9.8%
Total	11.7-13.7%	19.3-24.1%

Sources: [1] For the U.S.A. (all items): Woolhandler and Himmelstein, 1991; [1] For the Netherlands the calculation was based on WVC, 1989; [2] For the Netherlands: National Hospital Institute, 1988; [3] For the Netherlands: Personal Communication, 1991.

Regarding input, the number of physicians and beds in acute care hospitals per citizen in the U.S.A. is slightly smaller than for the Netherlands and Sweden (table 4.3), which cannot explain higher costs. Turning to output, the length of stay in acute care hospitals in the U.S.A. is lower than in the Netherlands and comparable to Sweden, which would rather point to lower costs. A really striking difference is the "high intensity" of U.S.A. medicine, illustrated by the rate of inpatient surgical procedures which is for the U.S.A.

about twice as high as for the Netherlands (table 4.3). Studies of pacemaker implantation (Greenspan et al., 1988), endoscopy (Kahn et al., 1988) and carotid endarectomy (Winslow et al., 1988) indicate that overtreatment of insured patients is not exceptional in the United States. The costs of 'defensive medicine' in the U.S.A. were estimated to be $ 15 billion US $ in 1989 (Center for Health Policy Research, 1990).

This style of practising medicine may also be related to the competition between U.S.A.-hospitals concentrating on "quality" rather than on price, resulting in the duplication of expensive equipment and services. Robinson has shown convincingly that in the U.S.A. the amount of medical equipment within each hospital increases with the number of other hospitals in the neighbourhood (Robinson and Luft, 1985). In the Netherlands regional dispersion of expensive services is regulated by licensing, which reduces the inefficient use of medical equipment.

Comparing prices of health care services between countries is not without problems, but an OECD study indicates that for the equivalent of one US $ significantly more medical services can be bought in other OECD countries than in the U.S.A. (OECD, 1987). Viewing the mentioned data on volumes and prices, it is not surprising that average earnings of physicians in the U.S.A. are three times as high as in Sweden and twice the Dutch level (OECD, 1991; Jendteg, 1991) (table 4.3). The formidable liability insurance premiums for US-physicians, which have grown by about 15% per year in recent years (Centre for Health Policy Research, 1990), contribute to the rise in health care costs, although the physician's net disposable income is not affected.

Demography and costs

For the years 2010 and 2030, we calculated the consequences for health care costs of expected demographic developments in the Netherlands, using the estimated average costs per person by age and sex for 1988. As presented in table 4.5, total costs will rise by 1%point per year during 1988-2030. Seventy percent of this increase (0.7%point), is due to ageing, the remaining 0.3% being a result of population growth. During 2010-2030 ageing is responsible for 90 percent of the increase in costs: 0.9% per year. For the period 1980-2040, Rice estimated U.S.A. health care costs to grow 1.1% per year, of which only 0.3% caused by ageing (Rice, 1989).

Table 4.5 Indices for predicted of health care costs by disease, due to demography in the Netherlands, for 2010 and 2030, 1988=100.

Disease category	Year 2010	2030
All diseases	121	139
Prostate cancer	141	206
Dementia	148	203
Heart failure	143	196
Stroke	140	193
Lung cancer	142	185
Male genital diseases	136	181
Diabetes	134	174
Ischaemic heart diseases	141	174

Diseases for which costs are expected to rise faster than average are shown in table 4.5. The forecasts for these diseases are quite reliable, as they depend largely on the life expectancy of the living population and for most of these diseases the pattern of care and costs is not expected to change radically. Costs of dementia, lung and prostate cancer, stroke and heart failure will rise substantially up to 2010, but will increase more rapidly still in the decades thereafter, as the Dutch postwar baby boom reaches the age of 70 at which age medical consumption will accelerate (figure 4.2).

The 0.7%point annual growth of costs due to ageing is quite modest when compared with expected long term economic growth: 2.8%. Other forces than ageing seem to be more important in raising health care costs. For example, in the majority of OECD-countries, including the U.S.A., health care prices *structurally* tend to rise faster than in the rest of the economy (OECD, 1991; Baumol, 1985; CPB, 1987). In part this is due to lower productivity growth in health care than in manufacturing industries, but it also may reflect improved quality of medical services. This relative price increase of health care services acts as a continuous upward pressure on health care costs. During 1980-1990 health care prices in the U.S.A. rose 2.75% more than general inflation, in Sweden 1.0%, whereas in the Netherlands no difference occurred (OECD, 1991).

Discussion

The international comparability of disease specific data has been questioned recently. Stehbens demonstrated that diagnostic errors in certified causes of death may result in unreliable mortality statistics, invoking wrong conclusions on epidemiologic trends (Stehbens, 1987). It is not known whether the misclassification bias as reported for causes of death is equally important for diagnosis related to medical consumption. However, the international comparison presented is limited to very broad disease categories: the 17 ICD-chapters. Although differences in case-mix between countries *within* ICD-chapters may be substantial, misclassification *between* ICD-chapters appears to be less likely.

We used only the primary diagnosis of patients to assign costs to disease. This facilitates the comparison of results between diseases, a frequent problem in cost of illness studies, but it may underestimate costs of diseases which often prevail as comorbidity, such as diabetes. Reliable data on comorbidity and its influence on medical consumption may improve the cost estimates, especially for elderly people, but at the moment these are not available.

If disease specific data on pharmaceuticals become available for the Netherlands, it is unlikely that the ranking of diseases by costs will change substantially, because drugs only represent 9% of total health care costs in the Netherlands. Lindgren's data for Sweden indicate that, except for mental diseases, the ranking of total costs and costs of drugs by disease is quite similar (Lindgren, 1990).

The most recent cost estimates for the U.S.A. date from 1980, but may still be valid, as comparison of the distribution of U.S.A. costs by disease for 1980 and 1972 also showed only minor differences (Rice et al., 1985; Cooper and Rice, 1976). This also holds for the Swedish results regarding 1975 and 1983 (Lindgren, 1990 and 1981).

Although we do not provide a complete analysis of international differences in health care costs, numerous findings point to more aggressive medical practice, higher administration costs and higher prices as the main causes of unparalleled costs in the U.S.A. Further research is urgently needed to shed more light on the contribution of epidemiological causes. However, the high costs in the U.S.A. do not seem to produce better health care and public health for the *average American citizen*. In the U.S.A. the amount of care in psychiatric hospitals and nursing homes is substantially smaller than in the Netherlands and Sweden (United Nations, 1987) (table 4.3). Life expectancy at birth is lower in the U.S.A. than in the Netherlands and Sweden, while infant mortality is higher (WHO, 1991). This may be closely related to the fact that about 14% of Americans are uninsured (Congressional Budget Office, 1991).

In the Netherlands the number of uninsured is negligible (0,3-0,5%) and Sweden has a national health care system ensuring universal access to health care. Preventive care in the U.S.A. is underdeveloped, resulting in, for example, low immunization rates for infants (Starfield, 1991; Children's Defend Fund, 1987).

With respect to the impact of demographic change on costs we assumed that the relation between age, sex and health care costs for 1988 is a valid estimate for the future. In the absence of reliable evidence of the influence of epidemiology and technology on future medical consumption this estimate provides a reasonable baseline scenario, to be modified after more detailed research. If, for example, the health care costs for elderly will increase relatively more than for the younger, our calculations may underestimate the true costs of ageing. Further research on the development of the relation between age, morbidity and health care costs is needed to improve these estimates.

5 Current and future costs of cancer[1]

Summary

Cancer costs in the Netherlands amounted to 4.8% of health care costs in 1988. For five cancer types, and a sixth group, covering all other malignancies, costs were broken down by age, sex and disease phase, showing a remarkably similar pattern of medical consumption. Costs were linked to observed incidence and mortality and estimated prevalence, together allowing for prediction of future costs of cancer. In 2020, as a result of ageing, cancer costs will have increased much more rapidly than total health care costs, in particular for cancer of the lung and prostate. Colorectal cancer costs were predicted for epidemiological scenarios. Our model shows that an increase in future prevalence may bear quite different cost implications. If it is due to higher incidence, the costs will increase substantially. If due to survival improvement, the increase will be less prominent. Simply extrapolating costs based on future prevalence or mortality may produce serious errors.

Introduction

An optimal allocation of health care resources requires insight in the epidemiology and costs of diseases. Regarding cancer, we need to know both the absolute costs and the relative costs as compared to other diseases and total health care. Furthermore, the impact of demography, changes in medical practice and epidemiology on future health care costs should be analyzed. In this chapter we estimate the costs for five cancer categories: cancer of the lung, breast, colorectum, prostate, stomach; and all other malignancies in a sixth category. These cancer categories were selected, based on their importance for mortality, morbidity and medical consumption.

First, we calculated the total costs per type of cancer, age-group and sex for 1988. Next, these costs will be assigned to three disease phases: the

---

[1] Chapter based on Koopmanschap et al., Eur J Cancer, 1994

year following incidence, the year preceding death and the period in between. The estimated costs per patient by disease phase will be combined with incidence, mortality and prevalence, as calculated by our cancer disease model. This allows for prediction of future cancer costs for several possible scenarios: a demographic scenario and scenarios concerning expected trends in incidence and survival. The costs predicted by the 'three phase model' will be compared with outcomes using simple extrapolations of average costs per patient.

Material and methods

Total costs of cancer

In a recent study we estimated total health care costs for the Netherlands in 1988, for the 6 cancer types mentioned, together with 42 other disease categories, covering the entire ICD 9-CM classification (see chapter 4). Total health care costs were allocated to diseases, age-groups and sex, using utilization of services or the number of patients. Data on in-hospital and nursing home care were extracted from national registers with (nearly) complete coverage. For general practitioner's care, physical therapy and district nursing we used data from recent, large surveys. These results were used to calculate the costs of inpatient hospital care and non-hospital care for cancer.

Costs of outpatient hospital care consist of radiotherapy, chemotherapy and follow-up. Costs of Dutch radiotherapy-centres were assigned to the six cancer types, using the number of new patients receiving radiotherapy by type of cancer (Health Council, 1984). The costs of cytostatics administered in hospitals (Nepharma, 1989) were assigned to the six cancer types, in proportion to the distribution of in-hospital costs. The costs for monitoring patients without symptoms after diagnosis and primary treatment, were calculated in detail for breast cancer (de Koning et al., 1991) and were extrapolated to other cancers (see below).

Costs by disease phase

Several studies (Baker et al., 1989 and Bried et al., 1989) indicated that over the course of cancer, the costs show two peaks. The first peak, due to diagnosis and primary therapy; the second peak, due to palliation of often

severe symptoms in advanced disease. In between, medical consumption is modest. Hence, we discerned three disease phases, inspired by Baker:
- the first year following incidence, as a result of diagnosis and primary treatment;
- the last year of life for people dying of cancer (only for people who survived the first year, who get recurrent or metastatic disease and do not die from other causes);
- the period in between: 'the intermediate phase', during which people are disease-free or having already diagnosed metastases (in these terms disease-free is a mixture of cured patients and patients who currently have no disease symptoms, but who will later get recurrent or metastatic disease);

Costs for patients dying within one year after incidence were assigned to the incidence phase.

The average costs per patient by type of cancer, age, sex and phase were estimated as follows. For the *intermediate phase* we used detailed estimates for the Netherlands of annual follow-up costs per patient for breast cancer (De Koning et al., 1991). For the other types of cancer the *relative* follow-up *costs* compared to breast cancer, calculated for 87,000 patients in the United States (Baker et al., 1989), were combined with the costs for breast cancer, yielding the annual follow-up costs per patient.

The hospital costs in the *incidence phase* were primarily based on the length of stay in 1988, by type of cancer, age and sex, according to nation wide Dutch hospital register (SIG, 1990a). The length of stay for patients with metastases as primary cause for hospital admission determined the in-hospital cost during *the last year of life*. Although the last year of life and the period of metastases are not equivalent, hospital stays concerning metastases are often concentrated in the last year of life (Baker et al., 1989, Riley et al., 1987 and Van Ballegooijen et al., 1992). This length of stay is assumed to be age- and sex specific, but not cancer specific, because if metastases are coded as primary cause of admission, the site of the primary tumour is seldomly known.

The average hospital admission rate was derived by comparing the number of model based incident cases and deaths for 1988 (correcting for patients dying within the first year after diagnosis) to the number of hospital stays registered in 1988, for each combination of cancer, age and sex. The ratio of hospital stays versus incidence plus deceased served as the average admission rate. The admission rate was assumed to be equal for the incidence phase and the last year of life, because the data did not allow for separate estimates (see discussion). Combination of the model-based incidence and mortality with length of stay and admission rate yielded the number of 'model-based hospital days', which corresponds closely to the number of registered hospital days per cancer type, age and sex. Combining the number of hospital days with costs per day and adding other hospital costs (for surgery, radiotherapy, chemotherapy etc.) and non-hospital costs generated the average costs per patient.

Disease model

Our cancer model calculated the number of patients in each disease phase, per cancer type, age and sex (Bonneux et al., 1994). The model is a deterministic state transition model: it divides the population in subpopulations, defined by states. Each model year, the population 'ages' one year and the states are updated. Transitions from one state to another are governed by probabilities, extracted from registries and literature. Transition probabilities depend only on the current state defined by age, sex, cancer type, and disease phase.

A demographic module reproduces the population forecasts for the Netherlands (CBS, 1989) and generates first incidents of cancer by age and sex similar to the 1989 national cancer incidence rates for the Netherlands (Netherlands Cancer Registry, 1993).

After incidence patients first run a risk of immediate death as a consequence of primary treatment. The survivors are divided into two groups: a fraction which is cured and a fraction which is not. The not cured will enter the intermediate ('disease-free') phase of the disease, and subsequently will die of the disease, provided they do not die from other causes first. In practice of course cured and disease-free patients cannot be distinguished, and both will incur follow up costs.

The cancer-specific mortality of the not-cured is modelled by applying a lognormal survival to the relative survival rates. Relative survival rates correct the observed mortality in the diseased population for the expected mortality in the reference population with the same age and sex (Rutqvist et al., 1984 and 1985). Together this implies that survival after diagnosis can be characterised by 4 parameters: a fraction dead after treatment, a fraction cured, and the geometrical mean (equalling the median survival time) and variance of the lognormal distribution. A mathematical description of the survival is in appendix 5.

The demographic module takes care of all other causes of death, corrected for the studied diseases by using cause elimination life table methodology. This other causes mortality probability is applied to all states in the model, under the assumption that it is independent. Prevalences are calculated is the result from incidence, disease specific mortality and all other causes mortality.

Surviving patients run a higher risk of a second cancer episode. Therefore they are again subjected to the cancer incidence of the reference population, multiplied by the relative risks observed in the Connecticut cancer register (Boice et al., 1986).

The model was calibrated on aggregate data of cancer survival from Dutch and Scandinavian cancer registries, corrected for other causes of death. The fraction dead after treatment is from the Dutch hospital register (SIG, 1992), the other three survival parameters were estimated by an iterative nonlinear least squares regression, weighted for the numbers of deaths. Residual errors were small.

Scenarios

We have illustrated the relevance of epidemiology for future cost estimates with four scenarios for colorectal cancer. Scenario 1 only incorporates the expected demographic development. In scenario 2 we have added an improving prognosis after diagnosis. In the Norwegian and Finnish registries a very significant improvement over time of 5 year relative survival rates has occurred between 1954 and 1985 of 2% per year (Cancer Registry Norway, 1980; Langmark and Hagen, 1990; Hakulinen et al., 1981; Finish Foundation on Cancer Research, 1989). We have extrapolated an average relative decrease of cancer specific mortality of 2 % per year to 2005. Five year survival rates thus increased from 44% (1985) to 55% (1995) to 66 % (2005).

Scenario 3 involves, in addition to demography, a 2% annual relative increase in the incidence of all stages for men, and 1% for women. The trend in incidence is assumed to be 2% higher than the observed mortality trend in most western countries and is primarily caused by earlier diagnosis and increased detection of slower growing tumours (Doll and Peto, 1981).

Scenario 4, in our opinion the most realistic one, incorporates demography, increasing incidence (2.5% per year for men, 1.4% for women) and improved survival. As a consequence, mortality rates for men decrease by 0.4% annually and by 1.1% for women, as has been observed in the Netherlands over the period 1978-1989 (SIG, 1992).

All costs are given in millions of Dutch guilders. In 1988, the exchange rate of one Dutch Guilder (dfl) was approximately 0.3 British Pound and 0.5 U.S. Dollar.

Results

Total costs

Total costs of malignant cancer in the Netherlands amounted to 1894 million Dutch guilders in 1988 (table 5.1), representing 4.8% of total health care costs. West-Germany and Sweden have a comparable share: 4% and 5.1% respectively (Henke and Behrens, 1986; Lindgren, 1990). For the United States it is somewhat higher: 6.2% (Rice et al., 1985). In-hospital care takes about 60% of total costs. Outpatient hospital costs consist of radiotherapy (200 mln), chemotherapy (200 mln) and follow-up costs (143 mln).

Table 5.1 Health care costs for cancer, by sector in millions of Dutch guilders (in parentheses as % of total cancer costs). The Netherlands 1988.

Health care sector	Costs for 1988 in mln dfl	
In-patient hospital care	1,148	(61%)
Out-patient hospital care	543	(29%)
Non-hospital care	203	(11%)
Total health care	1,894	(100%)

About Dfl 1,000 mln was assigned to the five specific cancers, leaving almost 900 mln for the other cancers (table 5.2). Lung cancer is in first place: it is responsible for 16% of the costs of cancer (300 mln), and is mainly caused by men. Colorectal and breast cancer each represent about 14% of the total cancer costs, whereas stomach cancer and prostate cancer cause about 5-6% of the total costs. This ranking corresponds to recent estimates for Texas (Williams and Begley, 1992). Hospital costs predominate and costs are fairly equal for both sexes.

Table 5.2 Health care costs per type of cancer, sex and health care sector, for six types of cancer in millions of Dutch guilders. The Netherlands 1988.

Type of cancer	Total costs	% for men	% in hospital
Lung cancer	300	83 %	92 %
Breast cancer	253	0 %	82 %
Colorectal cancer	250	45 %	88 %
Prostate cancer	110	100 %	87 %
Stomach cancer	99	59 %	84 %
Other cancers	882	46 %	92 %
All malignant cancers	1,894	49 %	89 %

* 83%: 83% of total costs are incurred for men, leaving 17% for women.

Length of stay and admission rate
Figure 5.1 shows the average length of hospital stay by age and sex for colorectal cancer for the incidence phase and the last year of life. Only the results for colorectal cancer are shown, but the general pattern applies to all types of cancer.

Until age 55, the length of stay (per stay) is stable and fairly equal for both sexes. With increasing age, hospital stays become longer for both sexes, but at a quicker pace for women. For each type of cancer, older women in the Netherlands stay significantly longer in hospital than men of the same age, irrespective of the disease phase. Subsequent analyses have shown that this holds for the majority of diseases in the Netherlands (SIG, 1990a). Older women are more often single than older men (41% versus 15% in 1987 (CBS, 1990), limiting the opportunities for home care, which may cause longer hospital stays.

The average admission rate for colorectal cancer patients is 1.8 until age 45, but diminishes considerably with increasing age (figure 5.2). For people of age 80 and older it becomes smaller than 1, implying that a substantial number of these patients are not hospitalized. Consequently, they consume relatively more nursing home care, district nursing care and general practitioner's services, which we took into account. The admission rate is hardly sex-specific.

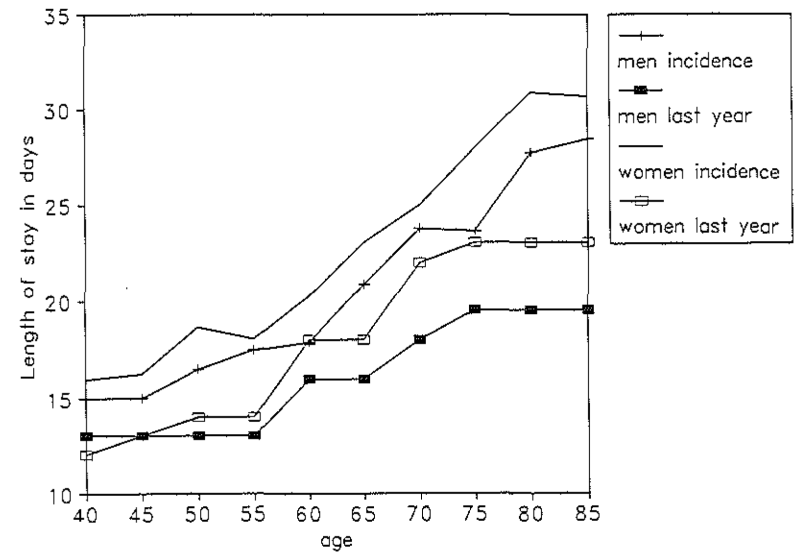

Figure 5.1 Average length of hospital stay in days for colorectal cancer, by age-group, sex and disease phase. The Netherlands 1988.

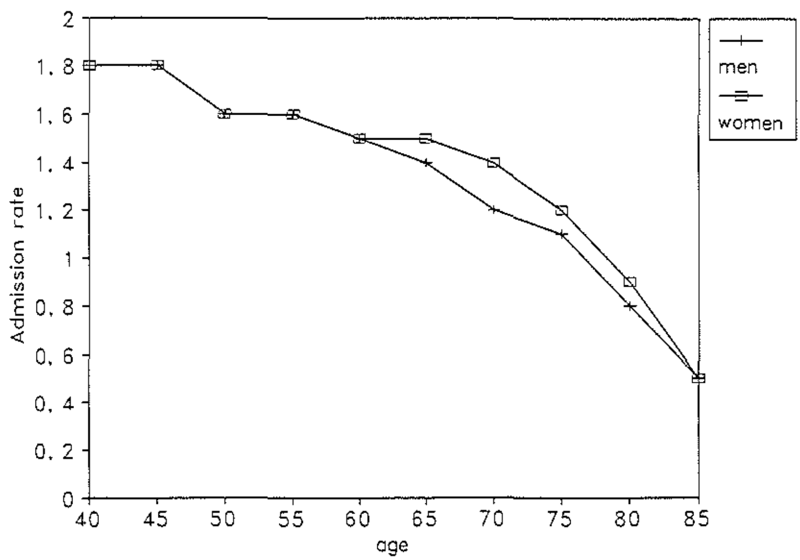

Figure 5.2 Average hospital admission rate per patient having colorectal cancer, by age-group and sex (model based outcomes). The Netherlands 1988.

Costs per disease phase

Figure 5.3 shows the average costs per patient having *colorectal cancer*, during the incidence phase and the last year of life respectively. Until age 55 the costs are stable, amounting to dfl 20-30,000 per patient. For women between age 60 and 75 they rise, due to the increase in length of stay which more than compensates for the falling admission rate. For people older than 80 the costs fall to 12-20,000, as the low admission rate then becomes the dominant factor. Relatively low costs for patients older than 80 years were also found by Riley and Lubitz (1989). During the incidence phase, costs (dfl 25-30,000) are higher than during the last year of life (20,000), irrespective of age and sex. For women over 60 the costs are about dfl 5,000 higher than for men of the same age.

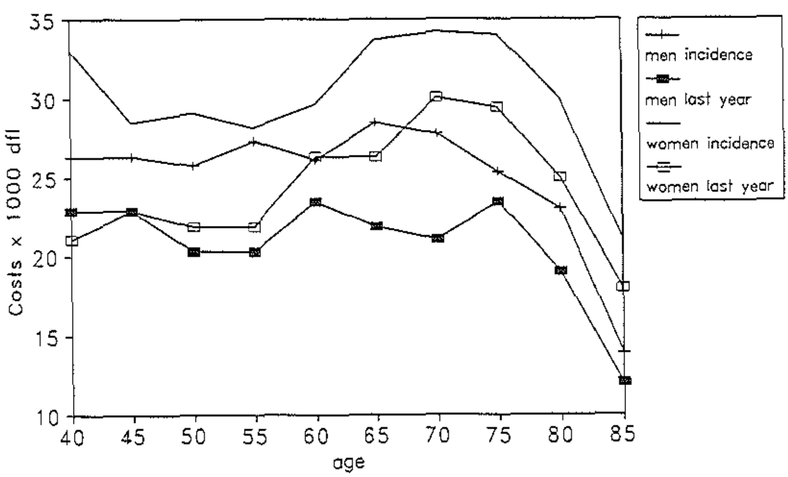

Figure 5.3 Average costs per patient having colorectal cancer, by disease phase, age-group and sex, in thousands of Dutch guilders (model based outcomes). The Netherlands 1988.

A considerable part of the results for colorectal cancer holds for all 6 cancer types, such as the lower costs for patients older than 75. The higher costs for women older than 60 apply to all 4 cancer types that are relevant for both sexes. Fairly stable costs until age 60 prevail for all cancers, except for 'other cancers', in which patients younger than 15 years incur considerably higher costs. This category is a mixture of very different types of cancer, both for young and for old patients. The high costs for people under 15 are mainly caused by leukaemia, brain cancer and non-Hodgkin's disease (Coebergh and van der Heijden 1991) for which therapy (immuno- and chemotherapy and bone marrow transplantation) is very expensive (OTA, 1981; Black, 1982; Office of Health Economics, 1980). The notable increase in costs between age 60 and 75 also applies to prostate cancer and lung cancer (women), but not to stomach cancer and other cancers. The incidence phase is not always more expensive than the last year of life. For breast-, prostate- and other cancers the last year of life appears to be more expensive. De Koning's study (de Koning et al., 1992) confirms this result for breast cancer.

During the incidence phase, the average costs per person are the highest for stomach cancer (dfl 30,000) and other cancers in younger people (30-50,000). Breast- and prostate cancer are relatively 'cheap' (21,000), leaving colorectal- and lung cancer in the middle-range (dfl 25-30,000). The variation in costs for the last year of life is much smaller, resembling Baker's findings for U.S.A. Medicare cancer patients (Baker et al., 1989), as well as Dutch results for breast and cervical cancer (van Ballegooijen et al., 1992; de Koning et al., 1992). We estimated these costs at about dfl 25,000. The only exception is 'other cancers', causing higher costs during the last year of life.

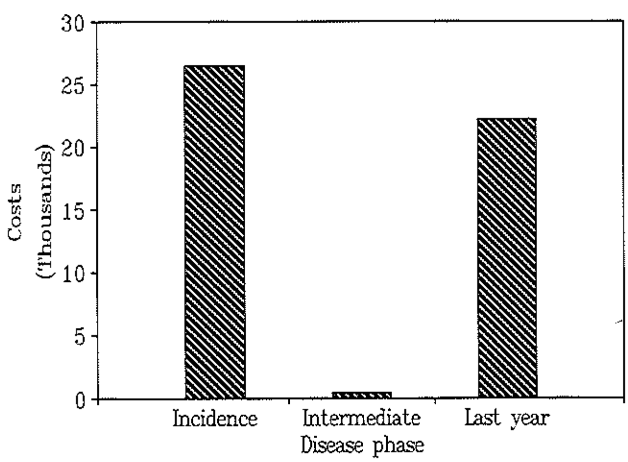

Figure 5.4 Average costs per patient by disease phase for colorectal cancer, in thousands of Dutch guilders (model based outcomes). The Netherlands 1988.

Costs over the course of cancer
Assembling the three disease phases illustrates the costs over the course of cancer (figure 5.4). The costs show two peaks, the first just after clinical detection, the second in the last year of life. In the intermediate phase, annual costs are very modest. The total costs of the intermediate phase depend chiefly on the length of disease-free survival. The total costs related to prevalence of cancer only become important if the intermediate phase is very long, say 15 years.

Table 5.3 Predicted costs per type of cancer for the Netherlands in 2005 and 2020, demographic scenario, in millions of guilders in 1988 prices, and as index compared to 1988 (1988=100).

Type of cancer	Costs in 2005		Costs in 2020	
Lung cancer	377	(126)	474	(158)
Breast cancer	311	(123)	352	(139)
Colorectal cancer	308	(123)	375	(150)
Prostate cancer	134	(123)	173	(157)
Stomach cancer	122	(123)	151	(150)
Other cancers	1,060	(120)	1,253	(142)
All cancers	2,312	(122)	2,778	(147)
All diseases	45,110	(116)	51,675	(130)

Demography and costs
Table 5.3 shows the predicted costs regarding all 6 cancer types, for the years 2005 and 2020, applying the demographic scenario. In 2005, the index for total cancer costs (in 1988 it was 100) is 122: a 1.2%point annual increase, comparable to total health care (index 116). Costs increase at about the same rate for each type of cancer. In 2020, the burden of ageing becomes heavier, as the Dutch postwar baby boom reaches ages 60-75. By then cancer costs (index 147) will have increased considerably more rapidly than total health care costs (index 130). The costs of lung- and prostate cancer will be most seriously affected by ageing (index 158 and 157 respectively). The consequences for breast cancer and other cancers are less severe as they

prevail at relatively young age. Still, their rate of cost increase is higher than for total health care.

Scenarios for colorectal cancer

The two-peaked cost pattern has important implications for future cancer costs. The results of four scenarios for colorectal cancer are presented for the year 2005 (table 5.4). According to our model, costs of colorectal cancer in 2005 amount to 308 million guilders, if only demography is taken into account. Improving disease-free survival (scenario 2), and consequently increasing the prevalence of cancer, will result in dfl 299 mln, since it only raises the number of patients consuming the relatively inexpensive follow-up (incidence remains the same) and cancer specific mortality declines, because of other causes mortality. On the other hand, if *the same increase in prevalence* of disease-free persons occurs, but due to a rise in incidence (scenario 3), the costs will rise substantially to dfl 395 mln. All extra patients will undergo expensive diagnosis and primary therapy. Furthermore, cancer mortality increases, causing extra costs as well. In case of the most realistic course of events (scenario 4), the costs will amount to dfl 417 mln.

Table 5.4 Predicted costs for colorectal cancer for the year 2005 applying 4 models to 4 scenarios, in millions of guilders. In parentheses: differences in % compared to outcomes of the 3-phase model.

	1988	2005 Scenario 1	2005 Scenario 2	2005 Scenario 3	2005 Scenario 4
Three phase model	250	308	299	395	417
No phase model prevalence based	250	309	358 (+20%)	358 (-9%)	413 (-1%)
No phase model mortality-based	250	310	218 (-27%)	401 (+2%)	279 (-33%)
No phase model incidence-based	250	290 (-6%)	290 (-3%)	380 (-4%)	460 (+10%)

Scenario 1: only demography; Scenario 2: demography and survival improvement; Scenario 3: demography and incidence increase; Scenario 4: demography, incidence increase and survival improvement; For exact details on the scenarios, see the material and methods.

Table 5.4 also presents the predicted costs using simple extrapolations without discerning disease phases. Three variants are shown, based on constant costs per prevalent case, per death and per incident case, respectively. In case of scenario 1, incorporating only demography, the predictions are quite accurate, although the incidence method understates the costs by 6%.

The *prevalence-based method* predicts the same costs for scenario 2 and 3 since either produces the same prevalence in 2005. In case of improved survival (scenario 2), this leads to overestimating the costs by 20%. If the incidence increases (scenario 3), the costs are underestimated by 9%. The *mortality-based approach* is quite accurate if the incidence rises (and consequently mortality) but fails completely in predicting the costs of survival improvement. For scenario 2 and 4 the costs are understated by 27-33%. The *incidence-based method* performs reasonably in case of scenario 1 to 3, but for scenario 4 costs are overestimated by 10%.

Discussion

Epidemiology and medical consumption vary considerably between types of cancer as well as between individual patients. Recognising this variability, our analysis shows that the general pattern of medical consumption coincides remarkably well for the six cancer types described. The length of hospital stay increases uniformly with age, especially for women. On the other hand, the admission rate falls with age. For patients older than 75, the second factor dominates, resulting in lower costs. Over the course of cancer, medical consumption shows two peaks: during the first year after incidence and in the last year of life. In between the costs are modest.

We assumed the same admission rate for patients in the incidence phase and the last year of life. The length of hospital stay for patients with metastases was assumed to be age- and sex specific, but equal for all cancer types, which is supported by research in the U.S.A. and the Netherlands. Both assumptions are obviously rather crude, but sufficient to demonstrate the divergent consequences of several scenarios. Sensitivity analysis by substantially varying the most uncertain parameter -the admission rate during the incidence phase versus the last year of life- proved not to affect the basic two-peaked pattern of costs seriously. Consequently, the predictions of future costs for the scenarios described are robust.

The costs during the prevalence phase were based on Dutch calculations for breast cancer, combined with Baker's relative estimates for other types of cancer. As the costs are relatively low, errors in these estimates will not influence the results seriously.

The assumed survival improvement for colorectal cancer (scenario 2 and 4) seems high, but is the results of two processes: a true decrease of mortality due to more effective treatment and a spurious increase of survival due to earlier diagnosis (lead time) and increased detection of slow growing tumours which had passed unnoticed before (length time and pseudo-diagnosis) (Doll and Peto, 1981; Cole and Morrison, 1980).

If one is only interested in the costs of demographic scenarios, simple extrapolations of current costs are satisfactory, provided that reliable information on costs by type of cancer, age and sex is available. However, it is highly plausible that in the future, trends in incidence and survival will prevail, as they have over the past 20 years. Although our three-phase model is still somewhat crude, it takes these dynamics into account and thus provides useful predictions of the costs of epidemiological scenarios.

The demographic scenario, as described, is quite restrictive, since it does not account for future trends in risk factor exposure, like smoking for lung cancer and the influence of screening on breast cancer. Further refinement of the disease model and cost estimates is underway, in order to predict the influence of these trends on future morbidity, mortality and costs.

Part III Indirect costs of disease

6 Indirect costs in economic studies: confronting the confusion[1]

Summary

Indirect costs of disease often constitute a substantial part of estimated costs or savings in economic evaluations of health care programs. The human capital approach is almost unanimously used for estimating indirect costs, defined as production loss due to disease, although a growing number of authors questions its validity. This chapter discusses the relevance of indirect cost estimates for health policy and reviews the current empirical and methodological literature on this issue. It describes several important issues and controversies regarding indirect costs, such as the consequences of short term absence from work for productivity, reduced productivity without absence form work, the influence of unemployment on production loss, the relation between health effects and indirect costs and the possible medium term macro-economic consequences of absence from work and disability.

It concludes that indirect costs are relevant for health policy, provided that the estimates of indirect costs reflect the real changes in production due to disease, including the production of unpaid labour. Future research should focus on attaining these estimates. Indirect costs in economic evaluations should preferably be presented separately from direct costs, health effects and other study outcomes.

Introduction

In economic studies often a distinction is often made between direct and indirect costs (or indirect benefits in the case of savings), and sometimes the latter category is further divided into indirect costs within and outside the health care sector. An example of indirect costs within health care is medical consumption during prolonged period of life due to life-saving interventions.

[1] Chapter based on Koopmanschap and Rutten, PharmacoEconomics, 1993

We consider here only indirect costs outside the health care sector, also referred to as societal production losses as a consequence of illness (including its treatment), disability or death. This definition explicitly excludes the valuation of life *per sé*, as applied in cost-benefit analysis; we consider only the changes in the individual's contribution to societal production. Some authors incorrectly describe direct non-medical costs, such as costs to the family and patient transportation, as 'indirect costs'. Sometimes the nonmonetary costs of pain, suffering, discomfort etc. are also included in the concept of indirect costs. It is more appropriate to consider these items as health effects and to quantify them in terms of reduced quality of life. The latter procedure reduces the risk of double counting (both as costs and health effects).

In sum we take an extra-welfarist position (Culyer, 1991), suggesting that an economic appraisal assesses the contribution of the intervention in terms of health gains, however measured, and in terms of the opportunity costs to society of performing a health care intervention or implementing a health care program. Clearly, these opportunity costs include changes in the production of those individuals who undergo such intervention or participate in such program.

Indirect costs appear in cost-of-illness studies as well as in economic appraisals of interventions, and the appropriateness of their appearance in the latter type of studies is often debated. This is illustrated by the fact, that the Ontario guidelines for economic appraisal in support of reimbursement decisions regarding pharmaceuticals do not object to including indirect costs in economic evaluation studies (Detsky, 1993; Ontario Ministry of Health, 1991), whereas the earlier version of the Australian guidelines stated that indirect costs should not be included in studies submitted to the Australian pharmaceutical benefits advisory committee (Commonwealth of Australia, 1990). The most recent version of the Australian guidelines suggests that economic appraisals of pharmaceuticals should present results both with and without indirect benefits and costs included, but stresses that valuation of work time gained or lost should be made explicit (Henry, 1992).

Part of the confusion on the proper place of indirect costs in economic studies is related to uncertainty about the appropriate methodology to estimate their magnitude accurately and to the fact that, in many studies to date their magnitude has been grossly overestimated. This chapter reviews the current status of estimating indirect costs in economic studies, to clarify some methodological and policy-related issues and to suggest how the issue of indirect costs should be properly addressed.

A literature review

Before discussing methodological issues, it seems useful to see how economic evaluations practically deal with production losses. The MEDLINE literature retrieval system was scanned for studies considering indirect costs of disease, published between 1987 and 1992. We identified 49 articles related to indirect costs, of which 23 papers have appeared since 1991, mirroring the growing popularity of economic evaluations (table 6.1 and appendix 6). The majority has been written by U.S.A.-authors, whereas in Europe the Swedish have published many articles.

Cost-of-illness studies, followed by cost-effectiveness analyses, are most likely to include indirect cost estimates. Almost all disease categories are being addressed. Ten studies concern health care interventions related to drugs, 7 of these being complete economic evaluations (cost-effectiveness/cost-utility analysis or cost-benefit analysis).

Virtually all studies used the human capital approach, which estimates the value of potentially lost production (or potential lost income) as a consequence of disease. We use the world potentially because in case of permanent or premature death at a specific age, the total productive value (or income) from that age until the age of retirement is counted as indirect costs (McGuire et al., 1988). In several articles the methods used were not stated clearly. As a rule, costs of absence from work are taken into account, just like mortality costs, if relevant. It is often not clear whether long term disability was considered. Production losses due to unpaid labour are valued in only 14 of the articles, mostly at market prices.

Only 6 articles, mostly European, point at the problems associated with calculating indirect costs or criticize the use of the human capital method. Two authors stress that estimating production losses should not be confused with valuing lives (Fitzgerald and Gafni, 1990; Levin and Jonsson, 1992). Three authors point to the fact that the human capital method will overestimate indirect costs for patient groups in an economy with less than full unemployment (Levin and Jönsson, 1992 ; Gerard et al., 1989; Olsson et al., 1987). The influence of indirect costs on results was substantial. On average, indirect costs represented 52% of total disease costs or total costs saved by health care interventions.

Table 6.1 Overview of empirical studies concerning indirect costs, extracted from the Medline literature retrieval system and published during 1987-1992.

Year	1987	1988	1989	1990	1991	1992
Number of studies	5	7	5	8	12	11

Country	United States	Sweden	United Kingdom	Canada	Other countries
Number of studies	27	6	2	2	12

Health care sector / Study category	Total health care	Drugs	Prevention	Orthopaedics	Other	Total
Cost of illness	23	0	1	0	3	27
CEA/CUA	0	6	2	2	2	12
CBA	0	1	1	0	1	3
Cost analysis	0	3	1	0	0	4
Other	1	0	0	0	2	3
Total	24	10	5	2	8	49

Some controversies

Despite the widespread use of indirect costs in economic evaluations and the predominance of the human capital approach, many authors of methodological papers feel uncomfortable with the concept. Some have philosophical objections against using indirect costs. Including of production losses may favour health care interventions directed to (well paid) working people, which may conflict with equity considerations. This is because, on average, nonworking people already have a lower health status. This is true especially if the results of economic studies are used to support decisions at the level of individual patient treatment.

Others criticize the way in which indirect costs are estimated. They state that the human capital method estimates the value of *potential* production lost, whereas the *actual* loss for society may be much smaller. Drummond commented in this respect: " for short term absences, a given person's work may be covered by others or made up by the sick person on his return to work. For long term absences, an individual's work can be covered by someone drawn from the ranks of the unemployed. Therefore, while absence from work may cost the individual, or that person's employer, it may not cost society very much" (Drummond, 1992).

Production losses are not restricted to absenteeism. For several diseases, sick people may continue working at a slower pace. Thus far, only a few studies tried to estimate the extent of the production losses involved in such case. These studies concern the cost of migraine (Osterhaus et al., 1992) and the costs and effects of using cimetidine for peptic ulcer (BIPE, 1978).

If the societal perspective is used, estimates of indirect costs should incorporate production losses of unpaid labour. Nevertheless, such costs are often not taken into account, which may lead to an undesirable bias in health policy in favour of preventing and fighting diseases which occur predominantly in men.

Finally, it is sometimes suggested that including quality of life and indirect costs in economic evaluations may lead to double counting. Russell stated: 'Estimates of earnings help to describe the consequences of better health in greater detail, but they are not an addition to the health effects. To include them as a separate item is to count some of the effects twice, once as a gain in health and again as a saving to be subtracted from program costs' (Russell, 1986).

Why and how to include indirect costs

The philosophical objections could easily be met by excluding indirect costs from economic evaluations. However, this would deny a significant part of economic reality, since production losses do contribute to the scarcity of resources and hence decrease the wealth of society. It seems preferable to insist that economists estimate indirect costs as realistically and trustworthily as possible, incorporating production losses in relation to unpaid labour.

Indirect costs may be presented together with, but separate from, health care costs, health effects, consequences for equity and ethical and legal considerations. It is the responsibility of health care decision makers to decide on the relative weight that they want to attach to indirect costs, compared with other costs.

Short term absence

Regarding short term absence from employment, Drummond's opinion has not provoked much discussion among health economists yet, but it seems to have inspired health authorities. Australian guidelines for economic evaluations in health care that were issued in 1990 in the context of decisions on reimbursement of pharmaceutical products denied any production loss due to short term absence: first, because production will be made up on the return to work, and secondly because employers usually have excess capacity in the labour force to cover absenteeism (Commonwealth of Australia, 1990). This line of reasoning was obviously an oversimplification. We agree with Drummond that 'the real position is likely to vary from case to case. An approach that ascribes either full production losses or zero losses in every case is likely to be incorrect' (Drummond, 1991). In addition, if the excess labour force within a company is structural, part of these excess labour costs may be the result of absenteeism.

The exact relation between short term absence and production loss depends on the employee's profession, the type of organisation and the production process. For example, a bus driver must be replaced in order to deliver the transportation services, whereas a sick health economist may make up his or her work later.

Complete insight into the consequences of short term absence for indirect costs requires detailed study at company level. A more recent version of the Australian guidelines takes these objections into account and suggests that study results should be presented with and without indirect benefits and costs included. In addition, evidence is required that indirect benefits are really achieved through introduction of the drug (Henry, 1992). This requires detailed analysis of the relationship between illness and productivity, preferably at the level of the employer.

Generally, absence from work affects the effective labour time of workers. Numerous studies demonstrated that a reduction of annual labour time causes a less than proportional decrease in labour productivity per year.

The elasticity found for some large companies in the Netherlands ranges from 0.6 to 0.9. (de Koning and Tuyl, 1984). In the absence of specific data, these estimates of the relationship between labour time and productivity may be usefully employed to deal with production losses due to short term absence (see chapter 8).

Long term absence

With respect to long term absence, many health economists agree on the mitigating effect of unemployment on production losses due to long term absence (Drummond et al., 1987; Gerard, 1989; Lindgren, 1981; Williams, 1985). They restrict the argument to substantial, persistent unemployment. If unemployment is less than frictional, workers may not be replaced easily and production losses would approach the human capital estimates. Frictional unemployment is an inevitable part of unemployment, since filling vacancies takes time and some qualitative discrepancies between labour demand and supply always prevail. Generally, frictional unemployment is assumed to be 1-2% of the labour supply.

When authors recognize the influence of unemployment on production losses, they tend to incorporate this relationship rather crudely in their analysis. Thus, if unemployment is assumed to be less than frictional unemployment, all absence is counted as production loss, whereas if unemployment is substantial, the loss is assumed to be nihil. It seems more realistic to assume that production loss always occurs to some extent, but that the amount depends on the time span needed to replace a sick worker, or to reorganise the production process otherwise. The length of this time span depends on the availability of qualified personnel within firms and on the labour market. During this adaptation period the production falls, or the production remains equal at extra costs. Consequently, the indirect costs are proportional to the length of the adaptation period.

In our view, a suitable proxy for the length of this period is the duration of vacancies, since it is sensitive both to the unemployment rate and to the efficiency of the labour market (van Ours and Ridder, 1991). The use of this concept in estimating indirect costs of disease, the so-called "friction cost method", was introduced recently (Koopmanschap and van Ineveld, 1992). In cases of very low unemployment, the friction cost method is less obvious applicable, since it may underestimate the true production losses. On the other hand, even then production losses will not last forever, so the human capital approach will still overestimate indirect costs.

Production loss without absence

Production losses that may occur when an individual is not absent from work are even more difficult to measure. Data have to be collected by questioning individuals on their remaining productivity during sickness, and the opportunities to make up later for the productivity lost or to reorganise parts of the work. With respect to Osterhaus and colleagues' study (Osterhaus et al., 1992), Drummond (1992) questioned the reliability of asking a person to recall reduced productivity over the last year, and the associated heavy reliance on subjective assessments. These objections may be met partially by using diary methods or recalling over a short period, for example the past two weeks.

Unpaid work

With respect to unpaid work, data on work absence and disability are very scarce. At the moment, the best solution appears to use data from time surveys, combined with specific patient questionnaires. In addition, it is important to know what happens if an unpaid worker gets ill. As for paid labour, short term illness may also lead to postponement of work, which can be made up later. If the illness takes longer, the work may be taken over by professional workers, for example home helps, and/or by the patient's community (family, friends, neighbours etc.).

Professional replacement can be valued at market prices. Informal care supplied at the expense of paid labour may be valued by the workers labour costs. If supplying informal care reduces leisure time, the valuation should reflect the opportunity costs of leisure time. The latter costs are difficult to measure properly, and the possible *positive utility* (i.e. the satisfaction of the person who supplies informal care) of taking care of a partner, relative or friend should be taken into account too (Wright, 1987).

Double counting

Russell's fear of double counting of indirect cost and health effects (Russell, 1986) raises the question of the impact of disease on health and on the economy, respectively. The impact on health and that on production are to a high degree distinct consequences of health care programs requiring separate measurement. However, if income from labour is (among other things) determining health status, the influence on health and on the economy become mutually dependent and the danger of some double counting arises. It may be possible to disentangle these effects. In quality-of-life studies, respondents may be asked to value the effect on their quality of life of being able to work, apart from the value of extra earnings. According to Jönsson and colleagues: 'the existence of health insurance and cash benefits, ... make(s) it reasonable to assume that respondents take only limited account of the effect of health status on their income when asked to state their preferences for different health states' (Jönsson et al., 1991). This argument seems rather convincing for developed countries. However, in less developed countries where people are lacking adequate health insurance, labour income may be indispensable to the enjoyment of any quality of life, bringing about Russell's problem of double counting again.

Macro-economic consequences

Absence from work and disability may influence the national economy in several ways ('the macro-economic consequences of disease'), in addition to the direct consequences for production, as described above. Absence from work, by influencing labour productivity, directly affects the labour costs per unit of output, an important determinant of competitiveness on the world market. This may affect the level of production of the sector exposed to international competition and consequently the national income of a country, which is an important indicator of a nation's wealth.

Reducing the incidence of disability will raise labour supply and unemployment, lowering labour costs, which again improves the production of the sector exposed to international competition and increases national income. Although disability benefits do not represent indirect costs, because they are income transfers, they may have an economic impact. The level of insurance premiums paid influences the labour costs per unit of output, affecting the performance of the exposed sector of the economy, as described above. Insurance premiums may also affect the net disposable income of workers, influencing private consumption and domestic production.

These macro-economic consequences are not simple to quantify, because several processes mutually interact, each having a different time lag. For example, a change in disposable income almost instantaneously influences private consumption, whereas the effect of labour costs on exports takes much longer. For a correct estimation of these macro-economic consequences a macro-economic model is indispensable (see chapter 8).

Indirect costs and health policy

Depending on the level and context of decision making and on the associated perspective on costs, policy makers are more or less inclined to take indirect costs in consideration when choosing between alternative strategies. But even when the societal perspective is regarded as the relevant one, there is still reluctance to accept indirect costs as a relevant factor in the decision-making process. Hence, there is still the challenge for health economists to clarify the concept of indirect costs and to present accurate and reliable estimates of their magnitude in economic studies.

Important in this respect is the fact that estimates of indirect costs of disease in cost of illness studies mostly refer to the *total costs* of absence from work, disability and mortality. These total costs of disease represent the possible savings of eradicating disease, which comes down to estimating 'the benefits of the unattainable' (Shiell et al., 1987). For health policy purposes, the consequences of *attainable changes* in absence from work and disability are highly relevant. In this respect, we need reliable estimates of the possible change in absence, disability and mortality, as a result of health care interventions. This again illustrates the greater relevance of cost-effectiveness analysis compared to cost of illness studies (Drummond, 1992).

The estimates of indirect costs should preferably take into account the possible competition between diseases. If for example, absence due to migraine can be curbed, without simultaneously reduction of workplace stress (one of the possible underlying causes), absence due to other diseases related to stress may increase, offsetting part of the estimated cost savings.

The agenda for future research

In setting an agenda for future research concerning indirect costs of disease, the following issues may be considered. First, we observed that production losses may occur as a consequence of lost opportunities for both paid and unpaid labour. As the literature review revealed very little research on the value of unpaid labour, a first point on the research agenda may be the estimation of the amount and value of production losses due to unpaid labour.

A number of issues discussed led us to suggest the use of another concept in estimating indirect costs of disease, the so-called 'friction cost method', which seeks to estimate the real amount of indirect costs attributable to disease, taking into account the situation within firms and on the labour market. If this approach is to be followed, a number of questions need to be addressed: (a) what is the frequency of friction periods (dependent on institutional arrangements)? (b) what is the average length of the friction period (dependent on age, sex, and education level); and (c) what is the relationship between the valuation of lost production and the length of the friction period? Such an approach would take explicit account of the situation on the labour market, which will influence the length of the friction period (see chapter 7 and 8).

Further analysis of the consequences of short term absence from work for output, productivity and costs is needed, as well as a study of the extent of production losses that may occur during illness without absence from work. And finally, the macro economic consequences of absence from work should be quantified, taking into account the various interactions described.

Each of these topics would require significant research effort. We believe that such research effort is required because indirect costs will become accepted as an integral part of economic appraisal, both by the scientific community and by health policy makers, only when they can be estimated in a valid and reliable manner.

7 Towards a new approach for estimating indirect costs of disease[1]

Summary

Many researchers in the field of evaluation of health care doubt the usefulness of estimates of indirect costs of disease in setting priorities in health care. This paper attempts to meet part of the criticism on the concept of indirect costs, which are defined as the value of production lost to society due to disease. Thus far in cost of illness studies and cost-effectiveness analyses the *potential* indirect costs of disease were calculated. In the following a first step will be taken towards a new method for estimating indirect costs which are expected to be effectuated in reality: the friction cost method. This method explicitly takes into account short and long run processes in the economy which reduce the production losses substantially as compared with the potential losses.

According to this method production losses will be confined to the period needed to replace a sick worker: the so called friction period. The length of this period and the resulting indirect costs depend on the situation on the labour market. Some preliminary results are presented for the indirect costs of the incidence of cardiovascular disease in the Netherlands for 1988, both for the friction costs and the potential costs.

The proposed methodology for estimating indirect costs is promising, but needs further development. The consequences of illness in people without a paid job need to be incorporated in the analysis and also the relation between internal labour reserves and costs of disease should be further investigated. In addition, more refined labour market assumptions, allowing for diverging situations on different segments of the labour market are necessary.

[1] Chapter based on Koopmanschap and van Ineveld, Soc Sci Med, 1992

Introduction

In most cost of illness-studies and in some cost-effectiveness analyses the indirect costs of disease are considered and estimated. Not all researchers are convinced that indirect cost estimates contribute to the economic evaluation of health care programs. In this paper a first step will be taken in developing an alternative method of estimating indirect costs, which meets an important part of the criticism. The conceptual elements of this method will be described in detail. Some preliminary results will be presented for the indirect costs of the incidence of cardiovascular disease in the Netherlands for 1988. A comparison will be made with the traditional method of estimating indirect costs based on the human capital approach.

The concept of indirect costs

Indirect costs of disease can be defined in several ways, depending on the perspective and objective of analysis. In this paper the indirect costs of disease will be defined as the value of production lost to society due to illness, with respect to paid labour as well as unpaid labour. While defined this way, this paper deals only with paid labour.
Production losses can be the consequence of:
- temporary absence from work;
- disability;
- premature mortality.

Basically, production losses can even be relevant without absence from work if sick people continue working at a suboptimal level. Information on this topic is unfortunately not available, so it won't be analyzed in this paper. The specification excludes the "costs" of pain, suffering, discomfort etc. Next to the formidable problem of valuing these items in monetary terms, it seems more appropriate to consider these psychological factors as effect variables and to quantify the consequences in terms of reduced quality of life. This procedure will also avoid possible double counting.

As a consequence we will focus on the "pure" economic consequences of disease. Furthermore it should be stressed that we don't seek to assign a value to human life per se, as in some cost-benefit analyses (Schoenbaum et al., 1976). This definition of indirect costs is therefore compatible with social cost-effectiveness analyses and cost-utility analyses.

Almost all studies estimating indirect costs use the human capital approach. The way in which the human capital approach is usually being applied leads to an estimate of the value of *potentially* lost production (or the potentially lost earnings) as a consequence of disease. We use the word potentially because in case of permanent disability or premature mortality at a specific age the total productive value (or earnings) from that age until the age of retirement is counted as indirect costs.

This concept of indirect costs in the economic appraisal in health care is not uncontroversial (Drummond et al., 1986 and 1987; McGuire et al., 1988; Hartunian et al., 1981). The critical comments focus on four topics:
1) earnings may be a poor estimate for the amount of production forgone to society. A better, though not perfect, estimate used in the rest of this paper is labour costs.
2) health care programs directed to working people will always generate more gains in production than programs for mentally handicapped or elderly. This could lead to priority setting enlarging existing inequalities in health, which may conflict with policy objectives.
3) sometimes it is not clear if costs are direct or indirect and this may cause double counting of costs.
4) the real production losses can be much smaller than the potential losses because sick people can often be replaced at little cost.

Notwithstanding this critique the fact that production losses due to illness (and production gains due to health care) influence the scarcity of resources and therefore the wealth of society is not disputed. In this respect there exists no difference in assessing direct and indirect costs; both should be incorporated in economic evaluations of health care programs.

The first and fourth criticism don't overthrow the concept of indirect costs but merely stress the importance of using estimates of the production losses which are expected to be effectuated in reality.

Addressing the first point of critique, an example in which labour costs are a poor estimate for the value of production is unpaid labour in the household. Reliable estimates of production losses should therefore include the production involved in this type of labour.

With respect to the second critical comment, the gain in society's wealth will indeed be larger for programs directed to economically active people. Explicit estimates of these gains will be more helpful in decision making than using implicit notions. Next to this, estimates of indirect costs should never be weighed in isolation, but always as a part of the political decision making process, in conjunction with other criteria like direct costs and equity. In addition, including indirect costs has implications for equity.

If a health care service is actually implemented estimates of production gains can be useful in the discussion about how the program needs to be financed.

An example of the third comment: if a sick person is a day absent of work and needs to visit a physician one could approximate the value of the time spent on the visit by the value of production lost in this time-span. If one also calculates the production lost due to a day of absence from work, the sum of direct and indirect costs is clearly an overstatement. We think this problem can be avoided by continuous awareness of the possibility of simultaneous sickness absence and medical consumption.

The last item can be met when taking into account the economic processes which make that sick people can be replaced, mitigating the amount of production losses. Several examples can be given in which the expected production loss due to illness may be much smaller than the potential loss.

- If unemployment, registered as well as hidden unemployment, is well beyond the level of inevitable frictional unemployment, sick persons can be replaced by unemployed.
- If labour reserves exist within firms with sick employees the loss in production may be negligible.
- Non-urgent work may be cancelled or postponed without consequences for the level of production.
- If replacement is only feasible by persons already working, a reallocation of employees over jobs can induce a smaller production loss, resulting from a vacant job at a lower productivity.

The underlying notion is that processes within firms and on the labour market will lead to replacement of sick persons, after a period of adaptation, reducing the production loss substantially. Unemployment is important in this respect, but it is not the only factor causing the difference between potential and real production losses. The above mentioned processes are effective in the short run when the organisation of the production and the labour market conditions are merely fixed constraints.

In the longer run unemployment may fall substantially or enhanced efficiency within firms may reduce internal labour reserves, both inducing higher production losses. However, in the longer run labour shortages and production losses will be mitigated by increasing possibilities for restructuring the organization of the production, by reallocation of employees and as a result of laboursaving investments. To illustrate this process the analysis should preferably be linked to a macro-economic model incorporating vintages of capital goods in relation to the relative costs of capital and labour.

In the following the concept of indirect costs will be modified, setting a first step towards deriving more realistic estimates of production losses due to disease.

Operationalisation

Disease may cause losses in production, but in general this loss will be confined to a period needed to adapt to the changed situation. This period will be called the friction period. To clarify this notion further, it may be helpful to consider the possible economic consequences of disease on the micro-economic level. Due to illness, there are three possible outcomes for production and costs during the friction period.

- Production falls;
- Production remains equal at extra labour input and extra costs;
- Production falls in spite of extra labour input and costs.

The sum of production loss and possible extra costs can vary widely, depending on the way extra labour input is realized. Acquisition of labour from within the firm will often be cheaper compared with hiring external personnel. Exact data about the magnitude of losses and extra costs as well as the frequency of these possible outcomes are lacking. Therefore a macro-economic approximation of the indirect costs should tackle this problem. The labour costs of the sick employee during the friction period may be the best possible estimate of the average indirect costs. Additional costs for selecting and training personnel may be relevant if the sick person must be replaced permanently.

In the Netherlands the current level of registered and hidden unemployment is substantial. Recent labour market studies indicate that with growing employment many yet non-participating women acquire new jobs. Still the participation rate for women is low, so an important part of non-participating women can be considered as involuntarily unemployed (Dept. of Social Affairs and Employment, 1989; Teulings et al., 1987). Furthermore, Dutch studies indicate that 15-20% of the prevalence of persons receiving disability benefits should actually be considered as unemployed (Roodenburg and Wong, 1985; Vrooman and Kemp, 1990).

Therefore we assume that in general replacement of sick persons by unemployed is possible after a friction period. The production loss during the friction period is assumed to be equal to the labour costs of the sick person. For an estimate of the total indirect costs of a disease two items are to be analyzed: the costs per friction period and the frequency of friction periods.

The costs per friction period are calculated for the Netherlands 1988. Because of insufficient data on the quantity of unpaid labour and its consequences for absence from work and disability the analysis yet had to be restricted to paid labour.

The average value of production per labour-year has been approximated by combining the net domestic product, measuring the total net value added by all employed labour, with the number of labour years in 1988 (CPB, 1989). This average production value was made age- and sex-specific using data about the relative gross labour incomes per labour-year by age and gender (CBS, 1988a). We assumed a stable and linear relationship between gross income and the value added per labour year.

Data on the relative amount of full time and part time labour by age and sex were extrapolated on the basis of a detailed study (Teulings et al., 1987) and used to calculate the average production value *per working person* per year. Correction for the age- and sex- specific share of employed persons in the population yields the average production value *per person* per year by age and sex.

The value per working person is applied to data on absence from work and disability. The production value per person will be used to calculate the indirect costs of mortality. We assumed equal mortality rates for employed and not-employed.

The length of the friction period was derived from data on the average period necessary to fill vacancies: 2.5 months in 1988 (not age- and sex-specific) (CBS, 1989a).

The production values per year are applied in calculating the potential indirect costs. The values per friction period lead up to the friction costs after combining them with the frequency of friction periods.

We would prefer using estimates of the marginal productivity instead of average productivity. However, reliable data on marginal productivity are difficult to obtain, especially for the services sector.

Data on absence from work, disability and mortality provide information about the frequency of friction periods (CBS, 1988b; CBS, 1989b; GMD, 1988). These databanks are not linked at the level of individuals which makes some assumptions necessary. For the Netherlands a distinction must be made between employees and self-employed, as a consequence of differences in registration.

For employees the possible flows between healthy, absent of work, disabled and dead are depicted in figure 7.1. In case of "immediate death" (flow 1) and absence from work (flow 5) a friction period is relevant. Premature death while one is absent of work (flow 3) can induce double counting of friction periods if the absence from work is longer than the friction period. We can't correct for this, because the number of people dying during absence from work and the relation with the length of absence are unknown. However the results in table 1 show that the maximum level of double counting is still small. Flows 2 and 4 don't cause friction costs because in the Dutch system one acquires the status of disabled only after a stay of one year in the absence from work insurance. In this case the friction period took place in a previous year.

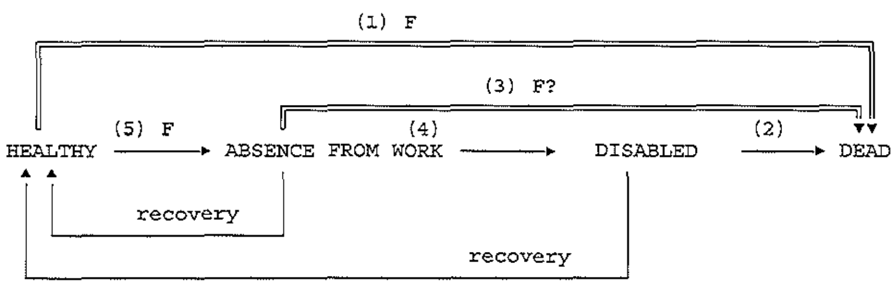

Figure 7.1 Flows between healthy, absence from work, disabled or dead for employees, including the relevance of friction periods (F).

For self-employed no absence from work register exists, so the frequency of friction periods must be derived from data on immediate death (flow 1) and incidence of disability (flow 3), see figure 7.2. Flow 1 was calculated by correcting total mortality for the number of disabled persons dying.

If internal labour reserves prevail, they can further reduce the frequency of friction periods and the level of production losses. A study of the Dutch industry for the period 1975-1982 estimated the level of *general* internal labour reserves as between 5 and 15% of the existing labour volume, concluding that 5 % seems a minimum level of general internal labour reserve of firms (de Koning, 1987). Given the heterogeneity of labour the reserve will be lower for each individual employee, but how much lower is unknown. Because of this uncertainty no further corrections are made for internal labour reserves.

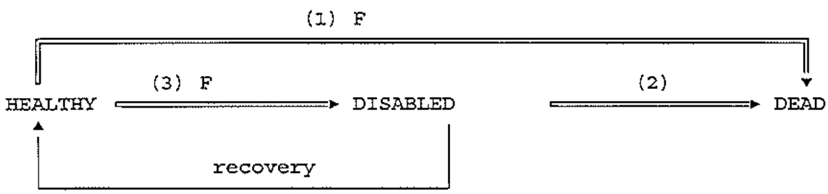

Figure 7.2 Flows between healthy, disabled and dead for self-employed, including the relevance of friction periods (F).

As far as the internal labour reserves exist for purposes of covering for sickness absence, these costs are clearly related to sickness. Internal labour reserves can also be useful in coping with fluctuations in the volume of production. In that case the costs are not caused by sickness. These remarks illustrate the need for further investigation of the role of internal labour reserves.

The assumptions for calculating the friction costs and the potential indirect costs can be summarized as follows. In both cases the incidence of absence, disability or mortality in 1988 is the viewpoint of analysis. In case of the friction cost method one friction period is counted per death until age 65, correcting for disabled persons dying. Incidence of disability is only relevant for self-employed persons. Days of absence from work are counted, but only within the friction period of two and a half months.

In estimating potential indirect costs production losses caused by mortality are counted until age 65. The average period of disability has been estimated for each 10-year age group and sex and was applied to the incidence of disability in 1988. All registered days of absence from work are counted. Costs beyond 1988 were discounted by 5% yearly.

Results

Table 7.1 shows the estimated indirect costs of the incidence of cardiovascular disease. The estimates are based on the production losses of paid labour only, due to insufficient data on unpaid labour. The results differ substantially depending on the method used, as could be expected. Total costs according to the human capital method are over 5 billion guilders, the friction cost methods results in somewhat more than 400 million. The major part of this difference is due to the cost of disability and mortality because of the great difference in the length of the period in which production losses are assumed to prevail. The results for the costs of absence from work differ less dramatically. The relative difference in costs of both methods corresponds with a Dutch study of the costs of traffic-accidents which used a variant of the friction cost-method (van Haselen, 1987).

Considering the types of cost per method, the importance of the costs of absence from work diverges completely. More than 80 % of the friction costs are caused by absence from work compared with only 12 % for the human capital method. The opposite holds for the shares of costs induced by mortality and disability. The above mentioned possibility of double counting in case of people dying while absent of work is by definition lower than 60 mln Dfl, the friction costs caused by mortality, see table 7.1.

Table 7.1 Indirect costs of incidence of cardiovascular disease according to the friction cost method and the human capital approach (gross variant) for the Netherlands in 1988 in millions of Dutch guilders. Costs beyond 1988 were discounted by 5% per year.

	Friction cost (mln dfl)		Human capital (mln dfl)	
Mortality	60	(14%)	2,072	(39%)
Disability	19	(4%)	2,615	(49%)
Absence from work	353	(82%)	624	(12%)
Total costs	421	(100%)	5,311	(100%)

Discussion

Comparing both methods of estimating indirect costs the following remarks can be made. Both methods can provide only global results as they use aggregate data.

In some studies estimates of production losses according to the human capital theory are corrected for the chance for an individual to become unemployed (Muller and Caton, 1983; Berk et al., 1978). This indeed influences the individual income, but does not necessarily have an impact on society's level of production. So this won't necessarily lead to indirect costs according to our definition.

A critical remark on the friction cost method may be that its credibility depends on the existence of substantial unemployment. However, the friction cost method is sensitive to the level of unemployment. If unemployment declines, the average period to fill vacancies will become longer, inducing higher costs per friction period. The situation on the Dutch labour market has actually developed in this direction since 1986. The costs per friction period in 1990 will therefore be larger than our estimates for 1988. Nevertheless, in parts of the labour market with serious shortages, production losses will certainly be higher than estimated by the friction cost method.

The most important shortcoming of the results presented here is the lack of a reliable estimate of the production loss of unpaid labour, understating the production losses of women and elderly people to a large extent. This could lead to undesirable policies in favour of intensifying efforts to prevent and fight diseases from which predominantly men are suffering. This again shows the implications of calculating indirect costs for equity. Data on unpaid labour are very scarce. This also holds for the amount of absence from work and disability with respect to unpaid labour. Both are indispensable elements in estimating the indirect costs.

With respect to the internal reserves of labour within firms three qualitative comments can be made. First, it is plausible that their size will be influenced by the expected sickness absence, especially if the work can't be postponed and the sick employee must be replaced immediately (e.g in case of busdrivers). This will diminish the amount of production lost by disease, but at extra costs for maintaining this labour reserve. The costs of this reserve are not necessarily high, because firms may be able to hire employees for part-time jobs, who can be employed more hours if necessary (Reich et al., 1973). Secondly, the size of internal reserves and the level of unemployment may be inversely related, because in times of high unemployment firms can more easily acquire new personnel to replace sick employees. On the other hand, in the Netherlands temporary employment agencies play an important role in providing a flexible supply of labour which can act as a buffer, reducing fluctuations in the size of internal reserves. Thirdly, internal labour reserves can be held partly in order to cope with fluctuations in the volume of production. In this case not all costs are directly related to sickness. It can be concluded that the numerous aspects of the relation between costs of disease and internal labour reserves should be studied in more detail.

In estimating production losses one should be careful when considering only the registered sickness absence, because sick people can continue working at a suboptimal level. The extent of this phenomenon is very difficult to quantify, but especially in times of high unemployment, it can be more than negligible.

Further analysis will be necessary on the costs of selection and training if vacancies have to be filled permanently and the level of internal labour reserves for specific jobs.

A stratification of the labour market assumptions seems useful to incorporate differences between parts of the labour market which are relevant for the level of production losses. The labour market can be split into educational levels, each with different production values and length of the friction period. This will refine the estimates of indirect costs and takes labour shortages in some market segments into account. For such refinement data on incidence and mortality by education level should be available.

Conclusion

In setting priorities in health care reliable estimates of direct and indirect costs of disease are indispensable. Thus far studies only provide estimates of the potential indirect costs of disease.

The friction cost method can be considered as a promising method of estimating the indirect costs which may actually result from the incidence of disease. This approach explicitly takes into account short run and long run adaptation processes in the economy. The friction cost method is therefore highly compatible with the methodology of cost-effectiveness analysis in which the real resource costs (of a disease or technology) are subject of analysis.

Because the costs per friction period depend on labour market conditions the estimates of the indirect costs can vary widely over time and across countries. In situations with a tight labour market the differences between the potential costs and the friction costs will become much smaller than shown in table 7.1. The differences also depend on the type of disease. In case of diseases causing predominantly disability and mortality the gap will be much wider than for diseases merely inducing absence from work.

The proposed methodology needs further development. Especially the consequences of illness in the longer run have to be considered in more detail, preferably by linking the analysis to a macro-economic model. Presumably, these consequences are minor, especially when illness occurs in more flexible and adaptive sectors of the economy. Furthermore, our empirical analysis is global and preliminary. Further refinement of the underlying labour market assumptions is necessary; both with respect to distinguishing different segments on the labour market as to a better analysis of the amount of sickness-related costs caused by internal labour reserves. Extension of the analysis to the consequences of illness in people without a paid job is needed.

8 The friction cost method for measuring indirect costs of disease[1]

Summary

A new approach for estimating the indirect costs of disease, which explicitly considers economic circumstances that limit production losses due to disease, is presented (the friction cost method). For the Netherlands the short term friction costs in 1990 amount to 1.5-2.5% of net national income (NNI), depending on the extent to which short term absence from work induces production loss and costs. The medium term macro-economic consequences of absence from work and disability reduce NNI by an additional 0.8%. These estimates are considerably lower than estimates based on the traditional human capital approach, but they better reflect the economic impact of illness.

Introduction

The inclusion of indirect costs of disease in economic appraisal is subject to considerable debate and scepticism. The disagreement on the relevance of indirect costs resulting from reduced productivity can be illustrated by considering recent guidelines for economic evaluations in health care, that are issued by Australian and Ontarian health authorities in the context of decisions on reimbursement of pharmaceutical products. The Australian guidelines suggest to exclude indirect costs, whereas the Ontarian seem to support inclusion (Commonwealth of Australia, 1990; Ontario Ministry of Health, 1991). Apart from ethical objections to any form of incorporating indirect costs in economic evaluation, criticism concentrates on the predominating estimation method: the human capital approach. Defining indirect costs as the value of production lost, more precisely as wealth lost, to society due to disease (and hence not as valuing lives per sé), the human capital method

---

[1] Chapter based on Koopmanschap et al., J of Health Economics, 1994

estimates the *potential* production lost, which may overestimate the actual production lost to a considerable extent.

We introduced an alternative approach: the so called "friction cost method", which takes into account several economic circumstances, that may reduce the estimated production losses substantially as compared with estimates based on the traditional human capital approach (see chapter 7). In this paper we will develop this friction cost method further, focusing on the influence of the labour market situation on indirect costs, the impact of absence from work on labour productivity and the macro-economic consequences of absence from work and disability. The estimation of indirect costs takes place in two steps. First, the short term effects of illness are considered by studying its possible consequences at the level of the firm.

Second, medium term effects occur because of changes in labour costs per unit of output, social insurance premiums and labour supply and these will influence national income, the most generally used yardstick for society's wealth, as well as other economic indicators such as (un)employment, inflation, the balance of payments and the government deficit.

As an illustration total cost of absence from work, disability and mortality respectively, will be estimated for the Netherlands in 1988 and 1990, using the friction cost method and the human capital approach. However, in this case estimates of total costs of disease represent the possible savings of eradicating disease and hence come down to estimating "the benefits of the unattainable" (Shiell et al. 1987). Therefore, the more relevant consequences of attainable reductions in absence from work and disability will be calculated. The impact of several assumptions will be investigated in a sensitivity analysis.

The friction cost method

Indirect costs disputed

Including indirect costs in economic appraisal of health care programs is not universally accepted. Opponents of inclusion state that health care programs directed to working people always generate more gains in production than programs for mentally handicapped or elderly. Even if allocation of resources to individuals is not driven by such considerations, this could well lead to priority setting enlarging existing inequalities in health. This argument should be taken seriously. Nevertheless, production losses due to illness do influence the scarcity of resources and therefore the wealth of society. In this respect there is no difference in assessing direct and indirect costs; both should be incorporated in economic evaluations of health care programs.

There may be a trade-off between equity and efficiency: society's wealth will indeed gain more from programs directed to economically active people. Explicit estimates of the extent of the trade-off may be more helpful in decision making than using implicit notions. In health policy decisions, estimates of indirect costs should be weighed in conjunction with other criteria like direct costs, ethical considerations and equity.

The "costs" of pain, suffering, discomfort etc. are sometimes included in the concept of indirect costs. In this paper we focus on the indirect "economic" costs of illness.

The human capital approach and the friction cost method

Almost all studies estimating indirect costs use the human capital approach (see chapter 6). This method estimates the value of *potentially* lost production (or the potentially lost income) as a consequence of disease. We use the word potentially because in case of permanent disablement or premature death at a specific age the total productive value (or income) from that age until the age of retirement is counted as indirect costs. Many authors have suggested that the real production losses for society may be much smaller (Drummond, 1992; Gerard et al., 1989; Lindgren, 1981; Williams, 1985). For short term absenteeism, work may be taken over by others, non-urgent work may be cancelled or made up for by the sick employee on his return to work. For long term absences, work can be taken over by someone drawn from the ranks of the unemployed or by reallocating employees over jobs.

More systematically, if an employee gets ill, the following outcomes are possible for a firm's production and costs (assuming some form of social insurance for absence):

1) Both the level of production and costs are unaffected. This situation may occur if work can be made up for by the sick employee on his return to work or if internal labour reserves exist, allowing work to be taken over by colleagues without extra costs. The opportunity costs of internal labour reserves depend on the probability for the internal labour reserve to be gainfully employed elsewhere. If unemployment is well above the level of frictional unemployment, these costs are very low. However, the existence of permanent internal labour reserves raises labour costs, which may have medium term macro-economic implications, as described below.

2) Production remains unchanged, but at higher costs, due to colleagues working overtime or hiring temporary workers, from a firm's own pool or from temporary agencies. In both cases the extra costs of maintaining production tend to be somewhat higher than average labour costs, as a result of higher wages paid for working overtime (reflecting the opportunity costs of leisure time) or the extra costs of using temporary agencies.

3) Production falls, while costs remain unchanged. The value of production lost is the relevant outcome.

4) Production falls, despite higher costs. The consequences are a mix of production loss and extra costs of permanent or temporary employees, which may be higher (or lower, although this is not to be expected) than the value of production of the sick employee.

In case of situation 1, zero costs are incurred in the short run, whereas the medium term consequences need to be analyzed, see below. Concerning the other three possible situations, the sum of production loss and extra costs varies from case to case, but on average it may well be approximated by the productive value of the sick employee during the period of absenteeism.

Because in situation 1 no short term costs prevail, we need to know its probability of occurrence. Exact empirical data on this probability are non-existent, but there is some material on the underlying determinants of the employment policies of firms. First, it should be known how often work can be made up for by the sick employee on his return to work. This possibility seems to be restricted to short term absence for jobs for which production can be postponed for a while and which output is not an indispensable input for other jobs. For example, a bus driver must be replaced in order to deliver the transportation services, whereas a sick health economist may make up for (most of) lost production a week later. Because we lack evidence on the magnitude of this phenomenon, we performed a sensitivity analysis regarding this issue.

The occurrence of situation 1 also depends on the degree to which firms maintain internal labour reserves, which in turn depends on the relative costs and risks of keeping internal labour reserves versus flexible forms of labour provision, such as temporary workers, flexible contracts etc. The most important determinants for internal labour reserves versus flexible labour appear to be (Williamson, 1981; Osterman, 1987; Mitchell and Zaidi, 1990):

- the macro-economic situation and the state of the relevant product market. The more uncertainty exists, the greater the employer's desire to make labour costs variable and execute flexible labour policies;

- the situation on the labour market. The tighter the labour market, the more a firm wants to maintain internal labour reserves, because flexible labour may become very scarce and expensive;
- the technology of the production process. The more delicate and risky the production technology used, the more skilled employees are required and the more suitable internal employment is;
- the firm-specificity of employee's skills. In a downturn market employees with firm-specific skills will not be laid off soon, since training new employees if the market recovers is expensive;
- the predictability of the level of production (business cycle) and the possibility of production in advance. The less predictable production, the higher the costs of a permanent internal labour reserve. If stock production is not possible, as often in producing services, flexible labour may be very useful to produce the output required;
- the heterogeneity of labour. The less homogeneous labour is, the larger the total amount of internal reserve required and the more expensive it is as compared to flexible labour reserves;
- the predictability and variance in time of absence from work. If the absence pattern is unpredictable and clustered in time, flexible labour reserves seems more attractive;
- the degree of competition. The more competitive a market, the larger the propensity to choose for less expensive, flexible labour reserves;
- the institutional setting. A liberal labour market legislation facilitates flexible labour policies, for example by allowing immediate dismissal.

In summary, choosing the appropriate mix of internal and flexible labour reserves depends on the trade-off between three different goals of a firm: - minimizing wages for any given level of output, - maximizing predictability i.e. the confidence of the availability of qualified labour at foreseeable prices and - flexibility with respect to staffing levels and abilities of the labour force (Osterman, 1987). Enhancing predictability always has a certain price in terms of higher labour costs and/or less flexibility and vice versa.

Empirical research for the United Kingdom, the Netherlands and the U.S.A. clearly shows that during the eighties the amount of flexible labour used has increased, even during the upturn of 1983-1990 (Hunter et al., 1993; OSA, 1987; Elfring and Kloosterman, 1993; Mitchell and Zaidi, 1990). In addition, Hunter et al. found that the firm's main motives for extending flexible labour are demands for short term cover as well as deliberate responses to increased product market uncertainty. Although these results unambiguously indicate a growing role for flexible labour, they can not answer the question on the relative quantity of internal labour reserve versus the use flexible labour arrangements. Hence, we assume for the moment that all short term absence spells induce indirect costs to some extent (for long term absence see below). Consequently, the estimates of friction costs as described below should be viewed as *upper bound estimates* of the short term indirect costs of disease. The lower bound estimates are described in the sensitivity analysis.

Returning to the friction cost method, its basic idea is that the amount of production lost due to disease depends on the time-span organisations need to restore the initial production level. We assume that if unemployment, registered and hidden, is beyond the level of frictional unemployment, sick employees can be replaced, after a period necessary for adaptation. Frictional unemployment is an inevitable part of unemployment, since filling vacancies takes time and some qualitative discrepancies between labour demand and supply always prevail. Production losses are assumed to be confined to the period needed to replace a sick worker: the friction period. This situation currently applies to the Netherlands, where the level of registered and hidden unemployment is substantial. In 1988 and 1990 registered unemployment was 10.2% and 8.2% respectively (CPB, 1992). In addition, several studies indicate that 15-20% of the persons receiving disability benefits should be considered as unemployed, thus adding another 3% unemployment (Roodenburg and Wong, 1985; Vrooman and Kemp, 1990).

In order to calculate the indirect costs of disease, using the friction cost method, the following questions need to be answered:

- when does a friction period occur?
- how long does a friction period last?
- how can the indirect costs during a friction period be determined?
- how can the medium term macro-economic consequences of illness be estimated?

Operationalisation of concepts

The frequency of friction periods

The possible flows between healthy, absent from work, disabled and dead, which depend on the specific social security arrangements, determine when friction periods occur (see figure 7.1 and 7.2 for the Dutch case). Hence, in order to estimate the total number of friction periods one needs data on the frequency and length of absence spells, the incidence of disability and mortality, correcting for the number of deaths in the disabled population.

The length of the friction period

The length of the friction period is based on the average vacancy duration, which depends on the level of unemployment and on the efficiency of the labour market in matching labour demand and supply. Substantial unemployment in an economy as a whole does not rule out labour scarcity in specific parts of the labour market. Assuming a homogeneous labour market may result in underestimating indirect costs, and therefore one should consider at least different segments of the labour market separately, preferably according to education level. The education level has proved to be an important determinant of (un)employment (Teulings and Koopmanschap, 1989).

As a rule, the friction period is longer than the vacancy duration, since time may elapse between the emergence of production loss and the decision to create a vacancy. In addition, time passes between filling a vacancy and the first working day of the new employee, especially if he or she is already employed. Even filling a vacancy by an employee from the same firm takes time. We take both circumstances into account in estimating the length of the friction period.

Absence and productivity

As noted above, short-term absence may lead to production loss and/or extra costs to continue production, but the extent to which this occurs depends on various circumstances as discussed above. Complete insight in the consequences of short term absence for indirect costs requires more detailed study on the level of the firm. However, generally speaking, absence from work reduces the effective labour time. Numerous studies have demonstrated that a reduction of annual labour time causes a less than proportional decrease in labour productivity per year. The estimated elasticity for annual labour time versus labour productivity is between 0.6 and 0.9. (CPB, 1987a; de Koning and Tuyl, 1984; WRR, 1977). In our main variant we assume the elasticity to be 0.8. It should be noted however, that because these elasticities were estimated on the level of the firm, they inevitably reflect the composite result of two possible components: the diminishing returns on labour and a possible reduction of the internal labour reserve. On the level of the firm, these two components can not be disentangled.

The valuation of lost production

We assume that the production level is restored after (a part of) the friction period. The actual indirect costs of disease consist of: the value of production lost and/or the extra costs to maintain production and, if an employee is to be replaced permanently, the costs of filling a vacancy and training new personnel. We assume the costs to be 80% (the aforementioned 0.8 elasticity of labour time versus productivity) of the average value of production per employee, by age-group, sex and education level. Costs of absence from work shorter than the friction period are calculated as being 80% of the production value during the period of absence. Costs of absence spells longer than the friction period are equal to the costs of a friction period, since we assume replacement after completion of the friction period. The production value per time period per employee is multiplied with the length of friction periods due to absence from work and incidence of disability. Because we assume equal mortality rates for employed and non-employed, the average production value per person (the production value per employee, corrected for participation in paid labour) is used to calculate the indirect costs of mortality before age 65.

Macro-economic consequences

The friction cost method concerns the short term indirect costs of disease. In addition, absence and disability may also have medium term macro-economic consequences. Absence from work, reducing labour productivity, directly affects the labour costs per unit of output, which is an important determinant of competitiveness on the world market. This may change the production of firms in the "exposed sector" of the economy: exporting firms as well as firms producing for the domestic market facing foreign competitors. Consequently, national income and other macro-economic indicators may be affected.

The amount of insurance premiums related to absence has no separate economic impact in the Netherlands. A change in absence premiums paid only causes a redistribution between wages and sickness benefits, total labour costs per worker remaining equal.

Although disability benefits themselves do not represent indirect costs, as they are income transfers, they may have economic consequences in two ways. First, insurance premiums influence the labour costs per worker and per unit of output, affecting the performance of the exposed sector of the economy. Secondly, insurance premiums affect the net disposable income of employees and employers, influencing private consumption, domestic production and national income. The extent of this influence depends on the possibilities for employees and employers to pass on changes in premiums to each other.

Furthermore, disability may have other economic consequences, in addition to the "premium-route" just described. Reducing the incidence of disability will raise labour supply and unemployment. High unemployment has a negative medium term impact on the average wage level and thus on labour costs, improving the competitiveness of the exposed sector. On the other hand, the high prevalence of disability in the Netherlands has undoubtedly raised the productivity of the labour force, because in the Netherlands less productive workers run a higher risk of becoming disabled (Aarts and de Jong, 1990). Policies, effective in lowering the incidence of disability, may have a small negative effect on average labour productivity.

These macro-economic consequences of absence and disability can only be estimated by using a macro-econometric model, allowing for all interactions and time lags involved.

Data and methods for the Netherlands

Frequency of friction periods

We used national, aggregated data on the frequency and length of absence from work, covering 50% of employment. We corrected for the fact that the absence rate for persons for whom we lack exact data is somewhat lower (CBS, 1989c and 1991). Data on incidence of disability and mortality by age and sex were available, covering the entire population (CBS, 1989b and 1991a; GMD, 1989 and 1991a).

The relation between absence and education level (and age and sex) for 1988 was derived from a sample of 25,000 Dutch workers (NIA, 1993). For the risk of becoming disabled by education level, we used the estimates of Bloemhoff and de Winter (1991). Mortality rates for men by education level for the Netherlands were estimated by Doornbos and Kromhout (1990) and applied to men and women.

Length of the friction period

We split up the labour market into 5 education levels. This reduced heterogeneity considerably, because in the Netherlands unemployment is clearly higher for lower educated (CBS, 1991b). Van Ours' proportional hazards model showed that the longer the vacancy rate, the smaller the probability that a vacancy is filled (van Ours and Ridder, 1991). Following van Ours' model, we estimated the completed vacancy durations by education level for 1988 and 1990, based on quarterly data on the uncompleted vacancy durations and the number of vacancies, from a large sample of Dutch firms (CBS, 1988a and 1990a), adding an additional period to allow for time lags indicated above.

Absence and productivity
The elasticity for annual labour time versus labour productivity, found for the Netherlands is between 0.6 and 0.9. (WRR, 1977; de Koning and Tuyl, 1984; CPB, 1987a) We assume the elasticity to be 0.8 (de Koning and Tuyl, 1984), performing sensitivity analysis on this parameter.

Valuation of lost production
The average value of production per labour-year was approximated by combining the net national product and employment in labour years in 1988 and 1990 (CPB, 1992). This production value was made age-, sex- and education level-specific using data on the relative gross labour incomes per labour-year by age, sex and education level (CBS, 1989a and 1991c; LTD, 1993). We assumed a linear relationship between gross income and value added per labour year. Data on participation in full-time and part-time labour, based on the annual Dutch labour force sample (CBS, 1989d and 1991b) were used to calculate the average production value *per employee* and *per person* (age-, sex- and education-specific).

Macro-economic consequences
The macro-economic consequences of absence and disability were calculated, using the quarterly macro-econometric sectoral model FKSEC for the Netherlands, developed by the Dutch Central Planning Bureau (CPB, 1992a). This general equilibrium model has been chosen after careful comparison of several macro-econometric models on their overall performance as their suitability for estimating the macro-economic consequences of disease. The model FKSEC is the youngest member of a family of well known Dutch macro-econometric models and is currently the most frequently used in advising the Dutch government and parliament on short- and medium term economic development.

It explicitly models the production process on the sectoral level (6 sectors), taking account of the differences between sectors in labour intensity, in the degree of substitutability between capital and labour and in the speed with which different sectors adjust to shocks in demand. In the capital-intensive exposed sector, the production capacity is based on a vintage production function. Vintages of capital goods for which the labour productivity is lower than the real wage have a negative return and are scrapped. Employment in the exposed sector is related to labour demand at full capacity and the utilization rate of capital. Because in the sheltered sector and other sectors labour is the main factor of production, capacity is based on the current labour force and the lagged trend in labour productivity. The employment level is determined by production (given production capacity), real labour costs, the ratio between the minimum and the average wage, working time and labour augmenting technical progress.

A historical simulation for the period 1979-1988 was carried out with the model, showing that FKSEC is capable of reproducing the most crucial developments in the eighties, such as the decrease in export growth, investments and production in the beginning of the decade and the acceleration afterwards (CPB, 1992a). An extensive survey of the CPB macro-economic models is given by Bodkin (Bodkin, 1991).

The model does not have a separate variable for absence from work, but absence from work primarily influences labour time per year and hence labour productivity per worker. A change in the number of disabled principally affects labour supply, social insurance premiums and labour productivity. The model simulates the rest of the consecutive chain of events including changes in labour costs, production, consumption, exports, national income etc.

The macro-economic consequences of disease are calculated as differences between the simulation runs and the base run. The latter reflects the expected economic development, without serious changes in the development of absence and disability. The simulation period is 1991-1998, allowing all possible time-lags to effectuate. In the simulation runs changes in absence and disability are assumed to be structural.

For the *main variant* in the following results we have split the labour market into five education levels and assume the elasticity of labour time versus production to be 0.8 (de Koning and Tuyl, 1984). Absence from work, incidence of disability and mortality are also related to sex, age and education level.

Results

Short term friction costs

The average friction period was estimated to be 2.8 months in 1988 and 3.2 months in 1990, including the vacancy duration and 3-5 weeks to allow for the time lags mentioned (table 8.1). In both years the friction period is considerably longer for higher education levels, reflecting higher unemployment for lower educated and a longer period needed for hiring higher educated workers. In 1990 friction periods converged, as the labour market for higher educated became less tight, while at the same time lower and intermediary educated technicians became relatively more scarce.

Table 8.1 Estimated length of friction periods for the Netherlands in months, by education level, for 1988 and 1990.

	Education level					
Year	Overall	Basic	Extended basic	Interme-diary	Higher vocational	University
1988	2.8	2.2	2.5	2.7	3.5	3.8
1990	3.2	2.8	3.0	3.3	3.4	3.5

The average production loss per employee, per friction period is higher for men than for women and increases sharply with education level and with age, levelling off at age 50 (see appendix 8). This reflects the strong positive relation between hourly wages and both age and education level and the substantial amount of part time work performed by women in the Netherlands.

Absence from work is strongly related with education: absence rates for the lowest educated are about 4 times as high as for the highest education level. Absence increases with age and is higher for women under 45 compared to men, partly due to pregnancy. The lower educated also run a higher risk of becoming disabled: the relative risk of the lowest versus the highest education level is 2.3, after correction for age. The relative risk of dying before age 65 for the lowest educated is 1.5.

The friction costs of disease for the main variant (assuming a heterogeneous labour market and labour time elasticity of production = 0.8) are shown in table 8.2. The friction costs amount to 9.5 billion guilders in 1988 and 12 billion guilders in 1990, about 2.5% of net national income. About half of the cost increase between 1988 and 1990 is due to employment growth (raising the absolute number of absence days), productivity growth and inflation. The other half is the result of (on average) longer friction periods, mirroring a more tight labour market. The costs of absence from work predominate, since we assume production losses to be confined to the friction period and consequently disability and mortality hardly contribute to the costs. The distinction by education level has a small upward effect on costs. The higher costs per hour and longer friction period for higher educated are almost entirely compensated by their considerably lower rate of absence.

Table 8.2 Indirect costs of disease in the Netherlands for 1988 and 1990 (main variant) in billions of Dutch guilders (in parentheses as % of net national income).

Cost category	Friction costs 1988	Human capital costs 1988	Friction costs 1990
Absence from work	9.2	23.8	11.6
Disability	0.15	49.1	0.2
Mortality	0.15	8.0	0.2
Total indirect costs	9.5 (2.1%)	80.9 (18%)	12.0 (2.6%)
without education split	*9.3*		*11.6*

Table 8.2 also presents the indirect costs for 1988, according to the human capital approach, which amount to more than 80 billion guilders, 18% of net national income. The huge difference compared to the friction costs is due to the cost of disability and mortality, which are assumed to cause production losses in the long term. Costs of disability are very large, because the average duration of disability in the Netherlands is 15 years.

The costs of absence from work differ less dramatically. As the friction cost method assumes production loss to be limited to the short run, the friction costs for diseases that cause mainly disability and mortality are much lower than according to the human capital approach. For diseases entailing short term absence, the difference is much smaller.

Table 8.3 Macro-economic consequences of eliminating absence from work and incidence of disability respectively, during 1991-1998, in % of net national income and other macro-economic indicators (differences in 1998 between the simulations and the base run).

	NNI	Employ-ment x1000	Produc-tion	Export	Bal.of pay-ments % NNI	Infla-tion	Govern. budget % NNI
I Absence from work	+0.30%	- 112	+0.9%	+2.0%	+0.1%	-0.7%	-0.2%
II Disability total effect	+0.49%	+345	+2.2%	+4.0%	+0.8%	-4.2%	-1.5%
IIa premium effect	+0.07%	+ 71	+0.4%	+1.0%	+0.5%	-0.9%	+0.6%
IIb labour supply effect	+0.42%	+271	+1.9%	+3.1%	+0.3%	-3.3%	-2.1%

* The sum of the partial effects IIa and IIB is not equal to the total effect II, due to very small interaction effects

Medium term macro-economic consequences

The medium term macro-economic consequences of absence and disability are presented in table 8.3, as structural differences in 1998 and later years, between the specific variants and the base run. The average Dutch absence rate is 8%, so we increased working time by 9% (={100-92}/92) during 1991-1998 to illustrate the impact of eliminating absence from work. In 1998, overall labour productivity will be 3.3% higher than in the base run: a 5% increase in the exposed and construction sector and 2% in the sheltered sector. The production of the exposed sector will increase by 1.6%. In the sheltered sector and the construction sector the production increase is only 0.6%, since it is limited by domestic demand. As a consequence, employment falls, raising unemployment.

Labour costs per unit of output decrease, causing an increase in export (+2%) and overall production (+0.9%) from 1994 onwards. The increased unemployment will push down wages, so private consumption slightly falls. The decrease in consumption occurs almost immediately, whereas the increase in production and export is lagged. As a consequence, it takes 4 years for the national income to increase as compared to the base run. After 8 years the structural increment in national income is 0.3%, which is considerably smaller than the increase in export.

Assuming no incidence of disability implies that each year during 1991-1997 labour supply increases by 116,000 persons, who would otherwise have left the labour force. As a result, disability premiums and the corresponding cash benefits decrease by 3.1 billion guilders per year. The second row of table 8.3 shows the total consequences of disability. Labour costs per worker decrease substantially (-16% in 1998), due to lower premiums and increased labour supply (the "Phillips curve-effect"). Employment grows considerably, but it can only absorb one third of the extra labour supply, so unemployment rises. Production (+2.2%) and export increase (+4%), whereas private consumption decreases (-1.7%), due to lower wage rates.

After 5 years, national income will be higher as compared to the base run, in 1998 the difference is 0.49%. The last two rows of table 8.3 show that the increase in labour supply is responsible for the lion's share of the effects. The impact of the lower disability premiums is relatively small, due to the trade-off between disability benefits and unemployment benefits. Since the total effect of disability is about equal to the sum of both effects, interaction effects appear not to be significant.

Scenarios

Table 8.4 shows the friction costs and the macro-economic consequences of three attainable scenarios regarding reduction in absence from work and disability in the Netherlands. These scenarios appear to be attainable, witnessing recent trends towards more prevention and closer monitoring of absence from work in the Netherlands. Furthermore, there is scope for reduction as the level of absence from work in Belgium and Germany is lower than in the Netherlands (Prins, 1990).

Reducing absence from work by 2% (scenario 1), from 8% to 6% of labour time, will reduce friction costs by 2.3 billion guilders, about 0.5% of net national income. The medium term economic effects induce another 0.08% annual increase in NNI. Reducing the incidence of disability by 20% (scenario 2) hardly saves any friction costs, but the macro-economic consequences are more important: 0.1% of NNI. Reducing both absence and disability (scenario 3) will save 2.3 billion guilders friction costs and 0.18% of net national income in the medium term. The effect of scenario 3 turns out to be more or less equal to the sum of the effects of scenario 1 and 2. The other macro-economic indicators are affected in the same direction as shown in table 8.3.

Table 8.4 Change in friction costs and macro-economic consequences of 3 scenarios for absence from work and disability. Friction costs in billions of guilders (1990), in parentheses as % of net national income. Macro-economic consequences in % of NNI in 1998.

Scenarios	Change in friction costs in bln dfl	Macro-economic consequences in % of NNI
1 absence -2%point	- 2.30 (0.5%)	0.08 %
2 disability incidence - 24000	- 0.03	0.10 %
3 scenario 1 and 2	- 2.33 (0.5%)	0.18 %

Sensitivity analysis

The results of the sensitivity analysis are presented in table 8.5. The first two options represent the lower bound estimates for the friction costs of absence from work.

As we lack detailed data on the consequences of short term absence, we assumed quite radically that half of the absence spells shorter than one week will not cause indirect costs, because work can be cancelled or postponed and made up for later. In this situation, the friction costs of absence from work would amount to 10.2 billion guilders in 1990, which is only 12% lower as compared to the main variant.

If we alternatively assume the elasticity of labour time versus production to be very low: 0.5 instead of 0.8, (implicitly assuming an important role of internal labour reserves) the friction costs of absence are substantially lower: 7.3 billion guilders, about 1.5% of NNI, a 37% decrease as compared to the main variant. Hence, the friction costs are very sensitive to the level of this elasticity.

Table 8.5 Sensitivity analysis: indirect costs of absence from work (excl macro-economic consequences) in billions of Dutch guilders in 1990 and differences as compared to the main variant.

Type of sensitivity analysis	Indirect costs of absence	(% diff.)
Postponement of work	10.2	(-12%)
Elasticity of production 0.5	7.3	(-37%)
Friction period 6 months	14.7	(+27%)
Human capital approach	23.8	(+105%)

In order to show the relation between the labour market situation and friction costs, we assumed a very tight labour market, entailing a long friction period, say 6 months. The friction costs amount to 14.7 billion guilders, a 27% increase as compared to the main variant. Indirect costs of absence from work based on the human capital approach amount to 23.8 billion guilders, which is considerably more than both the main variant and the variant with the 6 months friction period. The latter difference stems from the fact that in the Netherlands the proportion of long term absence is quite large. More than 20% of all absence days are accounted for by persons who stay the maximum of one year in the absence insurance scheme.

Discussion

The friction costs and the macro-economic consequences of absence from work do not overlap, but are complementary. The friction costs are restricted to the short term consequences of disease. After completion of the friction period, no costs are considered. The influence of absence on labour time, productivity, labour costs and production (especially of the exposed sector) is a medium term economic process, which takes several years to occur. Furthermore, in the macro-economic analysis, the negative price elasticity of demand ensures that lower production costs stimulate foreign demand. The friction cost method does not take this interaction of price and demand into account.

Micro- and macro-economic effects are complementary in describing the comprehensive effects of absence from work and disability. The macro-economic situation influences the length of the friction period and the relative importance of internal labour reserves, both determinants of friction costs. On the other hand, the internal labour reserves influence labour costs, which has clear macro-economic implications in the medium term.

The exact consequences of short term absence from work for labour productivity and production need to be analyzed in more detail on the level of the firm. This analysis should deal with the main issues discussed: the possibilities of postponing work and making up for it later, the nature and extent of internal labour reserves versus flexible labour and the diminishing returns to labour on the level of the individual employee.

If unemployment is very low, the friction period is relatively long. By then the vacancy duration may become a less accurate proxy of labour scarcity, as some firms will give up hiring personnel. Consequently, the registered vacancy duration may underestimate the true vacancy length and the friction costs may underestimate the value of production losses. If an absolute labour shortage would occur on a particular part of the labour market, the friction period could become as long as it takes to educate new employees appropriately. For the exposed sector of the economy this situation could mean a loss in market share and output. Facing a tight labour market, firms tend to maintain more internal labour reserves, but at the same time hiring employees is more difficult, mitigating this tendency.

We suggest that the true indirect costs of disease are within the range between the minimum estimates as provided in the sensitivity analysis and the estimates of the main variant. The minimum estimates are very conservative estimates of indirect costs, as these incorporate quite radical assumptions concerning the opportunity for postponing work and the impact of absence on productivity, respectively. This range of estimates is more useful in presenting the economic consequences of disease in the real world as compared to human capital estimates, showing the potential consequences of disease for an economy in a state of full employment equilibrium only found in textbooks.

The level of friction costs in other countries depends on the labour market and the relation between absence and education. The macro-economic consequences for other countries may be somewhat smaller. The Netherlands is a small open economy with a large exposed sector, hence, economic growth is very sensitive to changes in labour costs. However, not only the exposed sector is affected but also firms producing intermediary supplies. Furthermore, the rapid international economic integration will intensify international competition for firms producing for the domestic market, enlarging the importance of the exposed sector worldwide. The Dutch level of disability is relatively high, so reducing disability may have relatively large macro-economic effects as compared to other countries.

Although our analysis is age-, sex- and education-specific, we used fairly aggregate data which may not reflect the indirect costs for specific groups of patients. For estimating indirect costs on the individual level patient-questionnaires seem more appropriate. The latter may also be useful in estimating productivity loss without being absent from work (e.g. for ulcer or migraine) which hardly have been studied thus far. Production losses without absence may increase if absence from work is reduced without a decrease in the prevalence of the underlying disease.

Estimates of indirect costs should incorporate production loss in relation to unpaid labour to prevent adverse equity implications. However, data on work absence and disability related to unpaid labour are extremely scarce. A combination of data from time-surveys and patient questionnaires, administered by those carrying out unpaid work may provide reliable data.

9 The impact of indirect costs on outcomes of health care programs[1]

Summary

The impact of including indirect costs of disease (as a result of absence from work, disability and mortality) on outcomes of economic evaluations of specific health care programs is analyzed. For eight health care programs, changes in indirect costs are estimated using the friction cost method, that seeks to estimate the economic losses due to disease or the economic gains of health care programs. The impact of indirect costs on outcomes varies considerably across programs. Indirect costs tend to play an important role if health care programs produce health effects in the short run, if (short term) absence from work is affected considerably and if a significant proportion of the target population is employed at the moment they benefit from the program. The possible induction of treatment related absence from work and disability may also be relevant.

Introduction

Health care programs may result in changes in the ability of patients to perform paid work and other daily activities. Production losses or -gains due to illness or its treatment influence the scarcity of resources, just as health care costs. Consequently, indirect costs are relevant for health policy, provided that the estimates of indirect costs reflect the real economic impact of disease. This chapter aims at *illustrating* the impact of including indirect costs of disease on outcomes of economic evaluations of specific health care programs. The changes in indirect costs as a result of eight different health care programs are estimated using the friction cost method, that seeks to estimate the real amount of production losses due to disease or the production

---

[1] Chapter based on Koopmanschap and Rutten (submitted)

gains of health care programs.

The impact of indirect costs on outcomes varies considerably across programs, in terms of the type of health care program, the disease category involved and the population aimed at. Information about the characteristics of those interventions where inclusion does make a difference, may help to decide on the resources to be devoted to estimating indirect costs in a particular evaluation study.

Methods

The basic idea of the friction cost method is that the amount of production lost due to disease depends on the time-span organisations need to restore the initial production level. We assume that if unemployment, registered and hidden, is beyond the level of frictional unemployment, sick employees can be replaced, after a period necessary for adaptation: the 'friction period'. The length of this friction period and the amount of production lost depends on the situation within firms and on the labour market. The estimates for the length of the friction period are based on the time it takes to fill vacancies (see chapter 7 and 8).

The influence of including indirect costs was estimated for eight specific health care programs, where reliable estimates of indirect costs were available (table 9.1). These studies were selected as they represent a wide spectre of health care interventions regarding the type of service (preventive and curative), disease category (acute and chronic) and target population (young, middle-aged and elderly). In this chapter the estimates of indirect costs are restricted to paid labour. For an indication of the relevance size of indirect costs we used a relative indicator, the percentage change in outcome as a result of including indirect costs in the analysis.

In order to calculate the friction costs three questions need to be answered:

- when does a friction period occur?
- how long does a friction period last?
- how can the costs of lost production during a friction period be determined?

Applying the friction cost method implies that absence from work only invokes production loss during the friction period. If the duration of absence is longer than the friction period, the production loss is assumed to be limited to the length of the friction period.

Table 9.1 Overview of eight health care programs by type of program, type of study, first author and the year the costs refer to.

Program	Type of study	Country	First author	Year
Breast cancer mass screening age 50-70, each 2 years	CEA	Netherlands	de Koning	1990
Cervical cancer mass screening age 37-73, each 6 years	CEA	Netherlands	Koopmanschap	1988
Ace-inhibitors in congestive heart failure	CUA	Netherlands	van Hout	1988
Cimetidine versus surgery in peptic ulcer	COI	Sweden	Jönsson	1983
ß2-agonist+Corticosteroid vs ß2-agonist+placebo in asthma/copd	CEA	Netherlands	Rutten	1989
Hepatitis B vaccination	CEA	Netherlands	Eelkman-Rooda	1992
Thrombolitics vs placebo after myocardial infarction	CUA	Sweden	Levin	1989
Total hip arthroplasty in osteo-arthritis etc.	Cost study	USA	Cushner	1986

CEA: Cost-effectiveness analysis
CUA: Cost-utility analysis
COI: Cost of illness analysis

In case of mortality for persons younger than 65, we assume that the production loss is equal to the production loss for a complete friction period, taking into account the participation in paid work for the specific group of patients.

Thus far, the length of the friction period has only been estimated for the Netherlands. For 1988 it was estimated to be 2.8 months, for 1990 3.2 months on average. In the absence of data on the length of friction periods for other countries we used the Dutch estimates for all studies.

The value of production loss is assumed to be 80% of the average value of a worker's production, taking into account that absence generally causes a less than proportional decrease in production (see chapter 8). The average value of production per worker by age and sex was estimated for the Netherlands for 1988 and 1990 (see appendix 8). For Sweden and the U.S.A. we used the estimates as reported in the original studies.

Health care programs

Breast cancer screening

A national breast cancer screening program in the Netherlands was analyzed by de Koning et al. (1990 and 1991). They assumed that during the period 1990-2017 all women of age 50-70 are invited every two years to be screened mammographically. Absence from work was assumed to occur for each working patient receiving primary therapy for invasive breast cancer and for each patient having a relapse or metastases. Because the treatment of these patients is burdensome and prolonged, all absence spells are assumed to be 3 months or longer, implying that the period of production loss is equal to the length of the friction period.

Since breast cancer screening is directed to early detection of malignancies, screening will increase rather than reduce the number of patients getting primary therapy. A possible reduction in indirect costs may result from a decrease in the number of patients with recurrent or metastatic disease.

Cervical cancer screening

Regarding cervical cancer screening, indirect costs were calculated for a strategy of screening women aged 37-73 every 6 years during 1988-2015, which was demonstrated to be an efficient screening policy for the Netherlands (Koopmanschap et al., 1990). For each type of treatment a specific length of absence from work was assumed, based on observed length of hospital stay and expert opinion on the length of time for recovery at home (see appendix 9). For each disease stage the probabilities for specific treatments (van Ballegooijen et al., 1990) were combined with the number of patients and the length of the absence spells. For patients having recurrent or metastatic disease we assumed the duration of absence to be longer than the friction period.

Ace-inhibitors for congestive heart failure

In the analysis of prescribing ace-inhibitors for congestive heart failure for a period of ten years versus the standard treatment, patients were classified in four classes of disease severity: New York Heart Association (NYHA) I-IV (van Hout et al., 1993). Patients in NYHA class I were assumed not to be absent from work, but if a patient enters a higher NYHA-class (i.e. the heart function deteriorates) productivity loss of an entire friction period is assumed to occur. Consequently, production gains due to ace-inhibitors were assumed to arise if the health status of working patients improved significantly, implying the entrance to a lower NYHA-class.

Cimetidine versus surgery in duodenal ulcer
The costs of cimetidine versus surgery in duodenal ulcer were compared by several authors (Jönsson and Carlsson, 1991; Culyer and Maynard, 1981; Bulthuis, 1984). Both Jönsson et al. and Culyer et al. found a significant reduction in the length of absence from work in the first year of the therapy. Jönsson et al. estimated an average reduction of 48 days, resembling the estimate of 5-8 weeks as found by Culyer et al. Jönsson et al. did not report the total frequency distribution of the reduction in absence. However, since 48 days is considerably shorter than the friction period (about 3 months) we assume that the entire reduction of absence represents a gain in production for working patients.

Asthma/COPD
For 182 patients aged 18-60 years having asthma or COPD, beta-2-agonist/corticosteroid therapy (BA+CS) was compared to beta-2-agonist/placebo therapy (BA+PL) during 2.5 years in a randomized controlled study (Kerstjens et al., 1992; Rutten-van Mölken et al., 1994)). Each 3 months the treatment effect on lung function was measured in Forced Expiratory Volume in 1 second (FEV_1) as a % of the predicted value.

During the first year BA+CS caused significantly more improvement in FEV_1 than BA+PL and the average number of symptom free days per patient-year was significantly higher for BA+CS. The mean number of restricted activity days per patient-year for BA+CS-patients was 5.2, compared to 10 days for patients receiving BA+PL. However, the reduction in restricted activity days was not statistically significant. Since our calculations serve as an illustration, we used this reduction in the number of restricted activity days to calculate the indirect benefits of BA+CS versus BA+PL. Because it was unknown how many patients had a paid job, we assumed the proportion of paid workers to be equal to that in the general population aged 18 to 60.

Hepatitis B vaccination
The costs and effects of several strategies concerning hepatitis B vaccination were estimated for the Netherlands (Eelkman Rooda et al., 1992). Vaccination of all newborns was compared to the current Dutch strategy: screening pregnant women and vaccinating babies of the HBsAG-positive mothers. Costs and health effects (lives and life-years) were calculated for the period 1990-2090. For each acute and chronic disease state, the authors assumed the risk of being absent from work, the specific length of absence and the risk of getting disabled.

Intravenous thrombolytic therapy in MI
Costs and quality of life for intravenous thrombolytic therapy (rt-PA, n=153) were compared to placebo treatment (n=160) in the first year after a suspected myocardial infarction (Levin and Jönsson, 1992). The 12-month mortality of rt-PA patients was 8.5% as compared to 15.6% in the placebo-group (p=0.05). No significant difference in quality of life was observed. In the long run, the gain in life expectancy due to rt-PA, given an annual mortality rate of 6%, was 1.25 years.

After one year 70% of the rt-PA patients and 54% of the placebo patients had returned to work (p=0.06). After 1 and 6 months the percentage of rt-PA patients returning to work is also higher, but the authors don't give information on the statistical significance. The average time of absence from work was 69.7 days for rt-PA patients and 81.1 days in the placebo group. This difference was not statistically significant. We used the exact frequency distribution of absence durations for both treatment arms in calculating the indirect costs of rt-PA compared to placebo.

Total hip arthroplasty
Cushner and Friedman investigated the costs of total hip arthroplasty (THA) in the U.S.A. for patients having osteoarthritis, fracture-dislocation, rheumatoid arthritis and osteonecrosis (Cushner and Friedman, 1988). They estimate the annual direct costs of THA, including preoperative care, surgery and six-month follow up to be US $ 4.6 billion. The indirect costs during hospital stay were US $ 20 million, including the value of household work. During the six month rehabilitation period indirect costs were estimated as being US $ 221 million, using the human capital method taking into account all potential losses and gains in production. The authors also calculated the indirect benefits after rehabilitation. These are twofold: formerly disabled may resume working,

saving US $ 553 million and the prevention of disability would save US $ 1300 million. The authors conclude that the total costs of THA are about US $ 3 billion (direct costs $ 4.6 billion, indirect costs $ 0.2 billion, indirect benefits $ 1.8 billion).

Results

For each health care program the direct costs and health effects are presented, followed by the impact of including indirect costs, see table 9.2.

Table 9.2 Percentual impact of including indirect costs on total costs or cost-effectiveness of eight health care programs.

Program	Outcome parameter	% Impact
Breast cancer screening	Cost per life-year gained	+ 2 %
Cervical cancer screening	Cost per life-year gained	+ 2 %
Ace-inhibitors for CHF	Savings (and life-years gained)	+ 1 %
Cimetidine vs surgery in peptic ulcer	Total costs	- 18%
BA+CS vs BA+PL in asthma/COPD	Cost per 10% FEV_1 increase	Savings instead of costs
Hepatitis B vaccination	Savings	+ 52 %
Thrombolitics after MI	Cost per life-year gained	- 16 %
Total hip arthroplasty	Total costs	+ 2 %

Breast cancer screening
The total incremental direct costs of breast cancer screening over the period 1990-2017 amount to 466 million Dutch guilders (Dfl) (discount rate 5%). In the long term the screening program saves 700 lives per year, reducing breast cancer mortality by 16% (de Koning et al., 1990 and 1991). The costs per life-year gained are Dfl 7,650. Including production losses raises total costs by Dfl 9 million, a 2% increase. The costs per life-year gained amount to Dfl 7,790. The increase in costs is mainly caused by advancing the moment of

treatment and a slight increase in the number of patients having primary therapy. On the other hand, production gains for patients not having metastases are limited, because these patients are often older than age 65, the age of retirement. The impact of including indirect costs is very modest.

Cervical cancer screening
For cervical cancer screening the story is analogous. Including production losses raises total incremental costs over the period 1988-2015 by Dfl 6 million to Dfl 330 million, a 2% increase. Again this cost increase represents a balance of extra costs and savings. The screening program causes Dfl 16 million extra indirect costs for women with pre-invasive lesions, whereas the indirect costs for invasive cancer and metastases decrease by Dfl 10 million, a 20% reduction. The costs per life-year gained slightly increase from Dfl 24,400 to Dfl. 24,800.

ACE-inhibitors for congestive heart failure
In the first ten years of subscribing ACE-inhibitors for men having congestive heart failure the total number of life years gained is estimated to be almost 7000 (van Hout et al., 1993). Despite the additional annual costs of prescribing ACE-inhibitors of Dfl 8 million, total costs of congestive heart failure (including drugs) decrease by Dfl 32 million per year. Indirect costs fall about Dfl 200,000, which is less than 1% of the expected savings per year.

Cimetidine versus surgery in duodenal ulcer
The annual incremental direct costs of cimetidine as compared to surgery in duodenal ulcer amount to 28 million Swedish kronor (SEK) in 1983. The introduction of cimetidine induced SEK 39.6 million additional drug costs, but it saved SEK 11.6 million costs of hospital care. The savings due to a reduction in absence from work amount to SEK 5 million. The total annual incremental costs of cimetidine versus surgery are SEK 23 million (=28-5), thus including the indirect benefits makes a difference of 18%.

Asthma/COPD
For patients having asthma or COPD, the additional direct costs of BA+CS (versus BA+PL) were estimated to be 427 Dutch guilders per patient-year. The difference in lung function, measured in FEV_1 is 10% (Kerstjens et al., 1992; Rutten-van Mölken et al., 1994). Estimating a reduction in restricted activity days of 4.8 days, the indirect benefits amount to 500 guilders per

patient-year. Hence, including the indirect benefits of BA+CS therapy would result in 75 guilders net savings per patient-year.

Hepatitis B vaccination
Vaccination of all newborns against Hepatitis B in the Netherlands will save about 91000 life-years and Dfl 1325 million direct costs (not discounted) over the period 1990-2090, assuming that the vaccine costs 90 Dutch guilders per unit. Since the effects of vaccination take a long time to effectuate, the first savings will prevail in the year 2020. Indirect benefits, according to the friction cost method, make up for an additional saving of Dfl 675 million, resulting in total savings of Dfl 2 billion. Hence, including indirect benefits increases savings by 52%. The costs of a vaccine are not known yet. If it would cost Dfl. 135 instead of Dfl. 90, the savings in direct costs would be Dfl 490 million. Including the indirect benefits of Dfl 675 million would then increase savings by 138%.

Intravenous thrombolitics in MI
The average direct costs per patient of intravenous thrombolitics during the first year after myocardial infarction amount to SEK 39,400, a significant increase of SEK 9,880 as compared to the placebo treatment ($p=0.03$). The reduced absence from work induces an average savings in indirect costs of SEK 1,560 per patient. As a result, the total additional costs of thrombolitics per patient during the first year amount to SEK 8320. In the long run, the direct costs per life-year gained (discount rate 5%) are SEK 14,400, whereas the total costs (direct and indirect) per life-year gained amount to SEK 12,100. Including indirect costs improves the cost-effectiveness by 16%.

Total hip arthroplasty
The costs of total hip arthroplasty would change considerably, when the indirect costs are calculated according to the friction cost method. The indirect costs during hospital stay would remain unchanged. However, since the rehabilitation after THA takes 6 months, this would exceed the friction period (2.8 months in 1988) and the indirect costs would be about half the original estimate: US $ 100 million. Indirect benefits in the long term would be non-existent, using the friction cost method. Total costs of THA amount to US $ 4.7 billion, only 2% more than Cushner and Friedman's estimate of direct costs ($ 4.6 billion). On the other hand, the estimate of $ 4.7 billion is about 55% higher than Cushner's estimate of total costs ($ 3 billion), because

Cushner assumed substantial savings in indirect costs.

Analysis of impact of indirect costs

It is interesting to see the variation across health care programs in the impact of including indirect costs. The factors underlying this variation can be grouped in three categories: factors related to the type of program, factors related to the specific disease and characteristics of the specific group aimed at.

Table 9.3 presents some relevant characteristics of the eight health care programs. It clearly shows the special position of preventive programs. The (indirect) benefits of prevention take time to effectuate. For cancer screening this implies that future benefits are limited due to discounting and retirement of patients before indirect benefits may occur. In addition, screening advances morbidity and absence from work due to early detection of lesions. In the absence of screening for cervical cancer, a part of the non-invasive lesions would regress spontaneously (van Ballegooijen et al,. 1990).

In case of hepatitis B vaccination the time lag for health effects and savings in indirect costs is very long. Standard discounting would make the program worthless, but the authors defensibly argue that health benefits to be reaped by future generations should not be discounted in the standard way (Eelkman Rooda et al., 1992; Krahn and Gafni, 1993). The group to benefit from vaccination is fairly young, improving the scope for possible indirect benefits.

For the treatment programs the bulk of effects materialises within the first year of the program. The indirect benefits of ACE-inhibitors are marginal still, since only a small fraction of the patients has a paid job. Total hip arthroplasty causes additional absence due to rehabilitation following surgery, reducing indirect benefits. The latter are relatively modest, since many patients are retired.

The programs for ulcer, asthma and myocardial infarction cause instantaneous health effects, they aim at all age-groups and induce no additional absence from work, which are all favourable conditions for indirect benefits to be substantial.

Table 9.3 Main characteristics of eight health care programs and impact of including indirect costs on outcomes.

Program	Program type	Effect period	Group to benefit	Absence induced	Impact
Breast cancer screening	Secondary prevention	>5 years	women age > 55	yes	small
Cervical cancer screening	Secondary prevention	>10 years	women age > 45	yes	small
Ace-inhibitors in CHF	Treatment	< 1 year	age > 55	no	small
Cimetidine in peptic ulcer	Treatment	< 1 year	all ages	no	large
BA+CS in asthma/ COPD	Treatment	< 1 year	all ages	no	large
Hepatitis B vaccination	Primary prevention	>15 years	all ages	no	large
Thrombolitics after MI	Treatment	< 1 year	age > 40	no	large
Total hip arthroplasty	Treatment	>0.5 year	age > 55	yes	small

Table 9.4 summarizes the data on absence, disability and mortality for the relevant diseases. Following the friction cost method (short term) absence is the main cause of indirect costs/benefits. Absence shorter than three months is relatively important for ulcer, asthma/COPD and hepatitis B, whereas absence for myocardial infarction and breast cancer is predominantly long term absence. The relative high mortality for myocardial infarction and breast cancer yields only a small impact on the indirect benefits of the programs related to these diseases.

Table 9.4 Relative number of cases registered for (short term) absence from work, incidence of disability and mortality in the Netherlands for the diseases related to eight health care programs. Index for number of absent cases = 1.

Disease	Absence <3 months	Total absence	Disability incidence	Mortality
Breast cancer	0.13	1	0.38	2.96
Cervical cancer	0.41	1	0.53	5.71
Congestive heart failure	0.65	1	0.11	3.31
Duodenal ulcer	0.75	1	0.03	0.08
Asthma/COPD	0.86	1	0.06	0.23
Myocardial infarction	0.09	1	0.46	4.19
Hepatitis B	0.87	1	NA	0.05

Sources: absence from work (except hepatitis B) (CBS, 1989c); absence from work for Hepatitis B (Eelkman Rooda et al., 1992); incidence of disability (GMD, 1990); mortality (CBS, 1989b).

Discussion

Some authors have objections against including indirect costs in economic evaluation. Inclusion may favour health care interventions directed to (well) paid workers, which may conflict with equity considerations, especially when the results of economic studies are used to support decisions at the level of individual patient treatment. These objections could be met by excluding indirect costs from economic evaluations. However, this would imply denying a significant part of economic reality, since indirect costs of disease do contribute to the scarcity of resources and hence decrease society's wealth.

Indirect costs may be presented together with health care costs, health effects, consequences for equity and ethical and legal considerations, but they should be separately identifiable. It is up to the decision-makers to decide on the relative weight that they want to attach to indirect costs, compared to other criteria.

Recently, Gerard and Mooney argued that including non health care costs in cost-utility analysis is problematical (Gerard and Mooney, 1993). According to them, the opportunity costs of health care resources is defined in terms only of the QALY's forgone. Hence, cost utility analysis is about generating QALY's and QALY's foregone by opportunities sacrificed. This would preclude considering non health care costs/savings, which could only be considered in cost-benefit analysis. These authors interpret the concept of opportunity costs in a rather narrow sense. Clearly, when program A requires a similar health care budget as program B, but uses less resources than program B from a societal perspective, the additional resources becoming available when applying program A as compared to B, could be added to the health care budget (or added to budgets for education, housing and working conditions, which may also contribute to health status), and produce QALY's. Therefore, the opportunity costs in terms of QALY's foregone of program B would be larger than those of program A. So there is no specific requirement to use cost-benefit analysis, invoking all its valuation problems, when considering non health care costs in an economic evaluation.

Economic evaluations should preferably also include the indirect costs of unpaid labour, but data on the absence and disability related to unpaid labour are very scarce. With respect to breast cancer screening, a rough calculation showed that including unpaid work implies a small increase in the indirect costs of Dfl 1.6 million, leaving overall results unaltered (de Koning et al., 1990). For other health care programs, the impact of indirect costs of unpaid work may be more substantial.

With respect to the total costs of total hip arthroplasty, Cushner and Friedman indicated the maximum potential indirect benefits on the long term: after 6 months rehabilitation many formerly disabled are assumed to resume working. In addition, due to THA 80% of the working patients would not become disabled, but would remain employed after 6 months rehabilitation. If unemployment is above the level of frictional unemployment, which is the current situation for Western Europe and North America, substantial long term indirect benefits of preventing disability and mortality cannot be expected. On the other hand, a reduction in the incidence of disability may have some beneficial macro-economic implications. Lower incidence may increase labour supply, which results in lower labour costs per unit of output, an increase in production, exports and national income (see chapter 8).

The influence of including indirect costs on the results of economic evaluations may vary substantially. Generally speaking, a significant effort should be made to estimate the indirect costs when health care programs bring forth health effects in the short run, when there is a major impact on (short term) absence from work and if a significant proportion of the target population is employed at the moment of receiving the benefits from the program. The possible inducement of treatment related absence from work and disability should also be considered.

If health care programs have a considerable impact on permanent disability and mortality, estimates of indirect costs according to the human capital approach only illustrate the program's potential economic impact, which may overestimate the true economic consequences to a considerable extent (see chapter 7 and 8).

10 Conclusions

This chapter presents the main conclusions that can be drawn with respect to the key questions of this thesis: the use of detailed and more aggregate economic evaluations of health care and the role of indirect costs in economic evaluation.

Detailed analysis

The cost-effectiveness analysis of cervical cancer screening as described above illustrates the merits of a detailed economic evaluation of an individual health care program. The analysis permitted the selection of technically efficient screening policies, characterized by a broad age range for women to be invited (age 30-70) and a relatively large interval between invitations: at least 5 years.

The study provided insight in the relation between cost-effectiveness and the size of the screening program. The costs per life-year gained are reasonable for efficient policies with 10 or less invitations in a lifetime. The marginal cost-effectiveness deteriorates sharply if women are invited more than 10 times, illustrating the economic 'law of diminishing returns'.

The attendance rate is crucial for the cost-effectiveness of cervical cancer screening as women, who are less likely to attend, tend to carry a higher risk. Since the organisation of the screening program appears to influence attendance considerably, investments in a sound organisation of screening will produce good value for money.

The methodology of this type of detailed cost-effectiveness analysis is reasonably developed and standardized. The analysis is capable of providing relevant information for health policy, witness the recent decision of the Dutch government to extend the interval between successive invitations from 3 to 5 years and to broaden the age-range from age 35-53 to age 30-60, corresponding to the study recommendations. On the other hand, the current non-uniform screening organisation without a reminder invitation appears unfavourable with respect to attendance and cost-effectiveness.

Aggregate analysis

The aggregate cost of illness study described shows that for the Netherlands and Sweden the health care costs of chronic non-fatal disease predominate, whereas in the United States this pattern is less pronounced. In addition to comparing costs between countries, a cost of illness study can provide more insight in the consequences of differences between countries in epidemiology, health care system and health policy. However, a thorough explanation of the differences in costs between countries requires more comparable epidemiological data.

During 1988-2030, ageing will be responsible for an annual increase in total health care cost of 0.7%point in the Netherlands, assuming constant costs per caput (age- and sex-specific), as a first baseline scenario. With respect to cancer, the estimated annual cost increase due to ageing is 1.5%point. Health care costs per caput for the elderly may increase faster than for younger people, due to compression of disease within the last years of life, combined with a tendency towards more extensive diagnostic and therapeutic interventions in old and very old patients. Hence, longitudinal analysis of health care costs for elderly appears to be the next step in refining the estimates of the costs of ageing.

The future costs of cancer also depend on the expected epidemiological development. According to the cancer disease model, cancer costs will increase sharply if incidence rises, whereas improved disease-free survival has a moderate impact on costs. Insight in the expected epidemiological trends is indispensable for an accurate projection of future costs of cancer.

The cost analyses as described are not full economic evaluations, but these still may be useful for health policy. Cost estimates incorporated in a disease model (as the cancer model described) can show the likely consequences of expected trends in demography and epidemiology. The model should preferably take into account the cost consequences of competing diseases and the costs of comorbidity, especially in elderly patients. If such a model is supplemented with data on effectiveness of the relevant health care programs, it could generate an indication of the cost-effectiveness of a total set of health care programs directed to a particular disease or specific subsectors of health care.

In sum, aggregate economic evaluation can serve several purposes. It may facilitate estimating the combined results of interacting programs for the same disease and the consequences of health care programs for the costs and epidemiology of other diseases. It can produce estimates of the future costs, morbidity and mortality of diseases for demographic, epidemiological and technological scenarios. Finally, aggregate estimates of cost-effectiveness may select candidates for more detailed analysis in specific parts of health care that need closer investigation of costs and health effects. In this respect, detailed and aggregate economic evaluation are complementary instruments for health policy.

For practical application of the aggregate analysis, a number of methodological problems remain to be solved. One of the most tough issues is the estimation of the aggregate effectiveness of existing interventions, requiring knowledge about the health status of patients in the situation with and *without* the existing health care programs.

Indirect costs

Indirect non-medical costs of disease as an element in economic evaluation are controversial. According to some authors, cost-effectiveness/-utility analysis is restricted to generating health effects/Qaly's within a given health care budget, implying that indirect costs of disease should not be included in the evaluation. This is a rather narrow approach, since indirect costs influence the scarcity of resources and hence society's wealth. In addition, indirect costs may have an impact on the budget for health care and other budgets which may contribute to health status. In this respect direct and indirect costs are alike and the latter should therefore be included in economic evaluations.

In order to present a balanced overview of the consequences of health care programs, estimates of indirect costs may be presented together with, but separate from, health care costs, health effects, consequences for equity and ethical and legal considerations. It is up to the decision-makers to decide on the relative weight that they want to attach to indirect costs, compared to other criteria.

Given the relevance of indirect costs, it seems preferable to insist that indirect costs are estimated as realisticly as possible, instead of estimating the potential economic consequences of disease in an exceptional state of full employment equilibrium, according to the human capital approach. In estimating indirect costs the following issues should be analyzed:
- the short term impact of absence from work on production and costs;
- the time-span during which production and costs are affected directly,

depending on the situation within the firm as well as on the labour market;
- the medium term macro-economic consequences of absence from work and disability.

For the Netherlands the short term friction costs in 1990 amounted to 1.5-2.5% of net national income, depending on the extent to which short term absence from work induces production loss and costs. The medium term macro-economic consequences of absence from work and incidence of disability were estimated to reduce net national income by an additional 0.8%.

The exact consequences of short term absence from work for production and costs need to be analyzed in more detail on the level of the firm. Such analysis should focus on the possibilities of postponing work and making up for it later, the nature and extent of internal labour reserves versus flexible labour and the diminishing returns to labour for the individual employee.

An *attainable reduction* of absence from work (from 8% to 6% of working time) and incidence of disability (-20%) will save about 0.5 % NNI friction costs and 0.18% NNI in the medium term. In order to improve the estimates of indirect costs, further research is necessary to estimate to what extent absence from work and disability can be prevented and the possible role of competing diseases which may decrease the potential for reduction in absence and disability.

In general as well as in this thesis, estimates of indirect costs of disease only refer to paid labour. An important point for further research is the amount and value of indirect costs due to unpaid labour, according to the friction cost method. However, reliable data on the amount of absence and disability related to unpaid labour are rather scarce and the issue of the value of unpaid production needs to be analyzed further.

In addition to the aggregate estimates of indirect costs of disease, the impact of indirect costs on cost-effectiveness of individual health care programs was analyzed. The influence of including indirect costs on the results of particular health care programs depends on the time lag between investments and health effects, the relative importance of (short term) absence and the labour participation of the population to benefit. For cervical cancer screening the impact is rather small. For therapeutic interventions reducing absence from work, like cimetidine for peptic ulcer and thrombolitics after myocardial infarction the savings in indirect costs appear to be substantial.

A significant effort should be made to estimate the indirect costs when health care programs bring forth health effects in the short run, when there is a major impact on (short term) absence from work and when a significant proportion of the target population is employed at the moment of receiving the benefits from the program. The possible inducement of treatment related absence from work and disability should also be considered.

Literature

Aarts LJM, Jong PR de. Economic aspects of disability behaviour, Dissertation Erasmus University Rotterdam. Rotterdam, 1990

Anderson GH, Boyes DA, Benedet JL et al. Organisation and Results of the Cervical Cytology Screening Programme in British Columbia, 1955-85. BMJ 1988;296:975-978.

Armstrong BK, Rouse IL, Butler TL Cervical cytology in Western Australia. Med J Aust 1986;144:-239-247.

Baker MS, Kessler LG, Smucker RC. Site-specific treatment costs for cancer: an analysis of the medicare continuous history sample file, in: Scheffler RM, Andrews NC (eds). Cancer Cost Issues: 127-138. Health Administration Press Perspectives. Ann Arbor Michigan, 1989.

Ballegooijen M van, Boer R, Oortmarssen GJ van, Koopmanschap MA, Lubbe JThN, Habbema JDF. Cervical cancer screening: age ranges and intervals; An updated cost-effectiveness analysis (in Dutch). Institute for Public Health. Erasmus University Rotterdam. Rotterdam, 1993.

Ballegooijen M van, Koopmanschap MA, Oortmarssen GJ van, Habbema JDF, Lubbe JThN, Agt HME van. Diagnostic and treatment procedures induced by cervical cancer screening. Eur J Cancer 1990;26:941-5.

Ballegooijen M van. Diagnosis and treatment of cervical cancer (in Dutch) in: Habbema et al. 1988: B1-B52.

Ballegooijen M van, Koopmanschap MA, Subandono A, van Oortmarssen GJ. Care and costs for advanced cervical cancer. Eur J Cancer 1992; 28a: 1703-1708.

Banta D, Bos M. The relation between quantity and quality with coronary artery bypass grafting. Health Policy 1991;18:1-10.

Baumol W. Productivity policy and the service sector. In: Inman RP, editor. Managing the Service Economy, prospects and problems. Cambridge: Cambridge University Press, 1985.

Beeck EF van, Mackenbach JP, Oortmarssen GJ van, Barendregt JJM, Habbema JDF, Maas PJ. Scenarios for the future development of accident mortality in the Netherlands. Health Policy 1989;11:1-17.

Berg P van der. Further specification of features of the Europe-scenario (in Dutch). Central Planning Bureau. 's-Gravenhage, 1991.

Berget A. Influence of Population Screening on Morbidity and Mortality of Cancer of the Uterine Cervix in Maribo Amt. Dan Med Bull 1979;26:91-100.

Berk A, Paringer L, Mushkin S. The economic cost of illness, Fiscal 1975. Medical care 1978;16:7-85-790.

Berrino F, Chiappa L, Oliverio S, Todeschin P, Turolla E, Vegetti P. Study of Women Who Did Not Respond to Screening for Cervical Cancer. Tumori 1979;65:143-155.

Birch S, Gafni A. Cost effectiveness/utility analyses. Do current decision rules lead us where we want to be? Journal of Health Economics 1992: 11: 279-296.

Black Report. Report of the working group on bone marrow transplantation. Chairman Sir Douglas Black. London, 1982.

Bloemhoff A, Winter CR de. The influence of socio-economic status on disability, a longitudinal analysis (in Dutch), NIPG/TNO. Leiden, 1991.

Bodkin RG, Klein LR, Marwah K. A history of macroeconometric model-building. Edward Elgar, Brookfield, 1991.

Boice JD, Storm HS, Flannery JT, et al. Multiple primary cancers in Connecticut and Denmark. Natl Cancer Inst Monogr 1986;68:13-244.

Bonneux L, Barendregt JJ, Bruinvels DJ, Looman CWN, Bonsel GJ, van der Maas PJ. Diverging trends in colorectal cancer morbidity and mortality. Earlier diagnosis comes at a price. Submitted, 1994.

Bonneux L, Barendregt JJ. Towards a comprehensive health policy model, in: Morgenstern W, Chigan E, Prokhorskas R, Rusnak M and Schettler G (eds). Models of noncommunicable diseases: 11-18. Springer Verlag. Berlin/Heidelberg, 1992.

Boyes DA, Morrison B, Knox EG, Draper GJ, Miller AB. A Cohort Study of Cervical Cancer Screening in British Columbia. Clin Invest Med 1982;5:1-29.

Bried EM, Scheffler RM. The financial stages of cancer in the elderly. Oncology 1989;6S:153-160.

Brindle G, Higham E, Roberts DJ, Wakefield J, Yule R. Response to a Mobile Cervical Screening Unit in an Industrial Area of England. Public Health (London) 1976;90:165-169.

Bulthuis R. Cimetidine and the cost of peptic ulcer in the Netherlands. Effective Health Care 1984;1(6):297-309.

Bureau D'Informations et de Previsions Economiques (BIPE). The economics of ulcers in France. Neuilly, Seine Cedex, 1978.

Cancer Registry of Norway. Survival of Cancer Patients, cases diagnosed in Norway 1986-1974. Norwegian Cancer registry. Oslo, 1980.

Casparie AF, Hoogendoorn D. Effects of Budgeting on Health Care Services in Dutch Hospitals. Am J Public Health 1991;81:1442-7.

CBS (Central Bureau for Statistics). Demographic forecast, medium variant. Monthly Statistics on Demography, december. Voorburg, 1989.

CBS. Data on absence from work by diagnosis, age and sex, from the Diagnosis Statistics 1988. 's-Gravenhage, 1989c.

CBS. Diagnose Statistics (on absence from work) 1987. 's-Gravenhage, 1988b.

CBS. Socio-economic Monthstatistics 1988, 's-Gravenhage, 1988a.

CBS. Socio-economic Monthstatistics 1989. 's-Gravenhage, 1989a.

CBS. Socio-economic Monthstatistics 1990. 's-Gravenhage, 1990a.

CBS. Socio-economic Monthstatistics 1991. 's-Gravenhage, 1991c.

CBS. Mortality by cause of death, age and sex 1988. Health Monthstatistics. 's-Gravenhage, 1989b.

CBS. Labour Force Survey 1988. 's-Gravenhage, 1989d.

CBS. Labour Force Survey 1990. 's-Gravenhage, 1991b.

CBS. Mortality by cause of death, age and sex 1990. Health Monthstatistics. 's-Gravenhage, 1991a.

CBS. Data on absence from work by diagnosis, age and sex, from the Diagnose Statistics 1990. 's-Gravenhage, 1991.

CBS. National Health Interview Survey 1981-1985. 's-Gravenhage, 1988.

CBS. Elderly in the Netherlands by family status and age group, per sex. Monthly statistics on demography (february). Voorburg, 1990.

Center for Health Policy Research. The costs of medical professional liability in the 1980s. Chicago, American Medical Association, 1990.

Children's Defend Fund. Who is watching our children's health: immunization status of American children. Washington DC, 1987.

Choi NW, Nelson NA. Results from a Cervical Cancer Screening Programme in Manitoba, Canada.. In: M. Hakama, AB Miller & NE Day., eds. Screening for Cancer of the Uterine Cervix. Lyon: International Agency for Research on Cancer 1986:61-67.

Coebergh JWW, van der Heijden LH (eds). Cancer incidence and survival 1975-1987. Eindhoven Cancer Registry. Eindhoven, 1991.

Cole PH, Morrison AS. Basic issues in population screening for cancer. J Natl Canc Inst 1980;64:12-63-72.

COBA: Second report: standards for cost benefit analysis (in Dutch). Dutch government: department of Finance. 's-Gravenhage, 1975.

Commonwealth of Australia. Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including submissions involving economic analyses. Department of Health, Housing and Community Services. Woden, 1990.

Congressional Budget Office. Rising health care costs: Causes, Implications and Strategies. Government Printing Office. Washington DC, 1991.

Cooper BS, Rice DP. The Economic Cost of Illness Revisited. Soc Secur Bull 1976;39:21-36.

CPB (Central Planning Bureau). Development of productivity in the health care sector; defining concepts and quantification for the period 1981-1985. 's-Gravenhage, 1987.

CPB. FKSEC: a macro-econometric model for the Netherlands. Stenfert Kroese, Leiden/Antwerpen,1992a.

CPB. Uncertainties on labour time reduction in the medium term (in Dutch), Working document 14. 's-Gravenhage, 1987a.

CPB. Central Economic Plan 1989. 's-Gravenhage, 1989.

CPB. Central Economic Plan 1992. 's-Gravenhage, 1992.

Culyer AJ. The normative economics of health care finance and provision, in: McGuire A, Fenn P and Mayhew K. (eds.). Providing health care: the economics of alternative systems of finance and delivery. Oxford University Press. Oxford, 1991.

Culyer AJ, Maynard AK. Cost-effectiveness of duodenal ulcer treatment. Soc Sci Med 1981;15: 3-11.

Cushner F, Friedman RJ. Economic impact of total hip arthroplasty. Southern Medical Journal 1988;81:1379-81.

Day NE. Effect of cervical cancer screening in Scandinavia. Obstetrics and Gynecology 1984; 63: 714-718.

Department of Social Affairs and Employment. Labour market report 1989 (in Dutch). 's-Gravenhage, 1989.

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.

Doornbos G, Kromhout D. Educational level and mortality in a 32-year follow-up study of 18 year old men in the Netherlands, International Journal of Epidemiology 1990;19:374-379.

Draper G. Screening for Cervical Cancer: revised policy. The recommendations of the DHSS committee on gynaecological cytology. Br J Family Planning 1982;8:95-100.

Drummond MF. Cost-of-illness studies: a major headache? PharmacoEconomics 1992;2:1-4.

Drummond MF, Torrance G, Mason J. Cost-effectiveness league tables: more harm than good? Soc Sci Med 1993;37:33-40.

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

Drummond MF. Australian guidelines for cost-effectiveness studies of pharmaceuticals: The thin end of the boomerang? Centre for Health Economics Discussion Paper 88. University of York. York, 1991.

Drummond MF, Ludbrook A, Lowson K, Steele A. Studies in economic appraisal in health care. Volume 2, Oxford University Press. New York, 1986.

Dupont A. 140 years of Danish studies on the prevalence of mental retardation. Acta Psychiatrica Scand (Supplement) 1989;348(8):105-112.

Eardley A, Elkind AK, Spencer B, Hobbs P, Pendleton LL, Haran D. Attendance for Cervical Screening -Whose Problem?. Soc Sci Med 1985;20(9):955-962.

Ebeling K, Nischan P. Organization and Results of Cervical Cancer Screening in the German Democratic Republic. In: Hakama M, Miller AB and Day NE (eds). Screening for Cancer of the Uterine Cervix. Lyon: International Agency for Research on Cancer 1986:251-266.

Eelkman Rooda M, Barendregt J, Hattum J van, Hout BA van. Vaccination for hepatitis B. Costs and effects of eradiction strategies (in Dutch). Institute for Medical Technology Assessment. Erasmus University Rotterdam. Rotterdam, 1992.

Elfring T, Kloosterman RC.. The Dutch 'job machine' (in Dutch), ESB 1989;736-740.

Elkind AK, Haran D, Eardley A, Spencer B. Computer-managed cervical cytology screening: a pilot study of non-attenders. Public Health (London) 1987;101:253-266.

Emanuel-Vink H, Tax B. Final Report on the Analysis of effects of publicity on cervical cancer screening (in Dutch). Institute for Public Health. Nijmegen, 1979:1-50.

Personal communication. Estimates of earnings by the Dutch national associations for general practitioners, self employed specialists and salaried physicians. 1991.

EVAC (Evaluation Commission of Cervical Cancer Screening). Cervical Cancer Screening in the Pilot Regions Nijmegen, Rotterdam and Utrecht (in Dutch). Leidschendam, 1987.

Evans D, Hibbard B, Jones J, Landsman J, Sweetnam P. The Cardiff Cervical Cancer Cytology Study: Enumeration and definition of population and initial acceptance rates. J Epidem Comm Hlth 1980;34:9-13.

Finnish Foundation for Cancer Research. Cancer in Finland 1954-2008. Helsinki: Cancer Society of Finland publication No. 42, 1989.

Fitzgerald JM, Gafni A. A cost-effectiveness analysis of the routine use of isoniazid prophylaxis in patients with a positive mantoux skin test. Am Rev Respir 142:848-853,1990.

Gerard K, Donaldson C, Maynard AK. The cost of diabetes, Diabetic-Medicine 1989;6(2):164-170.

Gerard K, Mooney G. QALY league tables: handle with care. Health Economics 1993;2:59-64.

GMD (Mutual Medical Service). Statistical information 1988, Amsterdam, 1988.

GMD. Statistical information 1990. Amsterdam, 1991.

GMD. Data on incidence of disability by age, sex and type of employment for 1988. Amsterdam, 1990.

Graaf Y van der. Screening for Cervical Cancer, the Nijmegen Project. Academic dissertation. Katholic University Nijmegen, 1987.

Greenspan AM, Kay HR, Berger BC, Greenberg RM, Greenspon AJ, Spuhler Gaughan MJ. Incidence of unwarranted implantation of permanent cardiac pacemakers in a large medical population. N Eng J Med 1988;318:158-63.

Habbema JDF, Putten DJ van, Lubbe JThN, Maas PJ van der, Oortmarssen GJ van. The MISCAN Model: A new method for evaluation of mass screening. Part II: screening on breast cancer (in Dutch) Technical Report. Dept. of Public Health and Social Medicine, Erasmus University Rotterdam. Rotterdam, 1983.

Habbema JDF, Lubbe JThN, Agt HME van, Ballegooijen M van, Koopmanschap MA, Oortmarssen GJ van. Costs and effects of mass screening on cervical cancer (in Dutch). Erasmus University Rotterdam. Rotterdam, 1988.

Habbema JDF, Lubbe JThN, Oortmarssen GJ van, Maas PJ van der. A simulation approach to cost-effectiveness and cost-benefit calculations of screening for the early detection of disease. European Journal of Operational Research 1987;29:159-166.

Habbema JDF, Oortmarssen GJ van, Lubbe JThN, Maas PJ van der. The MISCAN simulation program for the evaluation of screening for disease. Computer Methods and Programs in Biomedicine 1984;20:79-83.

Hakama M, Chamberlain J, Day NE, Miller AB, Prorok PC. Evaluation of screening programmes for gynaecological cancer. Br J Cancer 1985;52:669-673.

Hakama M. Cervical Cancer: Risk Groups for Screening. In: M Hakama, AB Miller & NE Day, eds. Screening for Cancer of the Uterine Cervix. Lyon: Internat. Agency for Research on Cancer 1986:213-216.

Hakulinen T, Hakama M. The effect of screening of the incidence and mortality of cervical cancer in Finland. Nowotwory 1985;35:290-296.

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, 13.

Haran D, Hobbs P, Pendleton LL, Elkind A, Eardley A, Spencer B. A computer-managed call and recall system for cervical screening. Public Health (London) 1986;100:105-115.

Hartunian N, Smart N, Thompson M. The incidence and economic costs of major health impairments. Lexington Books. Lexington, 1981.

Haselen HWJ van. Production losses due to traffic accidents (in Dutch). Netherlands Economic Institute. Rotterdam, 1987.

Health Council. Advice concerning radiotherapy (in Dutch). s'Gravenhage 1984.

Henderson J, McKenzie L, Haiart D. Uptake of a Mobile Breast Screening Project: Implication for a Breast Screening Service. University of Aberdeen. Aberdeen, 1988.

Henke HD and Behrens CS. The economic costs of illness in the Federal Republic of Germany in the year 1980. Health Policy 1986;6:119-143.

Henry DA. The Australian guidelines for subsidisation of pharmaceuticals. PharmacoEconomics 1992;2:422-424.

Hodgson ThA, Kopstein AN. Health care expenditures for major diseases in 1980 (including unpublished appendix). Health Care Fin Rev 1984;4:1-12.

Horbeek J, Veenman J. Temporary agencies as educators (in Dutch). ESB 1993:943-945.

Hout van BA, Wielink G, Bonsel GJ, Rutten FFH. Effects of ACE inhibitors on heart failure in the Netherlands. A pharmacoeconomic model. PharmacoEconomics 1993;3:387-97.

Huisman R. A simulation model for the care for the elderly (in Dutch). University of Limburg. Maastricht, 1990.

Hunter L, McGregor A, MacInnes J, Sproull A. The 'Flexible Firm': Strategy and Segmentation, British Journal of Industrial Relations 1993;31:383-407.

Jansen JH, Kremers W. Mass Screening for Cervical Cancer: Change in Participation between 1978 and 1981 (in Dutch). Medisch Contact 1985;40:165-167.

Jendteg S (Swedish Institute for Health Economics). Personal communication, 1991.

Johannesson G, Geirsson G, Day N, Tulinius H. Screening for Cancer of the Uterine Cervix in Iceland 1965-1978. Acta Obstet Gynecol Scand 1982;61:199-203.

Jönsson B, Horisberger B, Bruguera M, Matter L. Cost-benefit analysis of hepatitis-b vaccination: a computerized decision model for Spain. International Journal of Technology Assessment in Health Care 1991;7:379-402.

Jönsson B, Carlsson P. The effect of cimetidine on the cost of ulcer disease in Sweden. Social Science and Medicine 1991;33: 275-82.

Jorm AF, Korten AE, Henderson AS. The prevalence of dementia: A quantitative integration of the literature. Acta Psychiatric Scand 1987;76:465-479.

Kahn KL, Kosecoff J, Chassin MR et al. The use and misuse of upper gastro-intestinal endoscopy. Ann Int Med 1988;109:664-70.

Kerstjens HAM, Brand PLP, Hughes MD, et al. A comparison of bronchodilator therapy with or without inhaled corticosteroid therapy in obstructive airways disease. N Eng J Med 1992; 327: 1413-9.

Kjellgren O. Mass screening in Sweden for cancer of the uterine cervix: effect on incidence and mortality. Gynecol Obstet Invest 1986;22:57-63.

Klaassen LH. Studygroup discounting: The role of the discount rate in policy-making (in Dutch). Dutch government, department of finance, 's-Gravenhage, 1985.

Kleinman JC, Kopstein A. Who is Being Screened for Cervical Cancer. Am J Public Health 1981;71:73-76.

Koning HJ de, van Ineveld BM, de Haes JCJM, van Oortmarssen GJ, Klijn JGM, van der Maas PJ. Advanced breast cancer and its prevention by screening. Br J Cancer 1992;65:950-955.

Koning de HJ, Ineveld BM van, Oortmarssen GJ van et al. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. Int J Cancer 1991;49:531-7.

Koning de HJ, Ineveld BM van, Oortmarssen van GJ et al. The costs and effects of breastcancer screening (in Dutch). Dept of Public Health. Erasmus University Rotterdam. Rotterdam, 1990.

Koning J de, Tuyl FAWM. The relation between labour time, production and employment (in Dutch). Netherlands Economic Institute. Rotterdam, 1984.

Koning J de. Volume and causes of labour hoarding (in Dutch). Erasmus University Rotterdam. Rotterdam, 1987.

Koopmanschap MA, Ineveld BM van. Towards a new approach for estimating indirect costs of disease, Social Science and Medicine 1992;34(9):1005-1010.

Koopmanschap MA, Roijen L van, Bonneux L. Cost of illness in the Netherlands (in Dutch). Department of Public Health and Social Medicine. Erasmus University Rotterdam. Rotterdam, 1991.

Koopmanschap MA, FFH Rutten. Indirect costs in economic studies: confronting the confusion. PharmacoEconomics 1993;4(6):446-454.

Koopmanschap MA, Rutten FFH, Ineveld BM van, Roijen L van. The friction cost method for measuring indirect costs of disease. J Health Economics 1994 (in press).

Koopmanschap MA, Lubbe JThN, Oortmarssen GJ van, Agt HME van, Ballegooijen M van, Habbema JDF. Economic Aspects of Cervical Cancer Screening. Soc Sci Med 1990;30(10):1081-1087.

Koopmanschap MA, Roijen L van, Bonneux L, Bonsel GJ, Rutten FFH, Maas PJ van der. Costs of diseases in international perspective. Eur J Public Health 1994 (in press).

Krahn M, Gafni A. Discounting in the economic evaluation of health care interventions. Medical Care 1993;31:403-18.

Laara E, Day N, Hakama M. Trends in Mortality from Cervical Cancer in the Nordic Countries: Association with Organised Screening Programmes. Lancet 1987;1247-1249.

Langmark F, Hagen S. Kreft (in Norsk). Kreftregisteret, August 1990.

Levin LA, Jönsson B. Cost-effectiveness of thrombolysis - a randomized study of intravenous rt-PA in suspected myocardial infarction. European Heart Journal 1992;13:2-8.

Lindgren B. The economic impact of illness, in: Abshagen U, Munnich FE (eds). Cost of illness and benefits of drug treatment, pp 12-20. Munich, 1990.

Lindgren B. Costs of illness in Sweden 1964-1975. Liber, Lund, 1981.

LTD (Ministry of Social Affairs and Employment). Data on gross wages by age, sex and education level for the Netherlands 1990. 's-Gravenhage, 1993.

Luce BR. The implications of cost effectiveness analysis of medical technology, an application to cervical cancer screening. Office for Technology Assessment (OTA) Congres. Washington, 1981.

Maas PJ van der, Koning HJ de, Ineveld BM van, et al. The Cost-effectiveness of Breast Cancer Screening. Int J Cancer 1989;43:1055-1060.

Maas PJ van der, Ineveld BM van, Oortmarssen GJ van, et al. The Costs and Effects of Mass Screening on Breast Cancer. Erasmus University Rotterdam. Rotterdam, 1988.

MacGregor JE, Moss S, Parkin DM, Day NE. Cervical Cancer Screening in North East Scotland, in: Hakama M, Miller AB and Day NE (eds). Screening for Cancer of the Uterine Cervix. Lyon: International Agency for Research on Cancer 1986:25-36.

Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-316.

McGuire A, Henderson J, Mooney G. The economics of health care, an introductory text. Routledge & Kegan Paul. London, 1988.

Mitchell DJB, Zaidi MA. Macroeconomic conditions and HMR-IR practice, Industrial Relations 1990;29:164-188.

Muller C, Caton C. Economic cost of schizophrenia: a postdischarge study. Medical Care 1983;21-:92-104.

National Hospital Institute. Financial statistics on hospitals. Utrecht, 1988.

Nepharma. Annual report 1988 (in Dutch). Utrecht, 1989.

Netherlands Cancer Registry. Incidence of cancer in the Netherlands, 1989. SIG Health Care Information. Utrecht, 1993.

NIA. Absence rates by age, sex and education level for 1988. Amsterdam, 1993.

OECD. Financing and delivering health care: a comparative analysis of OECD countries. Paris, 1987.

OECD. OECD Health Data. CREDES. Paris, 1991.

Office of Health Economics. Leukemia towards control. London, 1980.

Office of Technology Assessment. Cost-effectiveness of bone marrow transplant and its policy implications. U.S. Printing Office. Washington, 1981.

Olesen F. Use of Pap Smear in General Practice by Case Finding or Screening Programme. Scand J Prim Health Care 1986;4:59-62.

Ontario Ministry of Health. Guidelines for preparation of economic analysis to be included in submission to Drug Programs Branch for listing in the Ontario Drug Benefit Formulary/Comparative Drug Index. Ministry of Health. Toronto 1991.

OSA. OSA-Trendreport 1987 (in Dutch). 's-Gravenhage, 1987.

Osterhaus JT, Gutterman DL, Plachetka JR. Healthcare resource and lost labour costs of migraine headache in the US. PharmacoEconomics 1992;2:67-76.

Osterman P. Choice of employment systems in internal labor markets, Industrial Relations 1987;26-:46-67.

Ours J van, Ridder G. Cyclical variation in vacancy durations and vacancy flows, an empirical analysis, European Economic Review 1991;35:1143-1155.

Parkin DM, Moss SM. An evaluation of screening policies for cervical cancer in England and Wales using a simulation model. Journal of Epidemiology and Community Health 1986;40:143-153.

Parkin DM, Leach K, Cobb P, Clayden AD. Cervical Cytology Screening in Two Yorkshire Areas: Results of Testing. Public Health (London) 1982;96:3-14.

Prins R. Sickness absence in Belgium, Germany (Federal Republic) and the Netherlands, a comparative study. Netherlands Institute for Working Conditions (NIA). Amsterdam, 1990.

Reich M, Gordon D, Edwards R. A theory of labor market segmentation. American Economic Review 1973;63,2:359-365.

Rice DP. Demographic realities and projections of an aging population. In: Andreopoulos S and Hogness JR, editors. Health care for an aging society. New York: Churchill Livingstone, 1989.

Rice DP, Hodgson TA, Kopstein AN. The economic cost of illness: A replication and update. Health Care Fin Rev 1985;7:61-80.

Riley GF, Lubitz J, Prihoda R, Rabey E. The use and costs of Medicare services by cause of death. Inquiry 1987;24:233-244.

Riley GF, Lubitz J. Longitudinal patterns in Medicare costs for cancer decedents, in: Scheffler RM, Andrews NC (eds). Cancer Cost Issues, pp 89-106. Health Administration Press Perspectives. Ann Arbor Michigan, 1989.

Ritchie K, Kildea D, Robine JM. The relationship between age and the prevalence of senile dementia: A meta-analysis of recent data. Int J Epid 1992;21(4):763-769.

Robinson JC, Luft HS. The impact of hospital market structure on patient volume, average length of stay, and the cost of care. J of Health Economics 1985;4:333-56.

Roijen L van, Koopmanschap MA, Bonneux L. Costs of diseases (in Dutch). NtvG 1992:2:74-80.

Roodenburg HJ, Wong W. The labour market component in the disablement benefit (in Dutch). Central Office for Planning. 's-Gravenhage, 1985.

Russell LB. Evaluating preventive care. The Brookings Institution, Washington D.C., 1987.

Russell LB. Is prevention better than cure? The Brookings Institution. Washington D.C., 1986.

Rutqvist LE. On the utility of the lognormal model for analysis of breast cancer survival in Sweden 1961-73. Br J Cancer 1985;52:875-583.

Rutqvist LE, Wallgren A, Nilsson BO. Is Breast Cancer a Curable Disease? Cancer 1984;53:1793-1800.

Rutten FFH, Hout BA van. Examples of applications of medical technology assessment in health care for students in Health Policy and Management (unpublished) Erasmus University Rotterdam. Rotterdam, 1991.

Rutten-van Mölken MPMH, Doorslaer EKA van, Jansen MCC, Kerstjens HAM, Rutten FFH. Cost-effectiveness of inhaled corticosteroid and bronchodilator therapy in asthma and COPD. Submitted.

Sansom CD, MacInerney J, Oliver V, Wakefield J. Differential Response to Recall in a Cervical Screening Programme. Brit J prev soc Med 1975;29:40-47.

Schieber CJ, Poullier JP, Greenwald LM. Health Care systems in Twenty-Four Countries. Health Affairs 1991;(10)3: 22-38.

Schoenbaum SC et al. Benefit-cost analysis of rubella vaccination policy. N Eng J Med 1976;294:3-06-310.

Shiell A, Gerard K, Donaldson C. Cost of illness studies: an aid to decision-making? Health Policy 1987;8:317-323.

SIG (Information Centre for Health Care). Yearbook on hospital registration 1989. Utrecht, 1990.

SIG. Hospital data on admissions, first admissions, interventions and way of discharge for cancer patients in the Netherlands for 1978-1989. Utrecht, 1992.

SIG. Data on patient stays in non-psychiatric hospitals, by primary diagnosis, age-group and sex for the Netherlands 1988. Utrecht, 1990a.

Starfield B. Primary care and health; a cross-national comparison. JAMA 1991;266:2268-71.

Stehbens WE. An appraisal of the epidemic rise of coronary heart disease and its decline. The Lancet 1987;1(8533):606-611.

Teulings CN, Koopmanschap MA. An econometric model of crowding out of lower education levels. European Economic Review 1989;33:1653-1664.

Teulings CN, Koopmans CC, Koopmanschap MA, Vedder L. Redistribution of employment and social insurance. Foundation for Economic Research, University of Amsterdam, in charge of the Department of social affairs and employment (in Dutch). 's-Gravenhage, 1987.

Torrance GW. Measurement of health state utilities for economic appraisal. Journal of Health Economics 1986;5:1-30.

Tosteson ANA, Rosenthal DI, Melton J, Weinstein MC. Cost effectiveness of screening perimenopausal white women for osteoporosis: bone densitometry and hormone replacement therapy. Annals of Internal Medicine 1990;113:594-603.

United Nations (1987). Statistics of road traffic accidents in Europe. New York, 1987.

US Bureau of the Census. Current Population Reports, series P-25. Government Printing Office. Washington, 1989.

Verbeek ALM. Population Screening for Breast Cancer in Nijmegen: an Evaluation of the Period 1975-1982. Inst.Soc. Health, Catholic University. Nijmegen, 1985.

Vooijs GP. On behalf of the cytology commission of the Dutch pathologists association. Lettre to the Dutch pathologists. 1987.

Vrooman JC, Kemp AA de. Disablement and labour market (in Dutch). ESB 1990;3778:942-943.

Wakefield J. Studies of Response to Cervical Screening. Tumori 1976;62:315-318.

Waldo D, Sonnefeld S, McKusick D, Arnett R. Health expenditures by age group, 1977 and 1987. Health Care Fin Rev 1989;10(4):116-118.

WHO (World Health Organisation). World Health Statistics Annual 1989, 1990. Geneva, 1990, 1991.

Williams A. Economics of coronary artery bypass grafting. BMJ 1985;291:326-329.

Williams AF, Begley CE. The costs of cancer in Texas. Texas Medicine/the Journal 1992;6:62-67.

Williamson OE. The economics of organization: the transaction cost approach. American Journal of Sociology 1981;87:548-577.

Wilson A. Cervical cytology screening: a comparison of two call systems. BMJ 1987;295:181-182.

Winslow CM, Kosecoff JB, Chassin M, Kanouse DE, Brook RH. The Appropriateness of Performing Coronary Artery Bypass Surgery. JAMA 1988;260:505-9.

Woolhandler S, Himmelstein DU. The deteriorating administrative efficiency of the U.S. health care system. N Eng J Med 1991;324:1253-8.

World Health Organization. International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM)). Geneva, 1975.

Wright K. The economics of informal care of the elderly. Centre for Health Economics Discussion Paper 23. University of York. York, 1987.

WRR (Scientific Council for Government Affairs). Do we make it work? Research of the relation between active and inactive people (in Dutch). 's-Gravenhage, 1977.

WVC (National Department of Health and Welfare). Financial Overview of Health Care 1990 (in Dutch). The Hague, 1989.

Yeargin-Allsopp M, Murphy CC, Oakley GP, Siker RK. A multiple- source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities Study. Pediatrics 1992;89:624-630.

Summary

The increasing conflict between limited health care budgets and the development of medical technology has forced policy makers to reconsider the resource allocation in health care. Economic evaluations can support the decision making process, by providing information on the costs and health effects of health care programs.

Health care can be analyzed on 'the detailed level': an individual program or, alternatively on 'the aggregate level': total health care, a specific sector of health care or a disease category. Either level of analysis has advantages and limitations. The detailed analysis is appropriate, when the policy decision is to choose among specific programs for a specific patient group. However, focusing exclusively on cost-effectiveness of individual programs bears the risk of overlooking interactions with other health care programs. In addition, it is not feasible to evaluate each possible health care program.

More aggregate economic evaluations are indispensable in providing a consistent and complete evaluation of health care, as a starting point for more detailed analysis in specific parts of health care that need closer investigation of costs and health effects.

The cost-effectiveness analysis of cervical cancer screening illustrates the merits of a detailed economic evaluation of an individual health care program. The costs of screening itself are dominant, the impact of screening on the costs of diagnosis and treatment is limited. The analysis permitted the selection of technically efficient screening policies (generating maximal health effects for a given level of costs or minimal costs for a given level of health effects). These efficient policies can be characterized by a broad age range for women to be invited: age 30-70 and at least a 5-year interval between successive invitations. Given technical efficiency, the cost per life-year gained are reasonable for policies with 10 or less invitations in a lifetime. The marginal cost-effectiveness deteriorates sharply if women are invited more than 10 times.

The attendance rate is crucial for the cost-effectiveness of cervical cancer screening, due to economies of scale related to the costs of screening

and the relation between risk and attendance. Additional literature research indicated that the way the screening program is organized may have a considerable impact on attendance. Hence, investments in a sound organisation of screening appear to be worthwhile.

The aggregate cost of illness study described shows that for the Netherlands and Sweden the health care costs of chronic non-fatal disease predominate, which is less prominent for the United States. The available data suggest that the differences in medical practice and health care system may explain a substantial part of the divergent results; demographic or epidemiologic aspects seem less important. Assuming constant cost per caput (age- and sex-specific), ageing induces a modest 0.7%point annual cost increase in the Netherlands.

However, health care costs per caput for elderly may increase faster than for younger people, due to compression of disease and a tendency towards more extensive interventions in elderly. Further analysis of the trends in health care costs for elderly may refine the cost estimates related to ageing.

The contribution of other forces in raising costs, such as the limited role of labour productivity growth in health care and the rapid development of medical technology is probably more important than ageing. A structural upward pressure on costs also prevails in the Netherlands and Sweden, but it is more prominent in the U.S.A., probably due to a large number of medical procedures and high administration costs.

With respect to cancer, the estimated annual cost increase due to ageing is 1.5%point. The future costs of cancer also depend on the expected epidemiological development. According to the cancer disease model, the costs of colorectal cancer will increase sharply if incidence rises, whereas improved disease free survival only has a moderate impact on costs.

Cost estimates incorporated in disease models as described, supplemented with data on effectiveness of the relevant health care programs, may give an indication of the cost-effectiveness of the total set of health care programs directed to a particular disease or specific subsectors of health care.

At the moment, the relevance of the indirect non-medical costs in economic evaluation (further referred to as indirect costs) is subject to considerable debate. In this thesis, indirect costs of disease are defined as the consequences of absence from work, disability or death for society's production and costs. This definition explicitly excludes the valuation of life as such, as applied in cost-benefit analysis.

Because indirect costs of disease contribute to the scarcity of resources and decrease society's wealth, they should be considered in economic evaluations. It seems preferable to insist that the estimated indirect costs represent the real economic consequences of disease, as far as possible. The friction cost method seeks to estimate the indirect costs realistically. Its basic idea is that the amount of production lost and/or extra costs incurred due to disease depends on the time-span organisations need to restore the initial level of production and costs. It is assumed that if unemployment is beyond the level of frictional unemployment, sick employees can be replaced, after a period necessary for adaptation. Indirect costs are assumed to be confined to the period needed to replace a sick worker: the friction period.

For the Netherlands the short term friction costs in 1990 amount to 1.5-2.5% of net national income, depending on the extent to which short term absence from work induces production loss and costs. The medium term macro-economic consequences of absence from work and incidence of disability are estimated to reduce net national income by an additional 0.8%. The total economic benefits of attainable reductions in absence from work (-2 percent points) and incidence of disability (-20%) are estimated to be about 0.5% of net national income.

The exact consequences of short term absence from work for production and costs need to be analyzed in more detail on the level of the firm. Further study should deal with the possibilities of postponing work and making up for it later, the nature and extent of internal labour reserves versus flexible labour arrangements and the diminishing returns to labour for the individual worker.

The impact of including indirect costs on results of particular health care programs depends predominantly on the time lag for health effects, the relative importance of (short term) absence and the labour participation of the population to benefit. For cervical cancer screening the impact is rather small. For therapeutic interventions reducing absence from work, like cimetidine for peptic ulcer and thrombolytic treatment for myocardial infarction, the estimated savings in indirect costs appear to be substantial.

Samenvatting

De toenemende spanning tussen de beschikbare middelen voor de gezondheidszorg en de voortgaande ontwikkeling van medische technologie noopt beleidmakers steeds vaker tot heroverweging van de allocatie van middelen in de gezondheidszorg. Economische evaluatiestudies kunnen dit proces ondersteunen, door informatie te verstrekken over de kosten en gezondheidseffecten van gezondheidszorgvoorzieningen.

De gezondheidszorg kan op verschillende niveaus worden geanalyseerd: op detailniveau: de individuele voorziening, of op meer geaggregeerd niveau: de totale zorgsector, deelsectoren of een ziektecategorie. Beide analyseniveaus kennen voordelen en beperkingen. De detailanalyse is de aangewezen methode indien de besluitvorming gericht is op de keus tussen specifieke programma's voor een bepaalde patiëntengroep. Wanneer men alleen individuele voorzieningen evalueert bestaat het gevaar dat belangrijke interacties tussen voorzieningen worden gemist. Daarnaast is het praktisch ondoenlijk om alle gezondheidszorgvoorzieningen apart te evalueren.

Geaggregeerde analyses zijn noodzakelijk om een consistent en compleet beeld te verkrijgen van de kosten en effecten van de gezondheidszorg. Dit kan dienen als startpunt voor de selectie van delen van de gezondheidszorg waar de kosten en effecten in meer detail kunnen worden geanalyseerd.

De kosten-effectiviteitsanalyse van baarmoederhalskankerscreening toont de kracht van een gedetailleerde analyse van een individuele voorziening. De kosten van de screening zelf zijn dominant en de invloed van screening op de kosten van diagnostiek en behandeling zijn gering. Het bleek mogelijk technisch efficiënte screeningpolicies te selecteren (deze leveren maximale gezondheidseffecten voor een gegeven kostenniveau, of een gegeven niveau van effecten tegen minimale kosten). Deze efficiënte policies hebben twee kenmerken gemeen: een brede leeftijdsgroep van vrouwen wordt uitgenodigd voor de screening (30-70 jaar) met een interval van tenminste 5 jaar tussen de uitnodigingen. Voor de efficiënte policies geldt dat de kosten per gewonnen levensjaar redelijk gunstig zijn bij 10 of minder uitnodigingen per vrouw. Bij meer dan 10 uitnodigingen stijgen de additionele kosten per extra gewonnen levensjaar snel.

Het percentage van de uitgenodigde vrouwen dat daadwerkelijk gescreend wordt is van groot belang voor de kosteneffectiviteit, vanwege de schaalvoordelen in de kosten en de negatieve relatie tussen risico en opkomst. Aanvullende literatuurstudie wees uit dat de organisatie van het screeningsprogramma van groot belang kan zijn voor het opkomstpercentage, zodat een goede screeningsorganisatie een waardevolle investering lijkt.

De geaggregeerde kosten van ziektenstudie laat zien dat in Nederland en Zweden de kosten van niet fatale, chronische ziekten zeer belangrijk zijn, terwijl dit patroon in de Verenigde Staten veel minder duidelijk zichtbaar is. De beschikbare data suggereren dat het uiteenlopende kostenpatroon eerder lijkt voort te komen uit verschillen in medisch handelen en het zorgsysteem dan het gevolg is van verschillen in demografie en epidemiologie. Bij gelijkblijvende kosten per hoofd (naar leeftijd en geslacht), zorgt vergrijzing voor een bescheiden kostenstijging van 0.7 %punt per jaar.

Het is echter goed mogelijk dat de kosten per hoofd voor ouderen sneller zullen stijgen dan voor jongeren, als gevolg van compressie van ziekten en een tendens tot intensivering van diagnostische en therapeutische zorg voor ouderen. Verdere analyse van de trend in de zorgkosten voor ouderen is noodzakelijk om de kostenramingen van vergrijzing te verbeteren.

Andere factoren, zoals de beperkte groei van de arbeidsproduktiviteit in de zorgsector en de ontwikkeling van medische technologie zijn wellicht belangrijker oorzaken van kostenstijging. Nederland en Zweden kennen ook een permanente stijging van de zorgkosten, maar de toename is veel pregnanter in de Verenigde Staten, onder meer als gevolg van de grote omvang van dure chirurgie en de hoge administratieve kosten.

Voor kanker bedraagt de geschatte kostenstijging als gevolg van vergrijzing 1.5 %punt per jaar, bij gelijkblijvende kosten per hoofd (naar leeftijd en geslacht). De toekomstige kosten van kanker hangen ook af van epidemiologische ontwikkelingen. Het beschreven ziektemodel voor kanker wijst uit dat de kosten van colorectumkanker sterk zullen stijgen bij toename van de incidentie, terwijl verbetering van de ziektevrije overleving een veel geringere invloed heeft op de kosten.

Het koppelen van kostenramingen aan ziektemodellen, gecombineerd met gegevens over effectiviteit van de bijbehorende gezondheidszorg kan een indicatie geven van de totale kosten-effectiviteit van de voorzieningen voor een bepaalde ziektecategorie of specifieke deelsectoren van de zorg.

Het belang van indirecte niet medische kosten in economische evaluaties is op dit moment niet onomstreden. In dit proefschrift worden indirecte kosten gedefinieerd als de gevolgen van ziekteverzuim, arbeidsongeschiktheid en sterfte voor de produktie en de bijbehorende kosten op maatschappelijk niveau. Deze definitie bevat geen waardering van het menselijk leven op zich, zoals toegepast in kosten-baten analyse.

Omdat indirecte kosten de schaarste van goederen en diensten beïnvloeden, dienen ze te worden geanalyseerd in economische evaluaties, maar de ramingen van indirecte kosten moeten zo veel mogelijk de werkelijke economische gevolgen weerspiegelen. De frictiekostenmethode vormt hiertoe een aanzet. Het basisidee van deze methode is dat de indirecte kosten bij ziekte afhangen van de periode die nodig is voor organisaties om het niveau van produktie en bijbehorende kosten te herstellen. Aangenomen wordt dat wanneer de werkloosheid hoger is dan het niveau van de frictiewerkloosheid, na een zekere periode, zieke werknemers in het algemeen kunnen worden vervangen. De indirecte kosten blijven beperkt tot deze periode van aanpassing: de frictieperiode.

Voor Nederland zijn de korte termijn frictiekosten voor 1990 geschat op 1.5 -2.5% van het netto nationaal inkomen, afhankelijk van de mate waarin kortdurend verzuim produktieverlies en kosten veroorzaakt. De macro-economische gevolgen op middellange termijn van verzuim en incidentie van arbeidsongeschiktheid worden geschat op 0.8 % van het netto nationaal inkomen. De totale economische voordelen van een haalbare reductie van ziekteverzuim (-2 procentpunten) en incidentie van arbeidsongeschiktheid (-20%) komen naar schatting uit op ongeveer 0.5% van het netto nationaal inkomen.

De precieze gevolgen van korte termijn verzuim voor produktie en kosten dienen in meer detail bestudeerd te worden op bedrijfs(tak)niveau. Hierbij zou speciaal gelet moeten worden op de mogelijkheid om werk uit te stellen en later in te halen, de omvang en aard van interne en externe arbeidsreserves en het effect van de afnemende meeropbrengsten van de individuele werknemer.

De invloed van indirecte kosten op de resultaten van zorgvoorzieningen hangt vooral af van de mogelijke vertraging in het optreden van de gezondheidseffecten, het belang van (kortdurend) verzuim en de arbeidsparticipatie van de doelgroep. Voor baarmoederhalskankerscreening is de invloed beperkt. Bij therapeutische voorzieningen die leiden tot afname van verzuim, zoals cimetidine voor maagzweren en thrombolitica voor myocard infarct, zijn de geraamde besparingen op de indirecte kosten aanzienlijk.

Appendix 4

Definition and costs of 48 disease categories, in mln of Dutch guilders, 1988.

Disease category	ICD-code	ICD-Chapter	Costs men	Costs women	Total costs
Infective dis exc AIDS	001-139 exc 042	1	175	195	370
AIDS	042	1	23	17	41
Stomach cancer	151	2	46	31	77
Colorectal cancer	153-154	2	91	113	203
Lung cancer	162	2	174	35	210
Breast cancer	174	2	1	162	162
Prostate cancer	185	2	77	0	77
Other malignant dis	rest 140-208	2	357	453	809
Benign cancers	210-239 + 173	2	74	199	274
Diabetes	250	3	153	289	442
Other endocrin dis	240-279 exc 250	3	60	165	225
Blood diseases	280-289	4	53	85	138
Dementia	290	5	297	1058	1356
Other psychiatric dis	291-316	5	1748	2059	3807
Mental retardation	317-319	5	1562	1178	2739
Eye diseases	360-379	6	172	242	415
Ear diseases	380-389	6	162	153	314
Other neurological dis	320-359	6	327	473	800
Hypertension	401-405	7	52	91	143
Ischemis heart dis	410-414	7	538	278	815
Stroke	430-438	7	443	706	1150
Congestive heart failure	428-429	7	179	250	429
Other cardiovascular dis	rest 390-459	7	477	444	921
Acute lung diseases	460-466, 480-487	8	210	195	404
Chronic lung dis	490-496	8	270	181	451
Other lung dis	rest 460-519	8	212	165	377
Dental diseases	520-529	9	783	942	1724
Duodenumulcer	531-534	9	48	46	94
Appendicitis	540-543	9	59	67	126
Hernias	550-553	9	160	53	213
Biliary and liver dis	570-576	9	104	195	299
Other dis of the gut	555-569	9	169	243	412
Other gastro-intest dis	rest 520-579	9	85	84	169
Nefritis/nefrosis	580-589	10	49	55	104
Other renal/urinary dis	590-599	10	164	223	387
Dis male genital organs	600-608	10	174	0	174
Dis female genital organs	610-627, 629	10	0	375	375
Fertility problems	628	10	3	18	21
Pregnancy and delivery	630-676	11	80	1310	1389
Dermatologic dis	680-709	12	148	223	371
Locomotor dis exc RA	710-739 ex 714	13	1020	1587	2607
Reumatoide artritis	714	13	37	122	159
Congenital anomalies	740-759	14	97	76	173
Perinatal diseases	760-779	15	191	163	354
Symptoms	780-799	16	710	1016	1726
Traffic accidents	E800-E848	17	204	138	343
Other accidents	E850-E869, E880-E929	17	376	659	1035
Other external causes of injury and death	E870-E879, E930-E999	17	141	189	329
Not assigned to specific diseases			4389	5629	10017

Appendix 5.1

Estimation of cancer survival

1. The survival was modelled by using the lognormal distribution:

$$f(t) = \frac{1}{t\sqrt{2\pi}\sigma} \exp\left[\frac{-(\ln t - \mu)^2}{2\sigma^2}\right]$$

Where t is the time after diagnosis, μ the median survival and σ^2 the variance.
We used a discrete approximation with a time step of one month. The proportion of survivors at any time t_i after diagnosis t_0 is then given by:

$$P(S)_{t_i} = (1-M_{op})\left(c + (1-c)\sum_{t_0}^{t_i}\left[\frac{1}{t_i\sqrt{2\pi}\sigma}\exp\left(\frac{-(\ln t_i - \mu)^2}{2\sigma^2}\right)\right]\right)$$

where M_{op} is the proportion dying in hospital after first admission and c is the fraction cured.

Appendix 5.2

Average costs per patient by cancer type, phase, age and sex, x 1,000 Dutch guilders, 1988.

Stomach cancer

	Men		Women	
Age	Incidence phase	Last year of life	Incidence phase	Last year of life
25-44	40	37	37	32
45-49	33	30	33	27
50-54	31	27	31	25
55-59	32	25	34	24
60-64	30	25	34	26
65-69	30	24	32	23
70-74	26	23	30	25
75-79	23	21	32	29
80-84	20	19	25	24
85+	15	15	14	15

The costs during the intermediate phase are dfl 514 per patient per year.

Colorectal cancer

	Men		Women	
Age	Incidence phase	Last year of life	Incidence phase	Last year of life
25-44	27	23	33	21
45-49	26	23	29	23
50-54	26	20	29	22
55-59	27	20	28	22
60-64	26	23	30	26
65-69	29	22	34	26
70-74	28	21	34	30
75-79	25	23	34	29
80-84	23	19	30	25
85+	14	12	21	18

The costs during the intermediate phase are dfl 443 per patient per year.

Lung cancer

	Men		Women	
Age	Incidence phase	Last year of life	Incidence phase	Last year of life
25-44	27	27	38	32
45-49	25	27	34	31
50-54	25	27	34	33
55-59	27	27	39	32
60-64	25	28	37	39
65-69	26	28	35	37
70-74	23	24	34	36
75-79	21	23	31	28
80-84	15	17	23	21
85+	10	11	10	10

The costs during the intermediate phase are dfl 536 per patient per year.

Breast cancer			Prostate cancer		
	Women			Men	
Age	Incidence phase	Last year of life	Age	Incidence phase	Last year of life
25-44	19	22	40-44	28	25
45-49	22	24	45-49	23	25
50-54	23	26	50-54	23	25
55-59	23	25	55-59	22	24
60-64	22	27	60-64	22	29
65-69	22	25	65-69	23	28
70-74	21	26	70-74	21	27
75-79	21	27	75-79	20	28
80-84	18	23	80-84	19	24
85+	11	15	85+	11	13

The costs during the intermediate phase for breast cancer (prostate cancer) are dfl 375 (435) per patient per year.

Other cancers

	Men		Women	
Age	Incidence phase	Last year of life	Incidence phase	Last year of life
0-4	55	51	54	47
5-9	46	51	49	47
10-14	38	45	35	35
15-19	35	41	32	35
20-24	33	41	33	35
25-29	31	38	33	35
30-34	31	38	34	35
35-39	27	28	33	35
40-44	27	28	32	35
45-49	25	26	34	35
50-54	23	24	32	36
55-59	24	24	33	36
60-64	22	25	34	38
65-69	21	25	35	37
70-74	19	23	32	39
75-79	18	24	33	39
80-84	15	20	31	34
85+	14	18	25	27

The costs during the intermediate phase are dfl 488 per patient per year.

Appendix 6

References of empirical studies concerning indirect costs

Birenbaum LK, Clarke-Steffen L. Terminal care costs in childhood cancer. Pediatric Nursing 1992;18:285-288.

Böstman O, Hirvensalo E, Partio E, Törmälä P, Rokkanen P. Impact of the use of absorbable fracture fixation implants on consumption of hospital resources and economic costs. The Journal of Trauma 1991;31:1400-1403.

Browne GB, Arpin K, Corey P, Fitch M, Gafni A. Individual correlates of health service utilization and the cost of poor adjustment to chronic illness. Medical Care 1990;28:43-58.

Colditz GA. Economic costs of obesity. Am J Clin Nutr 1992;55:503S-507S.

Cushner F, Friedman RJ. Economic impact of total hip arthroplasty. Southern Medical Journal 1988;81:1379-1381.

Daly E, Roche M, Barlow D, Gray A, McPherson K, Vessey M. HRT: An analysis of benefits, risks and costs. British Medical Bulletin 1992;48:368-400.

Dodson WC, Haney AF. Controlled ovarian hyperstimulation and intrauterine insemination for treatment of infertility. Fertility and Sterility 1991;55:457-467.

Fahs MC, Markowitz SB, Fischer E, Shapiro J, Landrigan PJ. Health costs of occupational disease in New York State. American Journal of Industrial Medicine 1989;16:437-449.

Fitzgerald JM, Gafni A. A cost-effectiveness analysis of the routine use of isoniazid prophylaxis in patients with a positive mantoux skin test. Am Rev Respir 1990;142:848-853.

Frymoyer JW, Cats-Baril WL. An overview of the incidences and costs of low back pain. Othopedic Clinics of North America 1991;22:263-271.

Gerard K, Donaldson C, Maynard AK. The cost of diabetes. Diabetic Med 1989;6:164-170.

Girolamo de G. Epidemiology and social costs of low back pain and fibromyalgia. The Clinical Journal of Pain 1991;7(Suppl.1):S1-S7.

Gorsky RD, Schwartz E, Dennis D. The mortality, morbidity and economic costs of alcohol abuse in New Hampshire. Preventive Medicine 1988;17:736-745.

Gorsky RD, Schwartz E, Dennis D. The morbidity, mortality, and economic costs of cigarette smoking in New Hampshire. J Comm Health 1990;15:175-183.

Goulston KJ, Dent OF, Mant A, Logan J, Ngu M. Use of H_2-receptor antagonists in patients with dyspepsia and heartburn: a cost comparison. Medical Journal Australia 1991;155:20-26.

Grimm RH. Epidemiology and cost implications of antihypertensive treatment for the prevention of cardiovascular disease. Journal of Human Hypertension 1989;3:55-61.

Guyer B, Ellers B. Childhood injuries in the United States. Mortality, morbidity, and cost. AJDC 1990;144:649-652.

Harlan LC, Harlan WR, Parsons PE. The economic impact of injuries: a major source of medical costs. AJPH 1990;80:453-459.

Hatziandreu EI, Koplan JP, Weinstein MC, Caspersen CJ, Warner KE. A cost-effectiveness analysis of exercise as a health promotion activity. AJPH 1988;78:1417-1421.

Havas S. Heart disease, cancer and stroke in Maryland. Southern Medical Journal 1992;85:599-607.

Hay JW, Ernst RL. The economic costs of Alzheimer's disease. AJPH 1987;77:1169-1175.

Huang LF, Cartwright WS, Hu TW. The economic cost of senile dementia in the United States, 1985. Public Health Reports 1988;103:3-7.

Jensen DM. Economic assessment of peptic ulcer disease treatments. Scand J Gastroenterol 1988;23(suppl 146):214-224.

Jonsson B, Rehnberg C, Borgquist L, Larsson SE. Locomotion status and costs in destructive rheumatoid arthritis. A comprehensive study of 82 patients from a population of 13.000. Acta Orthop Scand 1992;63:207-212.

Jönsson B, Horisberger B, Bruguera M, Matter L. Cost-benefit analysis of hepatitis-B vaccination. A computerized decision model for Spain. International Journal of Technology Assessment in Health Care 1991;7:379-402.

Jönsson B, Carlsson P. The effects of cimetidine on the cost of ulcer disease in Sweden. Soc Sci Med 1991;33:275-282.

Jönsson B, Haglund U. Cost-effectiveness of misoprostol in Sweden. International Journal of Technology Assessment in Health Care 1992;8:234-244.

Krosnick A. Economic impact of type II Diabetes Mellitus. Primary Care 1988;15:423-432.

Laupacis A, Labelle R, Goeree R, Cairns J. The cost-effectiveness of routine post myocardial infarction exercise stress testing. Can J Cardiol 1990;6:157-163.

Leigh JP, Richardson N, Beck R, Kerr C, Harrington H, Parcell CL, Fries JF. Randomized controlled study of a retiree health promotion program. Arch Intern Medicine 1992;152:1201-1206.

Lenfant C, LaRosa JH, Horan MJ, Passamani ER. Considerations for a national heart attack alert program. Clin Cardiol 1990;13:VIII9-11.

Levin LA, Jönsson B. Cost-effectiveness of thrombolysis - a randomized study of intravenous rt-PA in suspected myocardial infarction. Eur Heart J 1992;13:2-8.

Lubeck DP, Yelin EH. A question of value: measuring the impact of chronic disease. The Milbank Quarterly 1988;66: 444-464.

MacKenzie EJ, Morris JA, Smith GS, Fahey M. Acute hospital costs of trauma in the United States: implieations for regionalized systems of care. The Journal of Trauma 1990;30:1096-1103.

Mellis MG, Peat JK, Bauman AE, Woolcock AJ. The cost of asthma in New South Wales. Medical Journal Australia 1991;155:522-528.

Moore DA, Hopkins RS. Assessment of a school exclusion policy during a chickenpox outbreak. American Journal of Epidemiology 1991;133:1161-1167.

Olivera EM, Pérez Duhalde E, Gagliardino JJ. Costs of temporary and permanent disability induced by diabetes. Diabetes care 1991;14:593-596.

Olsson G, Levin LA, Rehnqvist N. Economic consequences of postinfarction prophylaxis with β blockers: cost effectiveness of metoprolol. British Medical Journal 1987;294:339-342.

Ozer H, Golomb HM, Zimmerman H, Spiegel RJ. Cost-benefit analysis of interferon Alfa-2b in treatment of hairy cell leukemia. J Nat Cancer Inst 1989;81:594-602.

Rutz W, Carlsson P, Knorring von L, Wälinder J. Cost-benefit analysis of an educational program for general practitioners by the Swedish Committee for the prevention and treatment of depression. Acta Psychiatr Scand 1992;85:457-464.

Ryden LA, Molgaard CA, Bobbitt SL. Benefits of a back care and light duty health promotion program in a hospital setting. J Comm Health 1988;13:222-230.

Saywell RM, Woods JR, Holmes GL, Sechrist ME, Nyhuis AW. Reducing the costs of patient transfers. JONA 1987;17:11-19.

Scitovsky AA, Rice DP. Estimates of the direct and indirect costs of acquired Immunodeficiency Syndrome in the United States, 1985, 1986, and 1991. Public Health Reports 1987;102:5-17.

Stool SE, Field MJ. The impact of otitis media. Pediatr Infect Dis J 1989;8:11-S14.

Washington AE, Johnson RE, Sanders LL. Chlamydia trachomatis infections in the United States. What are they costing us? JAMA 1987;257:2070-2072.

Washington AE, Katz P. Cost of and payment source for pelvic inflammatory disease. JAMA 1991;266:2565-2569.

Weiss KB, Gergen PJ, Hodgson TA. An economic evaluation of asthma in the United States. New Eng J Med 1992;326:862-866.

Williams AF, Begley CE. The cost of cancer in Texas. The Journal of Texas Medicine 1992;88:62-67.

Appendix 8

Average production value per employee per friction period, by age, sex and education level for the Netherlands 1990, in thousands of Dutch guilders.

Men

Education-> Age	Basic	Extended basic	Intermediary	Higher vocatio-nal	University
15-19	4.6	5.1	7.1	Not applicable	Not applicable
20-24	10.8	11.9	15.0	18.6	20.6
25-29	13.3	15.9	20.3	23.5	26.3
30-34	14.8	18.0	24.3	31.7	35.6
35-39	14.4	17.9	27.9	34.7	56.9
40-44	15.2	18.7	29.3	41.6	65.9
45-49	15.9	19.3	31.5	42.1	63.2
50-54	15.5	18.6	32.2	44.1	58.8
55-59	14.6	18.4	28.9	40.8	59.3
60-64	13.0	15.2	26.8	36.6	53.3

Women

Education-> Age	Basic	Extended basic	Intermediary	Higher vocatio-nal	University
15-19	3.5	3.7	6.3	Not applicable	Not applicable
20-24	7.6	9.4	12.6	17.1	19.7
25-29	9.1	11.2	15.6	20.1	23.6
30-34	7.8	10.1	15.4	20.4	24.0
35-39	8.3	10.5	17.1	22.1	32.8
40-44	7.7	9.6	16.4	24.3	33.6
45-49	7.8	10.2	17.8	26.5	38.0
50-54	7.4	10.1	18.6	27.1	37.9
55-59	6.7	9.0	16.8	24.6	33.5
60-64	6.2	8.2	15.8	23.1	31.4

Appendix 9

Average length of hospital stay and absence from work by type of treatment for cervical cancer; frequency of treatments by disease stage, the Netherlands.

Type of treatment	Hospital stay in days	Absence in days	Frequency stage CIN 1-2*	Frequency stage CIN 3*	Frequency invasive cancer	Frequency advanced cancer
Conserving treatments	0 (outpatients)	14	0.94	0.25	0	0
Exconisation (clinical)	5	21	0.08	0.68	0	0
Hysterectomy	13	50	0	0.11	0.2	0
Radical hysterectomy	23	65	0	0	0.6	0
Radiotherapy, palliation etc.	varying	> 85	0	0	0.2	1

* Sum of frequencies is greater than 1, due to exconisation as retreatment-procedure in case of conserving treatments as primary therapy.
Sources: SIG (hospital length of stay) and expert opinion (absence duration).

Curriculum Vitae

Marc Koopmanschap was born in 1961 in Amsterdam. After following secondary school at the Keizer Karel College in Amstelveen, he studied economics at the University of Amsterdam (1979-1986). At the Society for Economic Research (S.E.O.) of this university he assisted in labour market analysis (1985-1987). From 1987 onwards until June 1993 he was researcher at the institute of Public Health of the Erasmus University Rotterdam. Since July 1993 he combines research and teaching at the Institute of Medical Technology Assessment, part of the department of Health Policy and Management at the Erasmus University Rotterdam.

Acknowledgements

The analysis of cervical cancer screening (chapter 2 and 3) was financed by the Prevention Fund. The Dutch Ministry of Welfare, Health and Culture (WVC) financed the cost of illness studies in chapter 4 and 5. Dorothy Rice and Björn Lindgren supplied additional information on their cost of illness studies for the U.S.A. and Sweden respectively (chapter 5). The analysis of indirect costs was partially sponsored by Merck & Co Inc., and the Netherlands Institute for Health Sciences (NIHES). I am grateful to Marjolein van Ballegooijen, Ben van Hout, Martin van Ineveld, Lars Ake Levin and Maureen Rutten-van Mölken for providing information on the economic evaluations of some of the health care programs analyzed in chapter 9.

Thanks

Many people have contributed to this thesis, but some of them should be mentioned explicitly.

The members of the cervical cancer screening project supported me in analyzing screening in a systematic, model-based way. I would like to thank the researchers of the TAM-project, especially Luc Bonneux and Jan Barendregt, for their continuous stream of ideas and for their enthusiasm in suggesting new hot research topics.

Two persons should be mentioned separately. Special thanks are for Martin van Ineveld for on the job training in economic evaluation and for sharing my suspicion for neoclassical ideas. Leona van Roijen joined me in the extensive job of collecting data on disease costs and in exploring new issues in economic evaluation of health care.

I wish to acknowledge my supervisors for their support; Paul van der Maas for pushing the generalistic point of view in public health affairs and Frans Rutten for keeping me on a passable track and for correcting my english.

I am very grateful to Angélique Visser for her patience in performing the donkey work of this thesis: the layout.

Finally, Olga, Isabel and athletics frequently reminded me of the importance of other activities, which can't be overestimated.