Publication of this thesis was financially supported by the Department of Public Health, Erasmus MC, Rotterdam.

Cover design: Anne Henne Boone
Layout: Jaap van den Akker
Printed by: Print Partners Ipskamp, Enschede

ISBN 90-77283-03-x

# Mathematical Modelling in Evaluation of Screening for Breast and Cervical Cancer

## Het gebruik van wiskundige modellen bij de evaluatie van bevolkingsonderzoek naar borst- en baarmoederhalskanker

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de Rector Magnificus
Prof.dr.ir. J.H. van Bemmel
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 25 juni om 11.45 uur

door

Margaretha Elisabeth van den Akker-van Marle
geboren te Alphen aan den Rijn

**Promotiecommissie**

Promotor : Prof.dr.ir. J.D.F. Habbema

Overige leden: Prof.dr. Th. J. M. Helmerhorst
Prof.dr. F.F.H. Rutten
Prof.dr. A.L.M. Verbeek

Copromotor : Dr. M. van Ballegooijen

# CONTENTS

# 1

# INTRODUCTION

## CANCER AS A PUBLIC HEALTH PROBLEM

Cancer is an important public health problem. In the Netherlands, the cancer mortality rate was 271 per 100,000 men and 218 per 100,000 women in the year 1998 (1), or stated otherwise, 31% of all deaths in men and 25% of all deaths in women were due to cancer in that year. These numbers are slightly lower than the death rates for diseases of the circulatory system. In 1998, the mortality from these diseases was 313 deaths per 100,000 men and 321 per 100,000 women. Diseases of the respiratory system (influenza, pneumonia, chronic lower respiratory diseases and asthma) rank third, with considerably lower death rates of 97 per 100,000 men and 82 deaths per 100,000 women.
In men lung, prostate and colorectal cancer were shown to be the three principal cancers, accounting respectively for 86, 31 and 21 deaths per 100,000 men respectively. A similar pattern was reflected in the incidence rate, which showed that the incidence of lung cancer was 87, that of prostate cancer 83, and that of colorectal cancer 35 per 100,000 men.
In women, the three cancers with the highest mortality rates are breast, lung and colorectal cancer, which respectively accounted for 45, 25 and 22 deaths per 100,000 women in 1998. The incidence of breast cancer was 127 per 100,000 women in 1997, while the figures for lung cancer were 27 per 100,000 women and for colon cancer, 37 per 100,000 women (8). Breast cancer incidence and mortality rates have been affected by the national screening programme for breast cancer, which was launched around 1989.

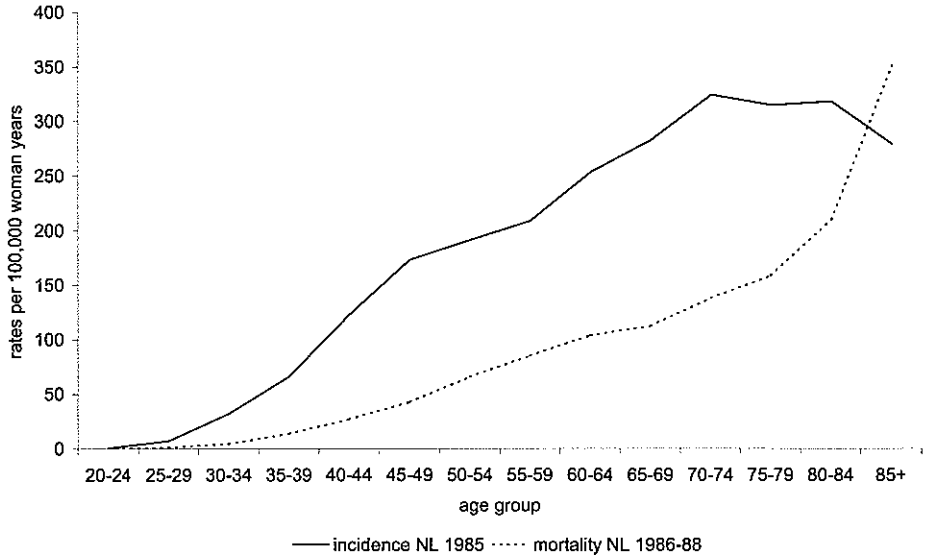
In 1985, prior to the implementation of the nationwide screening programme, the incidence of breast cancer in the Netherlands was assessed at 102 per 100,000 women on the basis of hospital admission data. Figure 1.1.a. shows the incidence and the mortality in this period by age group (4). The lifetime risk of developing breast cancer was at that time 10% and the burden of disease, as expressed by the number of life years lost due to mortality from breast cancer, was 17,500 per 100,000 women.
Incidence rates for breast cancer differ worldwide (see Figure 1.2.a) (9). In general, breast cancer incidence is highest in high-income countries. In countries with screening programmes or recommendations, screening will influence the incidence rate.

The incidence of cervical cancer in the Netherlands in 1965-1969, the period before the Pap smear was introduced in the Netherlands, was 23 per 100,000 women. The incidence data were obtained from three local registries in the (rural) province of Friesland and in the cities of Rotterdam and The Hague. Together these registries covered 8% of the Dutch population (3). In Figure 1.1.b the incidence and the mortality (national data) in this period is shown by age group. The lifetime risk of developing cervical cancer was 2% in that period, and the life years lost due to mortality from cervical cancer amounted to 3,500 per 100,000 women.

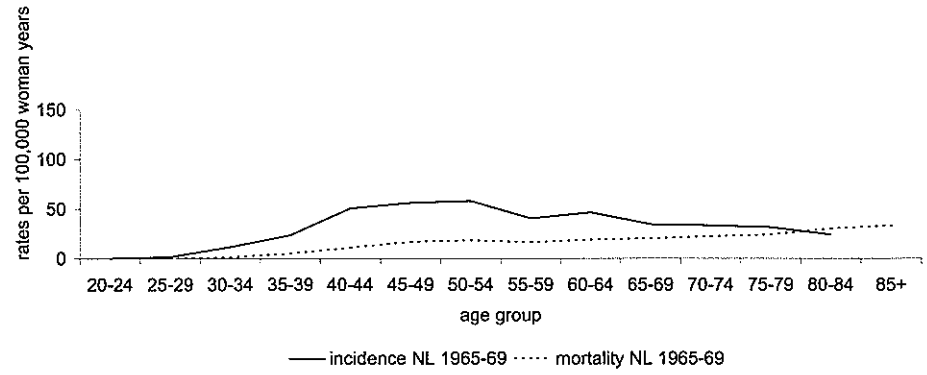
***Figure 1.1***

*The incidence and mortality per 100,000 woman years for breast (a) and cervical (b) cancer from the period before screening for the respective cancers became widespread(3-5).*

a



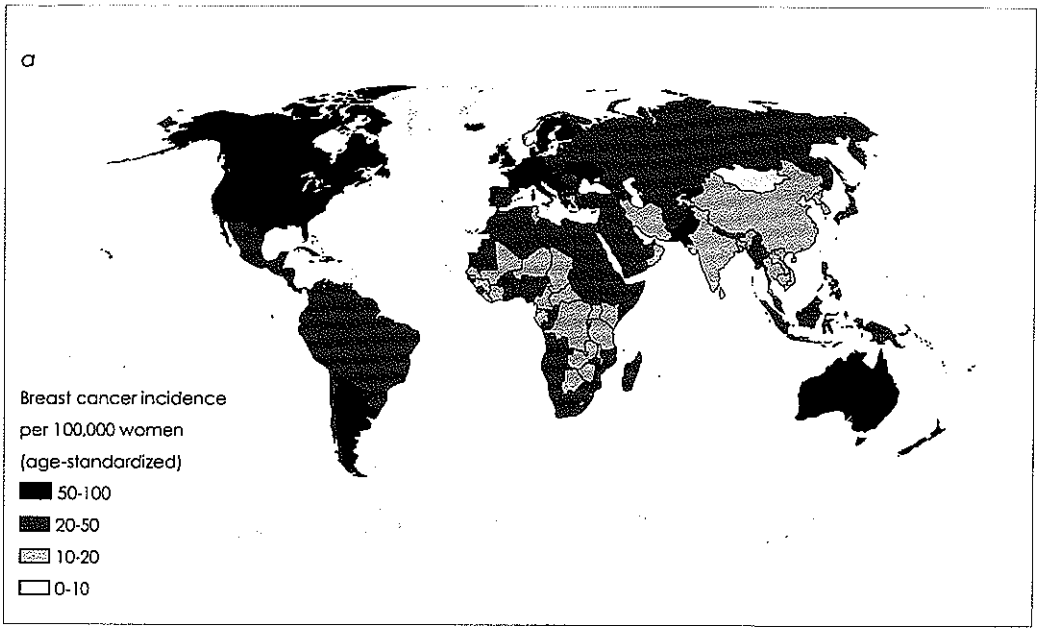

b




Otherwise than with breast cancer, the highest incidence rates for cervical cancer occur in low-income countries, as shown in Figure 1.2.b (9). In countries where screening is performed, whether on an opportunistic or organised basis, the incidence may be lower.

***Figure 1.2***

*Overview of incidence of breast (a) and cervical (b) cancer in the world. Estimates for the year 2000 (9).*



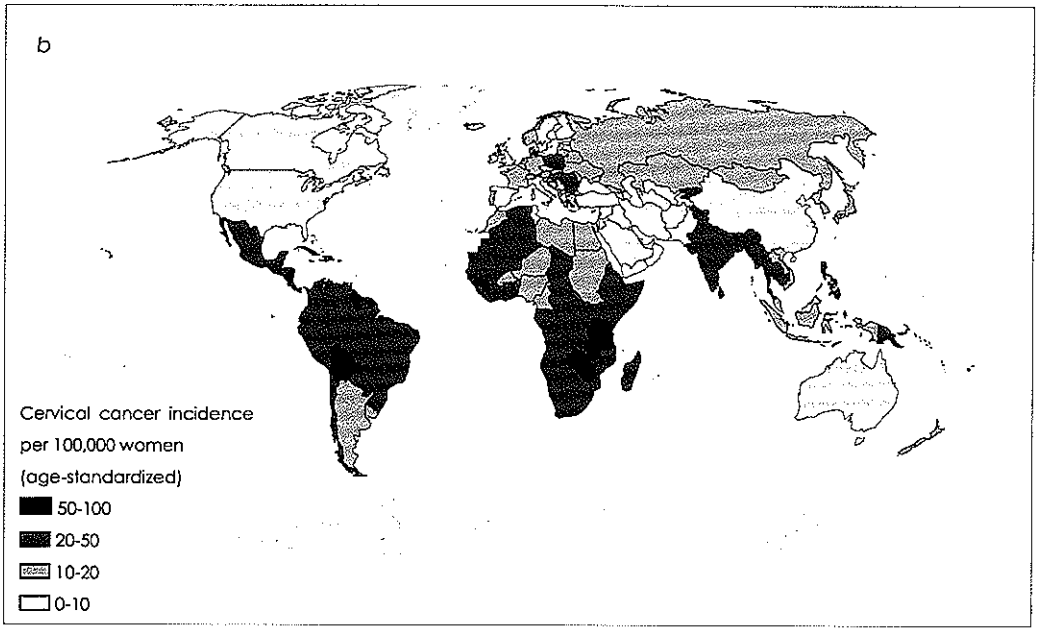
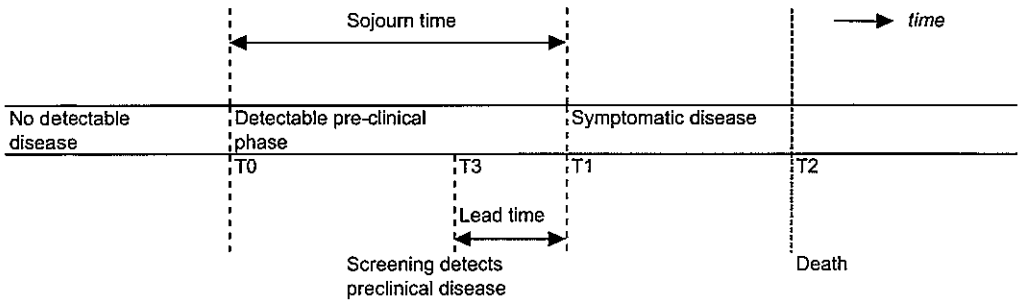



*Figure 1.3*

*The progression of chronic disease in relation to screening(2)*




# SCREENING FOR CANCER

## Screening as secondary prevention of cancer

The burden of disease may be reduced if prevention of the disease is possible. Two different types of prevention can be distinguished, namely primary and secondary prevention. Primary prevention comprises all interventions that aim at all presumably disease-free people at risk for the disease. By reducing the exposure of the population to risk factors, new cases of disease are avoided.

Secondary prevention is targeted at the early stages of the disease, in which the disease is already screen detectable, but has not yet become clinically manifest (detectable pre-clinical phase). This is represented by the period T0-T1 in Figure 1.3. If screening can postpone death from the disease (T2), early detection is effective. If the prognosis is not improved after screen-detection, the sole effect of screening is the additional time during which the patient is aware of the disease. This lead time is indicated in Figure 1.3.

The possible method(s) for prevention differ(s) by type of cancer. For primary prevention, risk factors of the disease must be known and interventions must be available to reduce exposure to the risk factors. For lung cancer, for example, smoking is an established risk factor, which means that public campaigns against smoking constitute primary prevention against lung cancer. For the possibility of secondary prevention several conditions must be fulfilled: a) the disease must have an identifiable latent phase or early symptomatic stage, b) early detection must result in a better prognosis, and c) an appropriate screening test must be available. Even if all these conditions are fulfilled, the effects and costs of the intervention must be carefully weighed against total effects and costs of health care and the cost and effects of other health interventions.

More than one method for prevention may be possible for a specific type of cancer. Next to the non smoking campaigns that serve as a primary intervention method for lung cancer, the use of spiral computed tomography for secondary prevention is currently under study *(10,11)*.

Although the foregoing classification of prevention is widespread, the distinction between the different categories of prevention is ambiguous. Detection of cervical intraepithelial neoplasia, which precedes invasive cervical cancer, can be classified as secondary prevention, while treatment of the same can be considered as primary prevention of invasive cervical cancer. The same applies to the treatment of adenomas in preventing colorectal cancer. Usually, screening is referred to as secondary prevention.

In several countries, screening programmes or national guidelines exist for breast and cervical cancer. Screening for breast and cervical cancer will be discussed below.

## Evaluation of cancer screening

Before a screening programme is introduced, its value must be demonstrated. In the Netherlands, this is regulated by the 1996 law on screening to protect people from screening examinations that have no proven benefit. To obtain a permit to run a screening programme, it must first be shown that the benefit of screening exceeds the risks.

Studies to evaluate cancer-screening programmes can be either experimental or observational. Experimental studies concern randomized controlled trials, and are the only way to obtain unbiased estimates of the effects of screening: lead-time bias, length time bias and selection bias do not influence the results, unlike observational studies, such as case-control studies (12), cohort studies and ecologic studies. Lead-time bias occurs when comparing screen-detected cases with clinically detected cases. For example, when comparing the survival after detection between both groups of cases, lead-time will cause a too favourable situation for screen-detected cases. Also, length time bias should be taken into account by comparing screen-detected and clinically detected cases. Length time bias is caused by the higher likelihood of detecting cases at screening with a longer preclinical detectable phase than fast growing cases. At screening, therefore, cases with a long sojourn time will be overrepresented. These may have a better prognosis than the more aggressive cases. Comparing the prognosis of screen-detected and clinical cases will then result in too favourable results for screening. Selection bias takes place as persons deciding to participate in screening are not representative for the general population. For example, women participating in cervical cancer screening are assumed to be at lower risk for the disease. Using the screening results as representative for the general population will lead to erroneous conclusions on the effectiveness of screening.

Experimental and observational studies are necessary to show the effectiveness of screening programs, but they have their limitations. They give an answer to the value of a specific screening policy in a particular situation. It is not feasible to do empirical studies for all alternative strategies. Nonetheless, translating these results to other screening policies or other situations is important. For example, if the effectiveness of a 2-yearly screening programme for breast cancer in women aged 50-69 for a certain country is assessed in a randomized controlled trial, a neighbouring country with, for example, a lower incidence of breast cancer, a different demographic structure and different health behaviour may be interested in whether they can expect the same effectiveness. Modelling

may then be used to combine the trial results with the data from the neighbouring country on demography, epidemiology in the situation without screening, expected screening quality, clinical practice as well as costs (*13*), in order to obtain estimates on the effectiveness for the neighbouring country.

By estimating the effectiveness of screening, which can be expressed in terms of reduction in morbidity or mortality, or increase in life-expectancy, it is possible to compare the effectiveness of different screening policies to each other, but also to compare the effectiveness of screening to other health interventions. Wright and Weinstein (*14*) compared the gain in life expectancy of different preventive interventions. The effectiveness of screening is in the middle of the range of gains in life expectancies for preventive interventions aimed at cardiovascular diseases (6-13 months) and vaccines for infectious diseases (0.01-0.26 months). However, from a public health perspective, it is also highly relevant to consider the costs in comparing the different interventions. In a cost-effectiveness analysis, a trade-off is made between cost and effects of interventions. To improve the comparability of cost-effectiveness estimates between different studies, recommendations guiding the conduct of cost-effectiveness analysis are developed by the Panel on Cost-Effectiveness in Health and Medicine (*15*).

## Use of mathematical modelling in evaluation of cancer screening

Pidd (*16*) defined a model as an external and explicit representation of part of reality as seen by the people who wish to use that model to understand, to change, to manage and to control that part of reality. This general definition also applies to mathematical models in cancer screening.

Major uses of mathematical models in cancer screening are for data analysis, evaluation/monitoring and planning/optimisation. Data analysis models are used to test hypotheses about the natural history of the disease, characteristics of the screening test and the association between early detection and risk of dying from the cancer. Evaluation and monitoring concern the assessment of the expected results of an existing screening programme, and their comparison to the observed results. These comparisons may concern early outcomes of screening programmes (coverage, detection rates, stage distribution) and late outcomes (mortality). From this comparison, recommendations may follow for improvements in the performance of the screening programme or if indicated, the model may be improved. If, for example, the observed detection rates of a screening programme are lower than expected, measures may be taken to improve the performance of screening, or if this is not feasible, assumptions on the sensitivity of the screening test in the model should be adapted, and consequences for the (cost-) effectiveness of screening should be reassessed. Using modelling for planning and optimisation results in cost-effectiveness estimates and optimal strategies for screening programmes to be implemented in the future.

These different uses of mathematical models in cancer screening are complementary to each other. Estimates on disease durations and sensitivity may, for example, be implemented in

evaluation models and models used for planning and optimisation. Subsequently, monitoring of screening programmes may lead to adaptations in the parameterisation of the models.

A possible classification of models used in cancer screening is suggested by Bross et al (*17*). They distinguished two types: surface models and deep models. Surface models consider only events that can be directly observed, such as clinical incidence, prevalence and mortality. Deep models incorporate hypotheses on the underlying processes that determine the observed events. Deep models can be further grouped in analytic and simulation models. In analytic models, the mathematical formulation of the problem can be solved, while in simulation models explicit formulation is not possible due to the complexity of the problem. The use of an analytical model for estimating parameters in a (simple) disease model may be feasible, but for more comprehensive models, the simulation approach should be applied. In simulation models, far more parameters can be used to describe the natural history of the disease, the characteristics of the screening test and other aspects needed to answer the research question.

However, more comprehensive models should not automatically be deemed to be better. According to Occam's razor an optimum is reached using a model with as few parameters as possible to give a plausible description of reality. Although this optimum is not objective, it may be used as a guideline in modelling.

A disadvantage of comprehensive models is that the input information required is often not readily available. For parameterisation of a model, different data sources usually have to be used, and if no data are available, a best guess is needed. It is clear that after the parameterisation of the model, the validity of the model has to be established, before it can be used for evaluation and prediction of the (costs and) effects of cancer screening.

Several types of validation of mathematical models can be distinguished. A first requirement is the face validity of the model. This means that the structure of the model makes sense to people who have a good knowledge of the problem, e.g. the clinician(s). The internal validity of a model is determined by the extent to which the model reproduces the data used to estimate the parameters for the model. Finally, the external validity concerns the comparison between predictions of the model with empirical data that are not used for parameter estimation of the model.

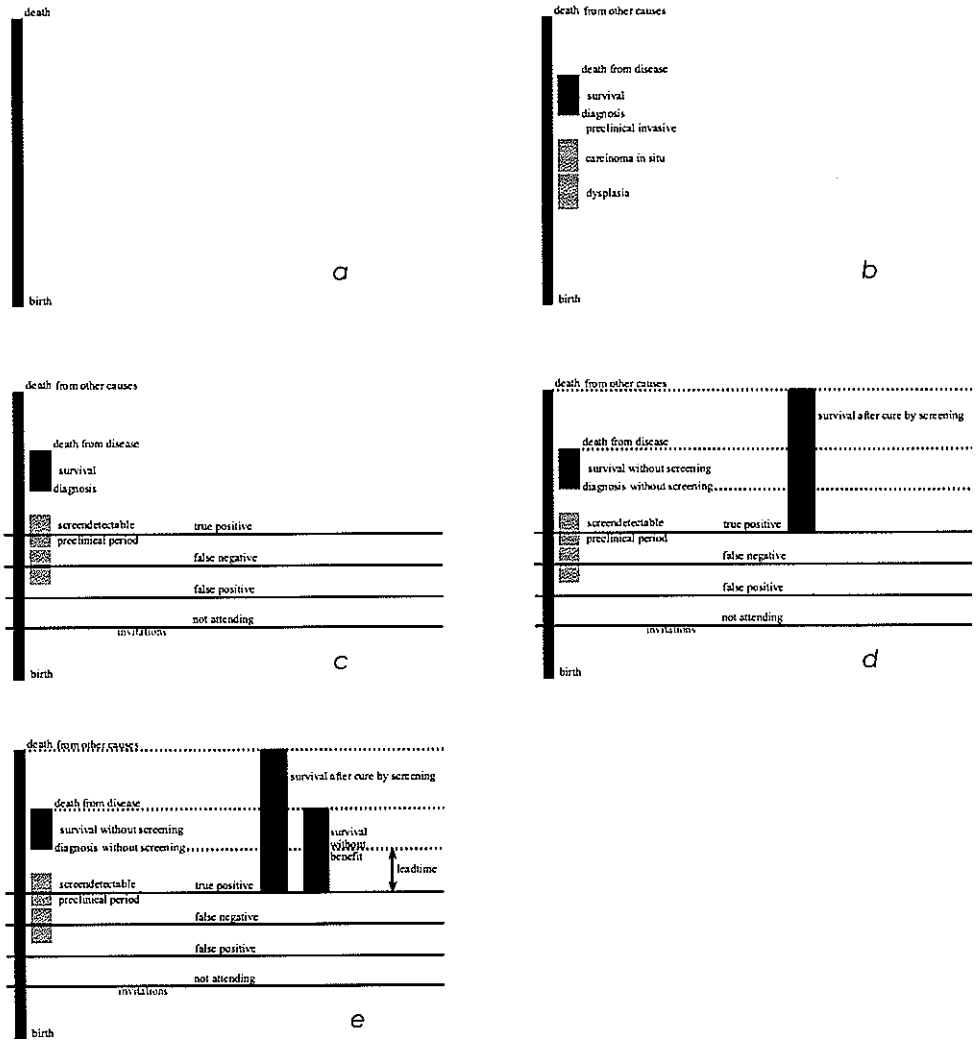
## Microsimulation Screening Analysis

The MISCAN (MIcrosimulation SCreening ANalysis) simulation program was developed for the evaluation of screening for disease, and has been applied to breast, cervical, colon and prostate cancer (*18,19*).

In MISCAN, both the situation with and without screening is simulated for the same population. First, a population is generated consisting of fictitious individual life histories without screening taking place (Figure 1.4.a). In some of these life histories cancer may develop (Figure 1.4.b). This results in output of age- and time-specific cancer incidence and mortality. Next, screening is simulated in

***Figure 1.4***

*Schematic representation of the MISCAN approach to screening for cervical cancer. Successive actions are respectively: a) A fictitious population of women is generated. b) In some persons cervical disease will develop, which in some cases will lead to death from cervical cancer. c) Screening is simulated in the fictitious population. d) Screening may postpone death due to early detection and cure of the disease. e) Screening may also detect disease that cannot be cured, as a result of which the period during which the woman is aware of her disease (lead time) is lengthened.*




this fictitious population (Figure 1.4.c). In some life histories preclinical lesions will be detected by screening, which may prevent further development of the disease and subsequent death from cancer (Figure 1.4.d). In other life histories, detection of the disease by screening will not lead to a better prognosis, and the patient will die at the same time as she would have without screening (Figure 1.4.e). The aggregated changes in life histories constitute the effectiveness of screening. The model specifications include demographic characteristics, the epidemiology and natural history of

the disease, and the screening characteristics. These characteristics are adapted to the population under study.

Furthermore, the costs of screening, diagnostic and treatment procedures can be specified to assess the cost-effectiveness. If the loss in quality of life in the different stages of screening and the disease are known, the cost-effectiveness of screening can be expressed as the costs per quality adjusted life year gained.

# SCREENING FOR BREAST CANCER

## Introduction

Some of the major established risk factors for breast cancer are a family history of breast cancer, early menarche, late age at first childbirth, late age at menopause, and exposure to ionizing radiation. Most of the above are associated with only a weak or moderately elevated risk for breast cancer (relative risk ≤ 3 (*20*)) and are hardly suitable for primary prevention. A family history of breast cancer, particularly if a diagnosis of breast cancer was made in first-degree relatives or in first-degree relatives at young age results in higher risks (*20*). Women with a BRCA1 or BRCA 2 mutation have a cumulative risk of invasive breast cancer of about 55-85%. For these women, prophylactic mastectomy may be an option. This may then be considered as primary prevention (*21-24*). Intensive screening (with or without chemoprevention) is another possibility. However it may be possible that surveillance may well fail to detect breast carcinoma at an early, curable stage in young women with a hereditary risk of breast cancer, because the growth rate in these women might be rapid (*25*) and the density of the breast tissue at this age may complicate detection by mammography.

The natural history can be described by a succession of several invasive stages, characterised by the size of the tumor, the involvement of lymph nodes, and the presence of metastases. In a number of cases, the invasive disease is preceded by the non-invasive stage known as ductal carcinoma in situ. The early invasive stages of breast cancer are often without clear symptoms of the disease. Only in later invasive stages do these symptoms manifest clearly, enabling a diagnosis of breast cancer to be made. This is usually when the woman has felt a lump or change in her breast. Metastasis of breast cancer commonly starts to the adjacent lymph nodes in the axilla. The likelihood of axillary-node metastasis is related to the size of the invasive cancer.

The main methods for early detection of breast cancer have been mammography and physical examination performed by a trained health professional. Mammography screening is generally used in screening programs for breast cancer in high-income countries. Sensitivities of mammography of between 63% and more than 90% have been reported (*26-30*). Next to actual differences in the manner in which screening is performed, the differences in reported sensitivity of mammography may be caused by characteristics of the women (age, breast density) (*28,29*), differences in screening methods (single vs. double view, single vs. double reading) (*30*) and differences in the assessment of the sensitivity caused by differences in follow-up period (*29,31*).

Physical examination performed by a trained health professional is sometimes used as screening method additional to mammography or instead of mammography, with variable results. In the HIP study performed in the 1960s in the USA, the sensitivity of physical examination was estimated to be considerably higher than the sensitivity of mammography, while in the Dutch pilot projects only a small increase in sensitivity was achieved by adding physical examination to mammography (32). Finally, Miller et al. found no additional effect on mortality if mammography plus physical examination was compared to physical examination alone (33). Currently the possibility of screening by magnetic resonance imaging (MRI) is under study. Especially in younger women where the performance of mammography is assumed to be less than in older women, screening by MRI may become an alternative (34,35). Breast self-examination is another potential method of screening, although studies on the value of breast self-examination have shown this to be relatively ineffective (36). Offering instruction in breast self-examination as primary screening method should, however, be distinguished from the process by which women perceive symptoms and seek care for them. The increase in health awareness in the last decades has improved the stage distribution of clinically diagnosed breast cancers.

## Evaluation of breast cancer screening

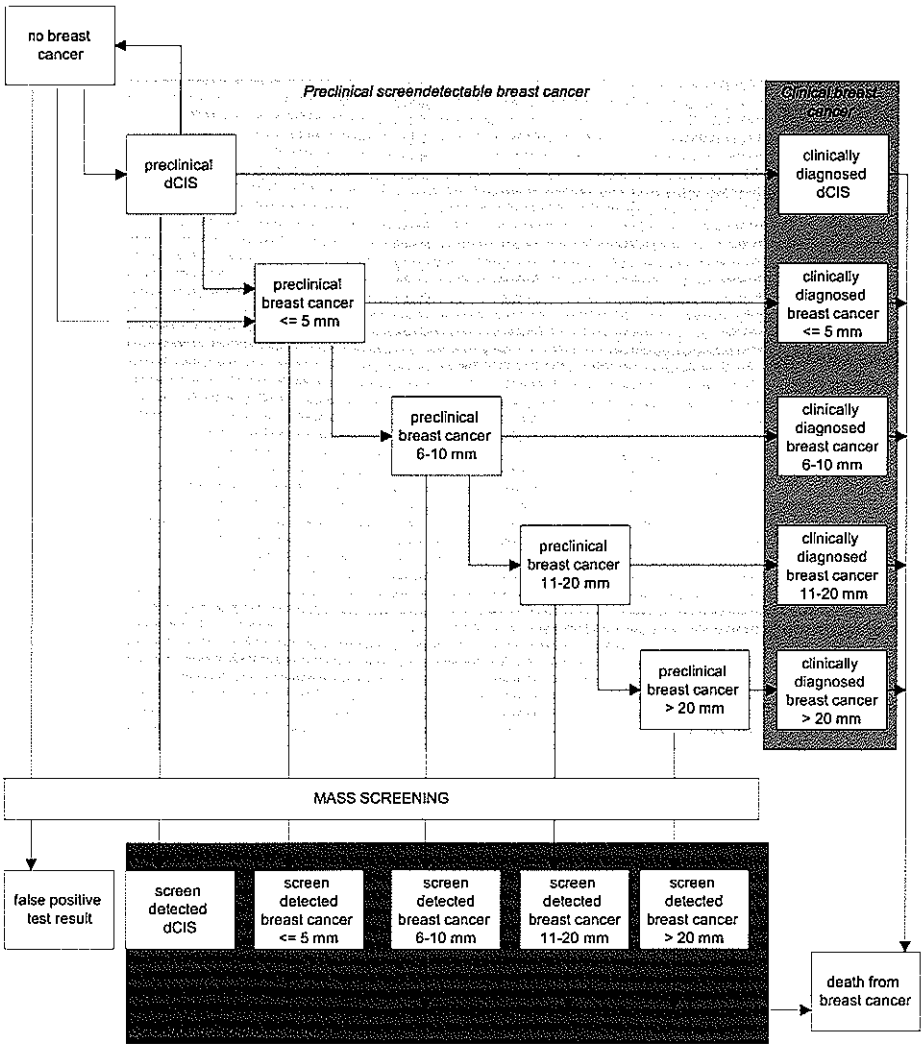
The effectiveness of mammography has been investigated in several randomized controlled trials in New York (USA), Edinburgh (Scotland), Canada, and Malmö, Kopparberg, Östergötland, Stockholm and Göteborg in Sweden. A meta-analysis of updated results of the five Swedish trials shows a breast cancer mortality reduction of 29% in women aged 50-69 years (37).
Recently, the results of these trials were criticised (38,39), but this criticism was convincingly rebutted (40-43).
For women over age 70, the positive effects of breast cancer screening are estimated to outweigh the negative effects until at least age 75. Extending the upper age limit from 69 to 75 may render the cost-effectiveness of screening less favourable than screening between ages 50-69, but might be acceptable. Due to the disproportional rise in negative effects of screening in older women, the estimated cost-effectiveness of screening women aged 75 and up will be reduced (44). For women under the age of 50, the effectiveness of screening has not yet been established (41,45).

## Use of mathematical modelling in evaluation of breast cancer screening

From the beginning, models for breast cancer were focussed on the influence of screening on the disease and the effects of different screening policies. Early on, Zelen and Feinleib (46) used an elementary model to demonstrate the existence of length bias sampling and the way it influences the lead time. Many models were fitted using the results of the HIP randomized trial, and subsequently used for predictions of cancer strategies (32,47-55). These differed in age range, interval between screenings, screening method (mammography, palpation and breast self examination), and different screening doses for mammography. Later on, the results of other trials in

***Figure 1.5***

*Structure of the MISCAN model for breast cancer*




Sweden, the Netherlands, the United Kingdom and Canada, also became available and were used for parameterisation of models on breast cancer screening (56-59).

## MISCAN breast cancer model

The basic structure of the MISCAN breast cancer model is shown in Figure 1.5. The first state is the state without breast cancer. During the lifetime of a woman, a transition may occur to preclinical screen-detectable breast cancer. A proportion of the cancers are assumed to be preceded by

*Table 1.1*

*Important model parameters for natural history and screening in the MISCAN breast cancer model*

| **Mean duration (years) of preclinical stage by age** | | | | | |
|---|---|---|---|---|---|
| *Stage* | | | *40 years* | *50 years* | *60 years* | *70 years* |
| Preclinical dCIS | | | 5.2 | 5.2 | 5.2 | 5.2 |
| Preclinical T1a (tumor <= 5 mm) | | | 0.1 | 0.1 | 0.1 | 0.2 |
| Preclinical T1b (tumor 6-10 mm) | | | 0.4 | 0.5 | 0.7 | 0.9 |
| Preclinical T1c (tumor 11-20 mm) | | | 0.8 | 1.0 | 1.5 | 1.8 |
| Preclinical T2+ (tumor > 20 mm) | | | 0.6 | 0.8 | 1.1 | 1.4 |
| **Long-term relative survival by clinical stage and age** | | | | | |
| *Age* | *DCIS* | *T1a* | *T1b* | *T1c* | *T2+* | |
| 40 | 1.00 | 0.86 | 0.79 | 0.63 | 0.42 | |
| 50 | 1.00 | 0.86 | 0.79 | 0.63 | 0.41 | |
| 60 | 1.00 | 0.83 | 0.75 | 0.56 | 0.31 | |
| 70 | 1.00 | 0.85 | 0.78 | 0.61 | 0.39 | |
| **Probability of surviving by time since diagnosis and stage** | | | | | |
| *Time since diagnosis* | | | *T1a* | *T1b* | *T1c* | *T2+* |
| 1 year | | | 0.94 | 0.95 | 0.95 | 0.91 |
| 3 years | | | 0.75 | 0.85 | 0.84 | 0.70 |
| 5 years | | | 0.60 | 0.63 | 0.61 | 0.52 |
| 7 years | | | 0.50 | 0.48 | 0.44 | 0.40 |
| 10 years | | | 0.39 | 0.30 | 0.21 | 0.30 |
| 20 years | | | 0.20 | 0.18 | 0.15 | 0.13 |
| 30 years | | | 0.12 | 0.11 | 0.08 | 0.07 |
| 50 years | | | 0.00 | 0.00 | 0.00 | 0.00 |
| **Sensitivity of mammography by stage and age** | | | | | |
| *Stage* | | | *40-44 years* | *45-49 years* | *>= 50 years* |
| Preclinical dCIS | | | 24% | 32% | 40% |
| Preclinical T1a (tumor <= 5 mm) | | | 39% | 52% | 65% |
| Preclinical T1b (tumor 6-10 mm) | | | 48% | 64% | 80% |
| Preclinical T1c (tumor 11-20 mm) | | | 54% | 72% | 90% |
| Preclinical T2+ (tumor > 20 mm) | | | 57% | 76% | 95% |
| **Reduction in risk of dying of breast cancer by stage in which (pre)cancer is detected** | | | | | |
| *Stage* | | | *Reduction in risk* | | |
| Preclinical dCIS | | | 100% | | |
| Preclinical T1a (tumor <= 5 mm) | | | 89.2% | | |
| Preclinical T1b (tumor 6-10 mm) | | | 81.4% | | |
| Preclinical T1c (tumor 11-20 mm) | | | 56.7% | | |
| Preclinical T2+ (tumor > 20 mm) | | | 39.5% | | |

the screen-detectable, ductal carcinoma in situ (dCIS) state. In this state, regression may also occur. Four invasive states with increasing tumor size are distinguished: <5 mm, 6-10 mm, 11-20 mm and > 20 mm. Invasive disease will progress through these states until clinical diagnosis or screen detection of the disease takes place.

The most important parameters for natural history and screening in the model are presented in Table 1.1.

# SCREENING FOR CERVICAL CANCER

## Introduction

By now, HPV has been established as the main risk factor for cervical cancer. Women with negative smear results but a positive HPV test result were found to have a relative risk of more than 100 to develop high grade dysplasia compared to women with negative smear and negative HPV test results (*60,61*). Classical risk factors mentioned in the literature were age at first sexual intercourse, number of sexual partners, use of oral contraceptives, parity and smoking. These risk factors are distributed over risk factors for acquisition of HPV infections (number of sexual partners) and cofactors for development of cervical neoplasia among HPV infected women (early age at first intercourse, use of oral contraceptives, parity and smoking) (*62-65*).

In consideration of the risk factors for cervical cancer, a potential method for primary prevention may be vaccination against HPV. Therapeutic and prophylactic vaccines against HPV are now in progress, but it will take several years before conclusions about the effectiveness of the HPV vaccines are achieved (*66*).

The natural history of cervical cancer is characterised by a long preclinical detectable stage, consisting of a pre-invasive and a pre-clinical invasive stage. The pre-invasive stage is usually subdivided into three stages, cervical intra epithelial neoplasia (CIN) I, CIN II, and CIN III. The median duration of the total preclinical detectable stage is assessed at about 15 years (*67-70*). HPV is found in (almost) all invasive cancers (*71,72*). These HPV infections are assumed to precede the occurrence of CIN. Another characteristic property of the natural history is the considerable part of pre-invasive lesions that will regress. This regression is assumed to be age dependent, with the highest regression rates at young ages (*68*). This is compatible with the high prevalence of HPV and the high clearance of HPV observed in young women (*73,74*).

The prognosis for (screen) detected CIN is very good. At this stage, the survival is close to 100%, even though retreatment is necessary in up to 25% of the women initially treated for CIN because of persistence or recurrence of disease (*75,76*). The prognosis in the local invasive stages is also good. This aspect, together with the long preclinical stage, combines to fulfil the conditions necessary for effective screening (i.e. an identifiable latent phase or early symptomatic stage and a better prognosis in case of early detection). The Pap smear is widely used as screening test. Sensitivities for cytology have been reported that range between 29% and 80% for CIN in screening populations (*68,77-80*). Next to differences in screening procedures, the differences in sensitivity may also originate from the evaluation of Pap smears, due to different cut-off points in definition of a positive smear and cervical disease, and method of determining the sensitivity. Intensive follow-up of slightly positive smears will lead to a higher sensitivity but decrease the specificity, and might disturb the delicate balance between benefits, harm and costs of cervical cancer screening. Currently, the possible role of the HPV test in primary screening is being investigated in several longitudinal studies.

The use of cytology or HPV tests in screening programs is not feasible in low-income countries, as this requires a high degree of organisation with laboratories and personnel. In these situations (aided) visual inspection of the uterine cervix may be an alternative solution (*81,82*).

Women participating in cervical cancer screening programmes are found to be at lower risk than non-participating women (*68,83,84*). However, screening results will suffer if high-risk women fail to be reached by screening. Currently, the use of self-sampling is considered as an alternative screening tool for unscreened women (*85*).

## Evaluation of cervical cancer screening

Unlike for breast cancer, no randomized controlled trial has ever been performed for cervical cancer. The effectiveness of cervical cancer screening has always been debated, a discussion that has been fanned by the controversial results reported in the literature. Cervical cancer screening was found to be effective in several observational studies (*86-88*). A spectacular 80% mortality drop was found in Iceland between 1965-82 (*87*). The fall in mortality in Finland and Sweden was modest (respectively 50% and 34%), while in Norway a decrease of no more than 10% was found. These differences are correlated with the width of the target age and the coverage of the screening programmes in the respective countries, and support the conclusion that screening is effective in reducing the cervical mortality. In England and Wales, however, no discernable impact of screening on the mortality rate had been able to be ascertained by the mid-1980s (*89*).

## Use of mathematical modelling in evaluation of cervical cancer screening

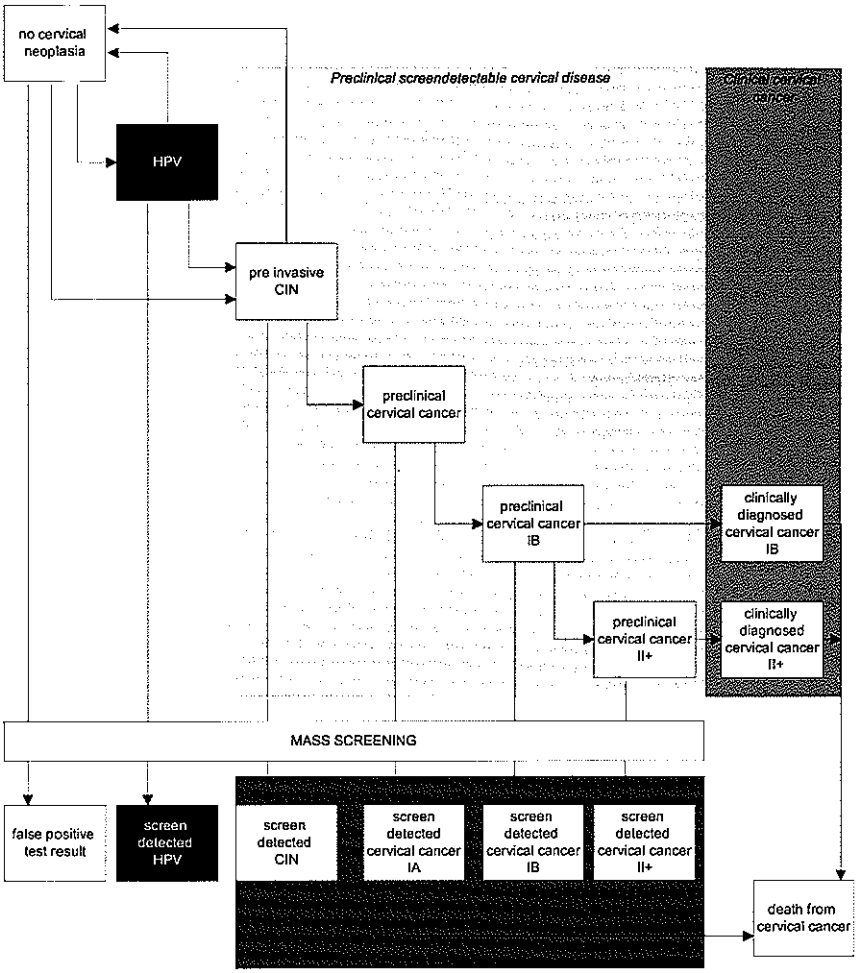
In cervical cancer screening a number of models attempt to draw inferences about the natural history of cervical cancer. A first attempt to study the natural history of cervical dysplasia and carcinoma in situ was made by Barron and Richart (*90,91*). More sophisticated models were designed in later years by Coppleson and Brown (92), Albert (*93*), Brookmeyer and Day (*94*), Gustafsson and Adami (*67*), and van Oortmarssen and Habbema (*68*). These models revealed some important features of the natural history of cervical cancer, as it became evident that the observed data could not be explained without including the possibility of regression of preinvasive cervical lesions and age-dependencies of transitions between stages in the model.

There are numerous models to estimate the (cost-) effectiveness of cervical cancer screening and/or identify optimal screening policies. These models vary in comprehensiveness. Examples of surface models are the age-period multiplicative model used by Hristova and Hakama (*95*) and the regression model by Forsmo et al. (96). Also, the articles of Chesebro and Everett (*97*), Waugh et al. (*98,99*) may be categorised as surface models, but these add some assumptions, for example, the percentage of women with preclinical invasive disease that will develop invasive cervical cancer.

A precursor of the recent detailed and comprehensive models, was the model of Knox (*100*). A model including the mean duration of dysplasia and carcinoma in situ and a false-negative rate

*Figure 1.6*

*Basic structure of the MISCAN model for cervical cancer*




for the screening test was used to calculate the best ages for carrying out cervical cancer screening. Examples of recent deep models on cervical cancer screening are the models of Eddy (*101*), Myers (*102,103*), Gyrd-Hansen (*104*), and Sherlaw-Johnson (*105*) used for cost-effectiveness calculations and the model of Gustafsson and Adami, which is used to estimate efficient screening policies. The MISCAN model, which is described below in more detail, is used for both cost-effectiveness calculations and for identification of efficient screening policies for cervical cancer. Over the past decade, several models were developed that included HPV in the natural history, with a view to deriving estimates of parameters for the natural history of the HPV infection and characteristics of the HPV test, as well as to estimate the cost-effectiveness of the use of the HPV

*Table 1.2*

*Important model parameters for natural history and screening in the MISCAN cervical cancer model.(6,7)*

| | | |
|---|---|---|
| **Duration of preclinical stages** | | |
| *Stage* | | *Mean duration (years)* |
| Pre-invasive neoplasia | | 11.8 |
| Micro-invasive IA | | 2.0 |
| Preclinical IB – clinical IB | | 1.9 |
| Preclinical IB - pre-clinical II+ | | 1.0 |
| Preclinical II+ | | 0.9 |
| Total pre-clinical | | 15.7 |
| **Long-term relative survival by clinical stage and age** | | |
| *Age* | *IB* | *II+* |
| < 25 years | 0.70 | 0.20 |
| 30 years | 0.81 | 0.50 |
| 50 years | 0.81 | 0.50 |
| >65 years | 0.62 | 0.00 |
| **Probability of surviving by time since diagnosis** | | |
| *Time since diagnosis* | | *Probability* |
| 1.5 years | | 0.64 |
| 4 years | | 0.21 |
| 7 years | | 0.14 |
| 99 years | | 0.00 |
| **Sensitivity of Pap smear by stage** | | |
| *Stage* | | *Sensitivity* |
| Pre-invasive neoplasia | | 80% |
| Micro-invasive IA | | 85% |
| Preclinical IB | | 85% |
| Preclinical II+ | | 90% |
| **Reduction in risk of dying of cervical cancer by stage in which (pre)cancer is detected** | | |
| *Stage* | | *Reduction in risk* |
| Pre-invasive neoplasia | | 100% |
| Micro-invasive IA | | 80% |
| Preclinical IB | | 40% |
| Preclinical II+ | | 20% |

test as a screening test (primary screening) or in the follow-up of screen positive women (secondary screening) (*106-111*).

## MISCAN cervical cancer model

The structure of the MISCAN cervical cancer model is shown in Figure 1.6. The cancer development process starts with a transition to pre-invasive cervical disease (CIN). Regression of CIN may occur, or the disease may progress to a preclinical invasive stage. Three preclinical invasive stages are distinguished: Ia, Ib and II+ which correspond with the FIGO stages (*112*). In the presence of screening, screen detection of preclinical disease may take place or the disease will surface clinically in the stages Ib or II+. In recent years the HPV stage has been added. The majority of the women developing invasive cervical cancer will have experienced an HPV infection.

The quantification of the most important parameters for the natural history and screening in the cervical cancer model is given in Table 1.2. A detailed description can be found in chapter 5.

# THIS THESIS

## Research questions

The objective of this thesis was to investigate important questions in screening for breast and cervical cancer. This was done by mathematical modelling, of which the usefulness was also evaluated.

Questions to be addressed are:

I What is the applicability of the MISCAN breast cancer screening model in new situations?

II How much breast cancer mortality reduction can be expected in the first years after introduction of breast cancer screening ?

III Which Pap smear-based cervical screening policies are optimal with respect to cost-effectiveness?

IV What is the cost-effectiveness of the cervical screening practice in different European countries?

V What is the incidence after negative screening in the Netherlands and does it correspond to the results from a multicountry analysis on which most important models for cervical cancer screening have based their assumptions?

VI What is the evidence for the (cost-) effectiveness of HPV testing in primary cervical cancer screening?

## Structure of the thesis

In Chapter 2, the screening programme for breast cancer in Navarra (Spain) is evaluated using MISCAN on the basis of first and second round screening results.

Chapter 3 reports the breast cancer mortality reduction that might be expected in the first years of a breast cancer screening programme. The observed mortality in the Netherlands and the United Kingdom is compared to the predicted mortality by MISCAN.

In chapter 4, the incidence of invasive carcinoma after negative screening is compared to the incidence in a situation without screening. These results provide an indication of the performance of the screening programme if compared to similar data of other screening programmes. Furthermore, the results provide useful information for validation of important parameters of a cervical cancer screening model.

In chapter 5, the costs and effects of a large number of screening programmes for cervical cancer are estimated using the MISCAN simulation programme, resulting in identification of efficient screening policies using data on demography, epidemiology, test characteristics and costs representative for the Dutch situation. The background characteristics are translated for other

countries with respect to incidence and price level to investigate whether the current wide diversity in screening programmes for cervical cancer screening could be explained.

In chapter 6, the cost-effectiveness of running cervical cancer screening programmes in the European Union is calculated on the basis of the observed screening intensity, consisting of recommended number of smears per lifetime and the number of excess smears on top of these recommendations.

Chapter 7 concerns a study on the natural history of HPV infections and the characteristics of the HPV test using cross-sectional data. The costs and effects of different screening strategies are calculated for two possible sets of parameters on the basis of these data. In the meantime longitudinal studies are being performed.

In chapter 8, the previous study is updated by using longitudinal data that has become available in the meantime.

In the general discussion (Chapter 9), the research questions underlying this thesis, as formulated above, are answered. Furthermore, the use of modelling in breast and cervical cancer screening is discussed.

## REFERENCES


(1) Netherlands Central Bureau of Statistics. Death by causes of death, age and gender 1998: Netherlands Central Bureau of Statistics, published annually.

(2) Alexander FE. Statistical analysis of population screening. Med Lab Sci 1989;46:255-67.

(3) Netherlands Central Cancer Registration (CCR). Incidence of carcinoma of the cervix uteri in Friesland, Rotterdam and The Hague, 1960-1970. Dutch Cancer Society; 1973.

(4) SIG- Information Centre for Health Care. LMR data on hospital admissions, 1963-1985. Utrecht.

(5) Netherlands Central Bureau of Statistics. Death by causes of death, age and gender. Series A1, 1950-1993. Voorburg: CBS; 1995.

(6) van Ballegooijen M. Effects and costs of cervical cancer screening. Department of Public Health. Rotterdam: Erasmus University; 1998.

(7) van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. J Natl Cancer Inst 2002;94:193-204.

(8) Visser O, Coebergh JWW, Schouten LJ, Van Dijck JAAM. Incidence of Cancer in the Netherlands 1997: ninth report of the netherlands cancer registry. Utrecht: Vereniging van Integrale Kankercentra; 2001.

(9) IARC. Globocan 2000: Cancer Incidence, Mortality and Prevalence Worldwide, Version 1.0. IARC CancerBase No.5. Lyon, France; 2000.

(10) van Klaveren RJ, J.D.F. H, Pedersen JH, de Koning HJ, Oudkerk M, Hoogsteden HC. Lung cancer screening by low-dose spiral computed tomography. Eur Respir J 2001;18:855-66.

(11) Patz EF, Black WC, Goodman PC. CT screening for lung cancer: not ready for routine practice. Radiology 2001;221:587-9.

(12) Connor RJ, Boer R, Prorok PC, Weed DL. Investigation of design and bias issues in case-control studies of cancer screening using microsimulation. Am J Epidemiol 2000;151:991-8.

(13) Boer R, de Koning HJ, van Ballegooijen M, van der Maas PJ. Important influences on effectiveness and costs to be considered in the evaluation of cancer screening. In Quantative methods of evaluation of cancer screening.

(14) Wright JC, Weinstein MC. Gains in life expectancy from medical interventions--standardizing data on outcomes. N Engl J Med 1998;339:380-6.

(15) Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, Inc.; 1996.

(16) Pidd M. Tools for thinking: modelling in management science. Chicester: John Wiley & Sons, Ltd; 1996.

(17) Bross I, Blumenson L, Slack N, Priore R. A two disease model for breast cancer. In Bunkler P, editor. Prognostic factors in breast cancer. Baltimore: Williams and Wilkins; 1968. p. 288-300.

(18) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.

(19) Loeve F, Boer R, van Oortmarssen G, van Ballegooijen M, Habbema J. The MISCAN-COLON simulation model for the evaluation of colorectal cancer. Comput Biomed Res 1999;32:13-33.
(20) Harris JR, Lippman ME, Veronesi U, Willett W. Breast Cancer (First of Three Parts). N Engl J Med 1992;327:319-28.
(21) Schrag D, Kuntz KM, Garber JE, Weeks JC. Decision analysis--effects of prophylactic mastectomy and oophorectomy on life expectancy among women with BRCA1 or BRCA2 mutations. N Engl J Med 1997;336:1465-71.
(22) Hartmann LC, Schaid DJ, Woods JE, Crotty TP, Myers JL, Arnold PG, et al. Efficacy of bilateral prophylactic mastectomy in women with a family history of breast cancer. N Engl J Med 1999;340:77-84.
(23) Hughes KS, Papa MZ, Whitney T, McLellan R. Prophylactic mastectomy and inherited predisposition to breast carcinoma. Cancer 1999;86:2502-16.
(24) Meijers-Heijboer EJ, Verhoog LC, Brekelmans CT, Seynaeve C, Tilanus-Linthorst MM, Wagner A, et al. Presymptomatic DNA testing and prophylactic surgery in families with a BRCA1 or BRCA2 mutation. Lancet 2000;355:2015-20.
(25) Breast Cancer Linkage Consortium. Pathology of familial breast cancer: differences between breast cancers in carriers of BRCA1 or BRCA2 mutations and sporadic cases. Lancet 1997;349:1505-10.
(26) Baines CJ, McFarlane DV, Miller AB. Sensitivity and specificity of first screen mammography in 15 NBSS centres. Can Assoc Radiol J 1988;39:273-6.
(27) Burhenne HJ, Burhenne LW, Goldberg F, Hislop TG, Worth AJ, Rebbeck PM, et al. Interval breast cancers in the Screening Mammography Program of British Columbia: analysis and classification. AJR Am J Roentgenol 1994;162:1067-71; discussion 72-5.
(28) Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Likelihood ratios for modern screening mammography. Risk of breast cancer based on age and mammographic interpretation. Jama 1996;276:39-43.
(29) Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. Jama 1996;276:33-8.
(30) Blanks RG, Wallis MG, Given-Wilson RM. Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography. J Med Screen 1999;6:152-8.
(31) Brekelmans CT, Collette HJ, Collette C, Fracheboud J, de Waard F. Breast cancer after a negative screen: follow-up of women participating in the DOM Screening Programme. Eur J Cancer 1992;28A:893-5.
(32) van Oortmarssen GJ, Habbema JDF, Lubbe JTN, van der Maas PJvd. A model-based analysis of the HIP project for breast cancer screening. Int J Cancer 1990;46:207-13.
(33) Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490-9.
(34) Stoutjesdijk MJ, Boetes C, Jager GJ, Beex L, Bult P, Hendriks JH, et al. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. J Natl Cancer Inst 2001;93:1095-102.
(35) Kriege M, Brekelmans CT, Boetes C, Rutgers EJT, Oosterwijk JC, Tollenaar RAEM, et al. MRI screening for breast cancer in women with familial or genetic predisposition: design of the Dutch National Study (MRISC). Familial Cancer 2001;1:163-8.
(36) Morrison AS. Is self-examination effective in screening for breast cancer? J Natl Cancer Inst 1991;83:226-7.
(37) Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O. Reduction in mortality from breasst cancer after mass screening with mammography. Lancet 1985;i:829-32.
(38) Gotsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355:129-34.
(39) Giard RWM, Bonneux LGA. Borstkankerscreening onvoldoende effectief. Ned Tijdschr Geneeskd 2001;145:2205-08.
(40) Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gotsche and Olsen. The Breast 2001.
(41) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002;359:909-19.
(42) de Koning HJ. Assessment of nationwide cancer-screening programmes. Lancet 2000;355:80-1.
(43) de Koning HJ, Fracheboud J, Verbeek ALM, Rutgers EJT, van der Maas PJ. De wetenschappelijke basis van het bevolkingsonderzoek naar borstkanker in Nederland. Ned Tijdschr Geneeskd 2002;146:1034-41.
(44) Boer R, de Koning HJ, van Oortmarssen GJ, van der Maas PJ. In search of the best upper age limit for breast cancer screening. Eur J Cancer 1995;31A:2040-3.
(45) Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet 1999;353:1903-8.
(46) Zelen M, Feinleib M. On the theory of screening for chronic diseses. Biometrika 1969;56:601-14.
(47) Chiacchierini RP, Lundin FE. Benefit/risk ratio of mammography. In Logan WW, editor. Breast carcinoma: THe radiologist's expanded role. New York: Wiley; 1977. p. 15-28.
(48) Chiacchierini RP, Lundin FE. Risk-benefit analysis for reduced dose mammography. In Muntz EP, editor. Reduced dose mammography. New York: Masson Publishing; 1979.

(49) Chiacchierini RP, Lundin FE, Scheidt PC. A risk/benefit analysus by life table modeling of an annual breast screening program which includes X-ray mammography. In Nieburgs HE, editor. Prevention and detection of cancer. Part II: Detection, Volume 2. New York: Marcel Dekker; 1980. p. 1741-62.
(50) Dubin N. Benefit of screening for breast cancer: application of a probabilistic model to a breast cancer detection project. J Chron Dis 1979;32:145-51.
(51) Dubin N. Predicting the benefit of screening for disease. J Appl Prob 1981;18:348-60.
(52) Dubin N. Effect of different mammographic radiation exposures on predicted effects of screening for breast cancer. Statis Med 1982;1:15-24.
(53) Shwartz M. A mathematical model used to analyze breast cancer screening strategies. Operations Res 1978;26:937-55.
(54) Knox EG. Simulation studies of breast cancer screening programmes. London: Oxford University Press; 1975.
(55) Eddy DM. A mathematical model on the efficacy of breast cancer screening. In McLelland R, editor. Breast carcinoma: Current diagnosis and treatment. New York: Masson Publishing; 1983. p. 339-49.
(56) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. Journal of Epidemiology and Biostatistics 1997;2:25-35.
(57) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the preclinical screen-detectable phase. Journal of Epidemiology and Biostatistics 1997;2:9-23.
(58) Day NE, Duffy SW. Trial design based on surrogate endpoints - application to comparison of different breast screening frequencies. J Royal Stat Soc, Series A 1996;159:49-60.
(59) Duffy SW, Chen HH, Tabar L, Day NE. Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. Stat Med 1995;14:1531-43.
(60) Rozendaal L, Walboomers JM, van der Linden JC, Voorhorst FJ, Kenemans P, Helmerhorst TJ, et al. PCR-based high-risk HPV test in cervical cancer screening gives objective risk assessment of women with cytomorphologically normal cervical smears. Int J Cancer 1996;68:766-9.
(61) Rozendaal L, Westerga J, van der Linden JC, Walboomers JM, Voorhorst FJ, Risse EK, et al. PCR based high risk HPV testing is superior to neural network based screening for predicting incident CIN III in women with normal cytology and borderline changes. J Clin Pathol 2000;53:606-11.
(62) Ho GY, Kadish AS, Burk RD, Basu J, Palan PR, Mikhail M, et al. HPV 16 and cigarette smoking as risk factors for high-grade cervical intra-epithelial neoplasia. Int J Cancer 1998;78:281-5.
(63) Deacon JM, Evans CD, Yule R, Desai M, Binns W, Taylor C, et al. Sexual behaviour and smoking as determinants of cervical HPV infection and of CIN3 among those infected: a case-control study nested within the Manchester cohort. Br J Cancer 2000;83:1565-72.
(64) Moreno V, Bosch FX, Munoz N, Meijer CJ, Shah KV, Walboomers JM, et al. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. Lancet 2002;359:1085-92.
(65) Munoz N, Franceschi S, Bosetti C, Moreno V, Herrero R, Smith JS, et al. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. Lancet 2002;359:1093-101.
(66) Schiller JT, Hidesheim A. Developing HPV virus-like particle vaccines to prevent cervical cancer: a progress report. J Clin Virol 2000;19:67-74.
(67) Gustafsson L, Adami HO. Natural History Of Cervical Neoplasia: Consistent Results Obtained By an Identification Technique [See Comments]. Br J Cancer 1989;60:132-41.
(68) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.
(69) van Oortmarssen GJ, Habbema JD, van Ballegooijen M. Predicting mortality from cervical cancer after negative smear test results. Bmj 1992;305:449-51.
(70) van Oortmarssen GJ, Habbema JD. Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative pap smears. Int J Epidemiol 1995;24:300-7.
(71) Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796-802.
(72) Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 1999;189:12-9.
(73) Melkert PW, Hopman E, van den Brule AJ, Risse EK, van Diest PJ, Bleker OP, et al. Prevalence of HPV in cytomorphologically normal cervical smears, as determined by the polymerase chain reaction, is age-dependent. Int J Cancer 1993;53:919-23.
(74) Evander M, Edlund K, Gustafsson A, Jonsson M, Karlsson R, Rylander E, et al. Human papillomavirus infection is transient in young women: a population-based cohort study. J Infect Dis 1995;171:1026-30.
(75) Alvarez RD, Helm CW, Edwards RP, Naumann RW, Partridge EE, Shingleton HM, et al. Prospective randomized trial of LLETZ versus laser ablation in patients with cervical intraepithelial neoplasia. Gynecol Oncol 1994;52:175-9.

(76) Mitchell MF, Tortolero-Luna G, Cook E, Whittaker L, Rhodes-Morris H, Silva E. A randomized clinical trial of cryotherapy, laser vaporization, and loop electrosurgical excision for treatment of squamous intraepithelial lesions of the cervix. Obstet Gynecol 1998;92:737-44.
(77) Hockstad RL. A comparison of simultaneous cervical cytology, HPV testing, and colposcopy. Fam Pract Res J 1992;12:53-60.
(78) Davison JM, Marty JJ. Detecting premalignant cervical lesions. Contribution of screening colposcopy to cytology. J Reprod Med 1994;39:388-92.
(79) Baldauf JJ, Dreyfus M, Lehmann M, Ritter J, Philippe E. Cervical cancer screening with cervicography and cytology. Eur J Obstet Gynecol Reprod Biol 1995;58:33-9.
(80) Ratnam S, Franco EL, Ferenczy A. Human papillomavirus testing for primary screening of cervical cancer precursors. Cancer Epidemiol Biomarkers Prev 2000;9:945-51.
(81) Sankaranarayanan R, Shyamalakumary B, Wesley R, Sreedevi Amma N, Parkin DM, Nair MK. Visual inspection with acetic acid in the early detection of cervical cancer and precursors. Int J Cancer 1999;80:161-3.
(82) Sankaranarayanan R, Wesley R, Somanathan T, Dhakad N, Shyamalakumary B, Amma NS, et al. Visual inspection of the uterine cervix after the application of acetic acid in the detection of cervical carcinoma and its precursors. Cancer 1998;83:2150-6.
(83) 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.
(84) Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-6.
(85) Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, Rozendaal L, Jaspars LH, Voorhorst FJ, et al. Primary screening for high risk HPV by home obtained cervicovaginal lavage is an alternative screening tool for unscreened women. J Clin Pathol 2002;55:435-9.
(86) Clarke EA, Anderson TW. Does screening by "Pap" smears help prevent cervical cancer? A case-control study. Lancet 1979;2:1-4.
(87) Laara E, Day NE, Hakama M. Trends In Mortality From Cervical Cancer In the Nordic Countries: Association With Organised Screening Programmes. Lancet 1987;1:1247-9.
(88) van der Graaf Y, Zielhuis GA, Peer PG, Vooijs PG. The effectiveness of cervical screening: a population-based case-control study. J Clin Epidemiol 1988;41:21-6.
(89) Patnick J. Cervical cancer screening in England. Eur J Cancer 2000;36:2205-8.
(90) Barron BA, Richart RM. A statistical model of the natural history of cervical carcinoma based on a prospective study of 557 cases. J Natl Cancer Inst 1968;41:1343-53.
(91) Richart RM, Barron BA. A follow-up of patients with cervical dysplasia. Am J Obstet Gynecol 1969;105:386-93.
(92) Coppleson LW, Brown B. Observations on a model of the biology of carcinoma of the cervix: a poor fit between observation and theory. Am J Obstet Gynecol 1975;122:127-36.
(93) Albert A. Estimated cervical cancer disease state incidence and transition rates. J Natl Cancer Inst 1981;67:571-6.
(94) Brookmeyer R, Day NE. Two-Stage Models For the Analysis Of Cancer Screening Data. Biometrics 1987;43:657-69.
(95) Hristova L, Hakama M. Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. Acta Oncol 1997;36:1-60.
(96) Forsmo S, Buhaug H, Skjeldestad FE, Haugen OA. Treatment of pre-invasive conditions during opportunistic screening and its effectiveness on cervical cancer in one Norwegian county. Int J Cancer 1997;71:4-8.
(97) Chesebro MJ, Everett WD. A cost-benefit analysis of colposcopy for cervical squamous intraepithelial lesions found on Papanicolaou smear. Arch Fam Med 1996;1996:576-81.
(98) Waugh N, Smith I, Robertson A, Reid GS, Halkerston R, Grants A. Costs and benefits of cervical cancer screening. I. The costs of the cervical screening programme. Cytopathol 1996;7:231-40.
(99) Waugh N, Robertson A. Costs and benefits of cervical cancer screening. II. Is it worthwile reducing the screening interval from 5 to 3 years. Cythopathol 1996;7:241-8.
(100) Knox EG. A simulation system for screening procedures. In McLachlan G, editor. The Future- and Present Indicatives, Problems and Progress in Medical Care (Ninth Series). London: Oxford University Press; 1973. p. 17-55.
(101) Eddy DM. Screening For Cervical Cancer. Ann Intern Med 1990;113:214-26.
(102) Bastian L, Datta S, Hasselblad V, Hickey J, Myers E, Nanda K. Evaluation of Cervical Cytology. Rockville, MD: Agency for Health Care Policy and Research; 1999.
(103) McCrory D, Matchar D, Bastian L, Datta S, Hasselblad V, Hickey J, et al. Evaluation of Cervical Cytology. Rockville, MD: Agency for Health Care Research; 1999.
(104) Gyrd-Hansen D, Holund B, Andersen P. A cost-effectiveness analysis of cervical cancer screening: health policy implications. Health Policy 1995;34:35-51.
(105) Sherlaw-Johnson C, Gallivan S, Jenkins D, Jones MH. Cytological screening and management of abnormalities in prevention of cervical cancer: an overview with stochastic modelling. J Clin Pathol 1994;47:430-5.

(106) Jenkins D, Sherlaw-Johnson C, Gallivan S. Can human papilloma virus testing be used to improve cervical cancer screening. Int J Cancer 1996;65:768-73.
(107) Sherlaw-Johnson C, Gallivan S, Jenkins D. Evaluating cervical cancer screening programmes for developing countries. Int J Cancer 1997;72:210-6.
(108) van Ballegooijen M, van den Akker-van Marle M, Warnerdam P, Meijer C, Walboomers J, Habbema J. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. Br J Cancer 1997;76:651-7.
(109) Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. Am J Epidemiol 2000;151:1158-71.
(110) Mandelblatt JS, Lawrence WF, Womack SM, Jacobson D, Yi B, Hwang YT, et al. Benefits and costs of using HPV testing to screen for cervical cancer. Jama 2002;287:2372-81.
(111) Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Jama 2002;287:2382-90.
(112) Changes in definitions of clinical staging for carcinoma of the cervix and ovary: International Federation of Gynecology and Obstetrics. Am J Obstet Gynecol 1987;156:263-4.

# 2

# BREAST CANCER SCREENING IN NAVARRA; INTERPRETATION OF A HIGH DETECTION RATE AT THE FIRST SCREENING ROUND AND A LOW RATE AT THE SECOND ROUND

## SUMMARY

*Objectives* - Evaluating the on-going European pilot project for breast cancer screening in Navarra, Spain, and predicting the effects and costs of the programme in the long run.

*Methods* - Observed results in Navarra, consisting of more than 100,000 screens, were compared with expected results. A MISCAN model was used, that included demographical, epidemiological and screening characteristics of Navarra. Alternative assumptions on epidemiological and screening characteristics were also addressed.

*Results* - The observed detection rate (5.9 per 1,000 screened women) in the first round was 18% higher than expected; the observed rate in the subsequent round (2.9) is 17% lower than expected. Longer pre-clinical durations, lower sensitivity or the existence of a high-risk group in Navarra could not satisfactorily explain the first and second round results together. Nevertheless, the programme will have an important health benefit for the women involved, due to an important trend in incidence in recent years and the relatively unfavourable clinical stage distribution in Navarra. The proportion T2+ cancers that will be prevented after 10 years of screening amounts to 36%. The annual mortality reduction in steady state is expected to range between 17% (if the observed rates in the second round indicate real screening performance) to 23% (if the first round indicates real performance).

*Conclusions* - Our results demonstrate that a high detection rate in the first round is insufficient to evaluate the quality of a programme. Interval cancer rates, results of the subsequent round and size distributions are also crucial indicators of the quality of the screening programme and should be analysed in their specific context.

# INTRODUCTION

Nationwide breast cancer screening programmes have been initiated in countries with earlier experience with breast cancer screening, such as Sweden, the United Kingdom and The Netherlands (*1,2*). In several other countries of the European Union without such long-standing experience, pilot projects started in the beginning of the 1990s (*3*). Breast cancer screening can substantially reduce breast cancer mortality rates for women aged 50 and older (*4,5*), but it is not evident for all European countries to have a breast screening programme. In addition to screening characteristics, the cost-effectiveness of a screening programme depends on demographical and epidemiological characteristics of the country concerned. Breast cancer screening is usually assumed as being cost-effective in countries with relatively high breast cancer incidence and mortality rates and if mammographic quality and substantial participation rates are guaranteed.

In Navarra, a region in the north of Spain, an Early Detection Programme started in 1990. This programme is directed at women aged 45-65 years who are invited to participate every 2 years. Striking results of the first round, in which almost 50,000 women were screened, are the high participation rate (85%) and high detection rates (nearly 6 per 1,000 women screened in the age group 45-65) (*6*). The detection rate at the first screening in Navarra is not much different from observed rates of 6 and 6.5 in Sweden (*7*) and The Netherlands (*8*), respectively, despite a much lower background incidence. Is this a promising result?

Models taking into account specific characteristics of the region concerned have been shown to be a valuable tool in the evaluation of a screening programme for breast cancer (*4,9,10*). Although Navarra is a circumscript region, it is the first European Union project to have detailed data of 4 years of screening, consisting of more than 100,000 screens, and background information on the region. This enabled us now to evaluate the pilot project in Navarra and to derive predictions on the effects and costs of the programme in the long run, using a validated model on breast cancer screening as reference (*1,4,11*).

# MATERIAL AND METHODS

## The Early Detection Programme in Navarra

The Early Detection Programme directed at women aged 45-65 years started in March 1990 with a 4-month pilot phase to adapt the units, procedures, etc., and normal operation began on 1 September 1990. The first round of screening ended on 1 July 1992. The second round of screening took place in the period 1 September 1992 to 1 July 1994. The average interval between the screen in the first round and the one in the second round was 2.06 years. In the second round, women aged 66 or 67 years were also invited if they had attended the first round, to give them the opportunity to attend at least twice.

## The MIcrosimulation SCreening ANalysis (MISCAN) approach

To evaluate the results of the current screening programme, and to derive predictions on the possible effects and costs of a long-term screening programme in Navarra, the simulation model MISCAN was used. A detailed description is given elsewhere (*4,12*). The MISCAN-model consists of 2 steps. The first step (the disease part of the program) reflects the natural course and epidemiology of the disease. In the model, several states of the disease are distinguished, including 4 invasive states (T1a, T1b, T1c and T2+) with increasing tumour size (respectively ≤5, >5 and ≤10, >10 and ≤20 and >20 mm, respectively). A proportion of the invasive breast cancers is assumed to be preceded by a screen-detectable, ductal carcinoma *in situ* (dCIS), state. It is also possible that regression might occur in this state. By incorporating demographical and epidemiological aspects of breast cancer, individual life histories are generated to constitute the population of Navarra. In this population the number of women, the age distribution of women, the incidence of clinically diagnosed breast cancer, the stage distribution of clinically diagnosed cancers, and the mortality due to breast cancer and to other causes are simulated as observed in Navarra.

In the second step (the screening part of the programme), various characteristics of screening policies can be defined, such as screening ages, interval and attendance (*12*). Furthermore, the screening performance - determined by the sensitivity and specificity of the screening-test, and improvement in prognosis after screen-detection are defined in this part. In the model, women with cancer have a reduced risk of dying of breast cancer due to screen detection, depending on the cancer size at detection. This improvement in prognosis after detection via screening was based on the results of analyses of Swedish randomized trials (*4,5*) and is calculated as 1 minus the ratio of the risk of dying of screen-detected breast cancer divided by the risk when the cancer had been diagnosed in the absence of screening. An equal improvement in prognosis after screen-detection is assumed for women younger than 50 years of age and those 50 years of age and older.

The results of the programme consist of the number of cases detected, the stage distribution of the screen-detected cases and the number of interval cases. The effects of screening, such as the number of life-years saved and the mortality reduction, are also results of the screening part of the programme.

## Demography, epidemiology of breast cancer, screening policy and performance of Navarra implemented in MISCAN

The 1989 population numbers in 5-year age-groups of Navarra (data not shown) were used. The death rates in 5-year age categories due to causes other than breast cancer in 1987-1989 were calculated from the number of deaths of all causes (data not shown), the number of deaths due to breast cancer (data not shown) and the population number during this period.

An estimate for the duration of the pre-clinical stages can be obtained from the clinical stage distribution in a period without screening. Breast cancer is a process in which one moves from relatively favourable stages to increasingly worse stages of the disease; therefore, clinical stage

*Table 2.1*

*Population numbers of Navarra, incidence and mortality of breast cancer in Navarra. Stage distribution of clinically diagnosed cancers and axillary lymph node status in Guipuzcoa.*

| Characteristic | Value |
|---|---|
| Female population (1989) | |
| Total | 262,731 |
| 45-49 | 13,998 |
| 50-54 | 13,593 |
| 55-59 | 14,729 |
| 60-64 | 14,271 |
| 65-69 | 12,828 |
| Breast cancer incidence (crude rate) | |
| Navarra (1987-1989) | 75.3 |
| Mortality due to breast cancer (crude rate) | |
| Navarra (1987-1989) | 28.4 |
| Mortality/Incidence ratio | |
| Navarra (1987-1989) | 0.38 |
| Stage distribution | |
| Clinically diagnosed breast cancers(%) | |
| Guipuzcoa | |
| DCIS | 5.7% |
| T1 | 21.1% |
| T2 | 44.6% |
| T3 | 19.6% |
| T4 | 8.9% |
| Axillary lymph node metastases | |
| (% of invasive tumours) | |
| Guipuzcoa | 50% |

distribution implies the duration of the disease process. Because there is no complete and precise information available of the stage distribution in the period before screening in Navarra, the stage distribution of Guipuzcoa (data not shown) was used. Guipuzcoa is a region neighbor to Navarra and is assumed to have a comparable stage distribution, although a slightly higher level of opportunistic screening may be present in Guipuzcoa. In Guipuzcoa, only 27% of the clinically diagnosed cancers have a tumor size ≤20 mm or are non-invasive, which is low compared with the percentages found around Florence (39%, (*13*)), in The Netherlands this percentage is 46% (*1*), and in Germany 38% (*11*). The clinical stage distribution is assumed to result from a hazard of growing to a next state of the disease and a hazard for clinical detection in a certain state. Therefore, assuming that the natural course of the disease is the same for different countries, the hazard of growing to a next state is equal for the different situations. The only factor that can influence the clinical stage distribution is thus the hazard for clinical detection. We adjusted the hazard for clinical detection in the different states downward to obtain the less favourable stage distribution. This resulted in estimates of the mean pre-clinical duration ranging from 3.2 years for women aged 45 at onset of the disease to more than 5.5 years for women aged 65 at onset of the disease.

***Table 2.2***

*Characteristics of the Navarrian screening programme used in the model for different age groups*

| Screening characteristic | Age 45-49 | Age 50-65 |
|---|---|---|
| Attendance rate (% of invited women) | | |
| 1st invitation | 85.5 | 84.4 |
| subsequent invitation | 85.9 | 86.7 |
| not attended previous round | 78.2 | 3.3 |
| attended previous round | 96.2 | 95.4 |
| | | |
| Hospital referral rate[a] | | |
| (% of screened women) | | |
| 1st screen | 1.1 | 1.0 |
| subsequent screen | 0.5 | 0.4 |
| | | |
| Biopsies (% of hospital referrals) | | |
| 1st screen | 87.7 | 93.4 |
| subsequent screen | 84.4 | 84.2 |
| | | |
| Positive predictive value advice for biopsy | | |
| 1st screen | 43.9 | 63.7 |
| subsequent screen | 70.4 | 78.5 |
| | | |
| Sensitivity[b]; principal analysis | | |
| DCIS | 0.32 | 0.40 |
| T1a | 0.52 | 0.65 |
| T1b | 0.64 | 0.80 |
| T1c | 0.72 | 0.90 |
| T2+ | 0.76 | 0.95 |
| | | |
| Reduced risk of dying from breast cancer[c] | | |
| screen-detected dCIS | 1.000[d] | 1.000 |
| screen-detected T1a | 0.310 | 0.892 |
| screen-detected T1b | 0.230 | 0.814 |
| screen-detected T1c | 0.070 | 0.567 |
| screen-detected T2+ | 0.050 | 0.395 |

[a] A woman with a positive screening test first underwent additional diagnostic procedures before she was referred to hospital.

[b] Based on the results of the Dutch pilot projects in Utrecht and Nijmegen.

[c] Defined as the proportion of women with cancer that have risk of dying from breast cancer reduced. This proportion depends on cancer size at detection and is based on the results of analysis of the Swedish randomized trials.

[d] If improvement in prognosis is assumed to be lower for women younger than 50 years compared to women aged 50 years and older

A 3% annual increase in breast cancer incidence was incorporated, based on incidence data from 1973-1989 ((*14-16*), data not shown). This was done by increasing the risk of contracting breast cancer during a lifetime for younger birth cohorts by 3% per birth year up to a maximum risk level; for women born after 1944, the risk of contracting breast cancer remains constant (according to López-Abante et al. (*17*)) at a level that is comparable with countries in northwestern Europe (*18*). This observed fast increase results in a high incidence level in Navarra compared with other parts of Spain (*14-16*).

There were no data available on the survival of breast cancer patients. Therefore, the stage specific survival is fitted on the breast cancer incidence and mortality rates of Navarra. In Table 2.1, the demography, incidence, mortality and clinical stage distribution are summarised.

The screening policy of the Early Detection Programme in Navarra, inviting women aged 45-65 years every 2 years, was implemented in MISCAN. The attendance rates and specificity were

based on the results of the first and second round of the programme. In first instance, estimates of the sensitivity were used, which were based on results of Dutch pilot projects and nationwide screening programme (*1,12*). Although this is not supposed to be a gold standard, it served as the first estimate in the analyses. A survey of the characteristics of the screening programme is given in Table 2.2.

### Evaluation of the current screening programme

After implementing the Navarrian characteristics in MISCAN, the screening results expected by MISCAN (detection rates and stage distribution of screen-detected cancers for the first and second rounds and the number of interval cancers) were compared with the results observed in Navarra. To assess the cost-effectiveness of the current screening policy, the predicted costs and effects in the total female population of a 27-year screening programme directed at women aged 45-65 were calculated. The costs and cost-effectiveness estimates are presented using a discounting rate of 5% per year, to establish a time-preference.

The cost calculations were primarily based on the detailed cost-analyses performed in The Netherlands (*19*), using the health-care-specific Purchasing Power Parity (with 1 dutch florin = approx. 47.90 pesetas), which is a conversion factor adjusted to health-care-specific price differences between Spain and The Netherlands (*20-23*). Some detailed costs for the screening itself were available from Spain (data not shown); these were quite consistent with the Dutch data on screening costs. The reported overhead costs of screening in Navarra are lower than in The Netherlands, this is caused partly by the larger number of screen units needed in a larger programme. Only rough estimates of cost of diagnosis, primary treatment and follow-up are available (data not shown).

## RESULTS

### Observed results of the first and second screening rounds compared to expectations

The number of breast cancer cases detected in the first round was 5.9 per 1,000 screened women. This number includes the cases found in the 9% of screened women who had a review within 1 year. Given the Navarrian characteristics, we would have expect a detection rate in the first round of 5.0 (Table 2.3). The difference between the observed and expected detection rates was also seen at the second round, in which the observed detection rate of women screened for the first time was 3.8, while the expected detection rate was 3.0. The latter are both women who were not invited in the first round and women who were invited in the first round, but did not attend. For women who had also attended the first round (the majority), a detection rate of 2.9 was observed at the second round; the expected detection rate was 3.5.

***Table 2.3***

*Screening results as observed in Navarra compared to simulated results from MISCAN*

| Screening results | Screened women | observed | expected |
|---|---|---|---|
| Detection rates (per 1,000 women screened) | | | |
| First round | | | |
| 45-49 | 11,238 | 4.8 | 2.4 |
| 50-54 | 10,365 | 5.2 | 4.2 |
| 55-59 | 12,097 | 5.9 | 5.8 |
| 60-64 | 11,550 | 7.2 | 7.6 |
| 65 | 3,151 | 7.6 | 8.3 |
| 45-65 | 48,401 | 5.9 | 5.0 |
| Second round (attended 1st round) | | | |
| 45-49 | 6,450 | 2.9 | 2.1 |
| 50-54 | 10,523 | 1.8 | 3.1 |
| 55-59 | 10,919 | 2.4 | 3.8 |
| 60-64 | 11,428 | 2.7 | 4.3 |
| 65-67 | 7,102 | 5.6 | 4.6 |
| 45-67 | 46,422 | 2.9 | 3.5 |
| Second round (not attended 1st round)° | | | |
| 45-49 | 6,397 | 2.8 | 2.1 |
| 50-54 | 627 | 6.4 | 4.7 |
| 55-59 | 562 | 7.1 | 6.3 |
| 60-64 | 554 | 5.4 | 6.9 |
| 65 | 183 | 16.4 | 9.1 |
| | 8,863 | 3.8 | 3.0 |
| Stage distribution first round (n=285) | | | |
| DCIS | | 17% | 12% |
| Invasive | | 83% | 88% |
| T1a | | 6% | 3% |
| T1b | | 27% | 20% |
| T1c | | 40% | 48% |
| T2+ | | 28% | 29% |
| Interval cancers (per 1,000 women screened) | | 0.7 | 1.4 |

° This concerns both women invited in the first round but not attending and women not invited in the first round

The observed stage distribution of screen-detected cancers in the first round was slightly more favorable than expected but worse in the second round. In Navarra, 36 interval cancers have been reported between the 2 screening rounds, 10 in the first and 26 in the second year after screening. The expected number of 69 interval cases in the first 2 years is almost 2 times higher.

*Table 2.4*

*Expected screening results for alternative analyses I-III, compared to the observed screening results in Navarra*

| Screening results | Navarra | alternative I[a] | alternative II[b] |
|---|---|---|---|
| Detection rates (per 1000 women screened) | | | |
| - first round | 5.9 | 5.9 | 3.8 |
| - second round (attended 1st round) | 2.9 | 3.7 | 3.0 |
| - second round (not att. 1st round)[c] | 3.8 | 3.3 | 2.3 |
| - second round (all) | 3.1 | 3.7 | 3.0 |
| Interval cancers (per 1000 screens) | 0.7 | 1.3 | 1.9 |

[a] Preclinical duration of the principal analysis+18%
[b] Sensitivity of the principal analysis -24%
[c] This concerns both women invited in the first round but not attending and women not invited in the first round

## Alternative assumptions for Navarra

Because the prevalence is primarily indicative for the pre-clinical duration of the tumours found at screening, we varied the pre-clinical duration to obtain the same observed and expected detection rate at the first round. When the screen-detectable pre-clinical period is extended by 18%, the observed and expected detection rate are both 5.9, but the detection rate at the second round would be expected to increase also, to 3.7 (Table 2.4, alternative I). The observed rate would then suggest a change in performance after the first round. Otherwise, if the prevalent figure is strongly inflated by cancers with borderline malignancy and length time sampling and the second round indicates real screening performance, the sensitivity has to be lowered by 24% to obtain comparable observed and expected detection rates at the second round (Table 2.4, alternative II). The possibility of a lower sensitivity in Navarra can be explained by differences in the performance of screening in Navarra (one-view mammography, only limited double reading) and other trials (double-view at first examination, double reading).

## International comparisons

In Table 2.5, the observed and expected results from Navarra are compared with the screening results of other programmes, including the Nationwide Screening Programme in The Netherlands (8), the Swedish Two County trial (7) and the Florence District Programme (24). The difference between the observed and expected detection rates in the prevalent round in Navarra is caused by the high observed detection rates in the youngest age-groups (45-49, 50-54). Compared with other programmes, the observed detection rates in the youngest age groups are also high, particularly when compared with the background incidence.

***Table 2.5***

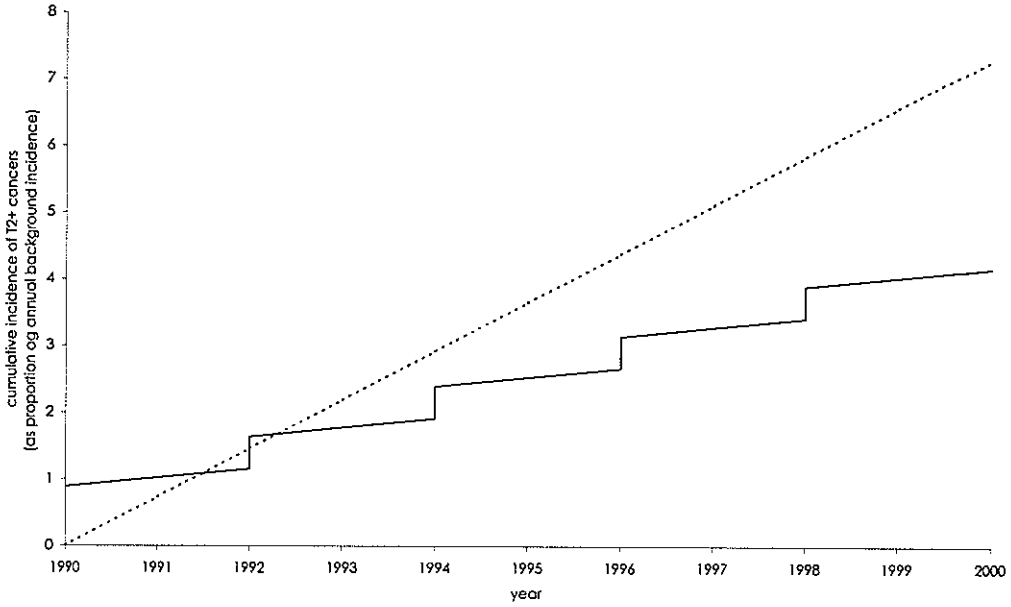
*Comparisons of the observed and expected screening results of Navarra to the results of other programs: the Nationwide Screening Programme in the Netherlands (8), the Swedish Two County trial (7) and the Florence District Programme (24).*

| Programme<br>*Age-group*<br>*Screening interval (S)*<br>*Number of views* | Age group | Detection rates 1st round (P) (per 1,000 screens) | Detection rates subsequent round/screen(s) (D) (per 1,000 screens) | Background incidence (I) (per 1,000 women) | P/I | D/I | Interval cancers as % of background incidence (C) | | | (D/I+C) /S |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 1st yr | 2nd yr | 3rd yr | |
| Navarra observed | 45-49 | 4.8 | 2.9 | 1.5[b] | 3.2 | 1.9 | | | | |
| *45-65* | 50-54 | 5.2 | 1.6 | 1.4 | 3.7 | 1.1 | 15 | 36 | - | 1.1 |
| *2.1 yrs* | 55-59 | 5.9 | 2.4 | 1.6 | 3.7 | 1.5 | | | | |
| *single* | 60-64 | 7.2 | 2.7 | 1.7 | 4.2 | 1.6 | | | | |
| Navarra expected | 45-49 | 2.4 | 2.1 | 1.3[c] | 1.8 | 1.6 | | | | |
| | 50-54 | 4.2 | 3.1 | 1.5 | 2.8 | 2.1 | 33 | 60 | - | 1.6 |
| | 55-59 | 5.8 | 3.8 | 1.7 | 3.4 | 2.2 | | | | |
| | 60-64 | 7.6 | 4.3 | 1.8 | 4.2 | 2.4 | | | | |
| NSPN 1994 | 50-54 | 4.5 | 3.0 | 2.1[d] | 2.1 | 1.4 | | | | |
| *50-69* | 55-59 | 5.8 | 3.8 | 2.3 | 2.5 | 1.7 | | | | |
| *2.0 yrs* | 60-64 | 7.3 | 3.6 | 2.5 | 2.9 | 1.4 | | | | |
| *two (1st ex)/single(subs ex)* | 65-69 | 10.0 | 4.8 | 3.0 | 3.3 | 1.6 | | | | |
| STC | 40-49[a] | 2.1 | 2.7 | 1.1[e] | 2.0 | 2.5 | 46 | 53 | - | 1.7 |
| *40-74* | 50-59 | 4.7 | 3.0 | 1.9 | 2.5 | 1.6 | 10 | 28 | 52 | 0.9 |
| *24-33 mnths* | 60-69 | 8.8 | 4.9 | 2.5 | 3.5 | 2.0 | 17 | 27 | 57 | 1.0 |
| *single* | 70-74 | 12.2 | 7.5 | 3.0 | 4.1 | 2.5 | 8 | 44 | 48 | 1.2 |
| FDP | 40-44 | 0.8 | 3.0 | 1.1[f] | 0.8 | 2.8 | 24 | 50 | 98 | 1.0 |
| *40-69* | 45-49 | 2.4 | 1.3 | 1.5 | 1.7 | 0.8 | | | | |
| *30 mnths* | 50-54 | 3.4 | 2.9 | 1.3 | 2.5 | 2.2 | 17 | 45 | 63 | 1.2 |
| *two* | 55-59 | 6.0 | 2.2 | 1.6 | 3.8 | 1.4 | | | | |
| | 60-64 | 5.8 | 3.8 | 1.8 | 3.3 | 2.0 | 9 | 17 | 39 | 1.1 |
| | 65-69 | 11.8 | 4.7 | 1.9 | 6.3 | 2.5 | | | | |

[a] ages at randomization
[b] clinical incidence observed in Navarra 1987-89
[c] expected clinical incidence 1987-89
[d] clinical incidence 1989 in the Netherlands corrected for undemotification at older ages
[e] control incidence
[f] expected incidence

*Figure 2.1*

*Detection rates at first and second round (for all women participating in the second round, irrespective of their participation in the previous round) as observed in Navarra and expected by MISCAN*




In the 2 oldest age-groups, the observed detection rates are lower than expected because there is a less steep increase in detection rates by age than expected. This steep increase at the prevalent screen is due to the assumption of a strongly age-dependent preclinical duration of breast cancer and to the increase of incidence by age. In other screening programmes, the increase by age of incidence and preclinical duration (the latter roughly indicated by the increase of prevalence/incidence ratio) was also observed. In contrast to the prevalent round, the observed detection rates in the subsequent round and the number of interval cancers are rather low compared with the other studies. The number of screen-detected cancers in subsequent rounds and the number of interval cancers depend on the background incidence, the screening interval and the performance of screening. Comparisons between studies having a different background incidence and screening interval can therefore only be made after dividing the sum of the screen-detected cancers in subsequent rounds and interval cancers, *i.e.* D/I + C, by the screening interval (S) (Table 2.5). The resulting ratio [(D/I+C)/S] observed in Navarra is comparable with other programmes, whereas a higher ratio is expected by MISCAN.

## Predicted effectiveness and cost-effectiveness of the current screening programme

An indicator for evaluating the effectiveness of a screening programme is the extent to which the occurrence of advanced cancers (stage T2+) is prevented. At the first round 1.4 per 1,000 women screened were found to have a cancer in stage T2+; at the second round this was 0.7 per 1,000

*Table 2.6*

*Effects on incidence, mortality and life-years gained and CE-ratio (5% discounting) of the Early Detection Programme in Navarra, starting in 1990 and directed at women aged 45-65 as predicted by three analyses assuming an equal improvement in prognosis for women below 50 years and women aged 50 years and older. For the prinicipal analysis also the effects are presented if a lower improvement in prognosis for women below 50 compared to women aged 50 and older is assumed. The effects of 27 years of screening are considered.*

| | Principal analysis | Alternative I[c] | Alternative II[b] |
|---|---|---|---|
| Breast cancers diagnosed | 9,780 | 9,850 | 9,610 |
| Screen-detected | 3,390 (35%) | 3,770 (38%) | 2,850 (30%) |
| Breast cancer deaths prevented | 1,100 | 1,200 | 910 |
| Life-years gained | 22,000 | 23,000 | 18,300 |
| Breast cancer mortality reduction in steady state (2015) | 20% | 23% | 17% |
| CE-ratio (pesetas per life-year saved) | 441,000 | 427,000 | 555,100 |

[a] Preclinical duration of the principal analysis +18%
[b] Sensitivity of the principal analysis -24%
[c] Lower improvement in prognosis for women under 50 years of age (see Table 2.2)

women screened. Between these screening rounds yearly 0.2 per 1,000 women were found to have a T2+ cancer. In the absence of screening, the incidence of T2+ would have been 1.1 per 1,000 women years. In Figure 2.1 the cumulative incidence of T2+ cancers as a proportion of the background incidence for women attending all screening rounds is compared with the situation without screening, with the assumption that the proportion T2+ cancers found in the second screening round and among the interval cancers in Navarra is representative for future screening rounds. In the first years of screening, the cumulative incidence of cancers with stage T2+ is higher for women attending the screenings compared with the situation without screening; however, after 10 years, the proportion of T2+ cancers prevented by screening was 36%.

When the screening programme introduced in Navarra in 1990 is extended to 2016, an annual reduction in the breast cancer mortality of 20% is predicted by MISCAN for the principal analysis from the year 2014 onwards. The 27 years of screening are expected to prevent 1,100 breast cancer deaths, corresponding to 22,000 life-years saved. Due to screening, the yearly number of newly diagnosed breast cancers is expected to increase (25) by 4% compared with the situation without screening. The screening programme is expected to detect 35% of all breast cancers diagnosed; 86% of the screen-detected cancers were ≤20 mm in diameter or non-invasive. The cost-effectiveness ratio of the present Navarra programme is estimated to be 441,000 pesetas per life-year gained; when Navarrian costs are used, the ratio is 319,000 pesetas per life-year gained.

The alternative analyses have different consequences for the (cost-)effectiveness of the programme. The predicted maximal annual reduction in breast cancer mortality attainable from the year 2015 onward ranges from 17%, if the observed detection rate in the second round indicates real screening performance (and the prevalent figure is strongly inflated by cancers with borderline malignancy and length time sampling), to 23% if the prevalent round detection rates

indicate the real situation and the currently reported number of interval cancers (and second round results) are too low. The cost-effectiveness ratio is assessed between 427,000 (alternative I) and 555,100 (alternative II) pesetas per life-year gained (Table 2.6).

## DISCUSSION

The high detection rate in the first and the low detection rate in the second round in Navarra were unexpected. In other analyses performed in The Netherlands, Germany, Italy and evaluations of the Nationwide Screening Programme in the Netherlands (NSPN *[11-13,26]*, the model reproduces the detection rates in the first and subsequent screening rounds and the incidence of interval cancers as observed, with the natural course and characteristics of the screening test appearing to be incorporated satisfactorily.

A comparison of the observed and expected results of Navarra with international results, taking into account the background incidence, demonstrates high detection rates in the youngest age groups and a less steep increase of the detection rates in the prevalent round by age than expected on the basis of the increase by age of incidence and duration of pre-clinical stage, which was also seen in other programmes. However, relating the sum of screen-detected cancers in the subsequent round and the interval cancers to the background incidence and screening interval (Table 2.5, ratio (D/I+C)/S) yields ratios comparable with other programmes, while a higher ratio is expected by MISCAN, which is partly caused by the increasing background incidence as observed in Navarra and the long pre-clinical duration. In Table 2.5, the background incidence of the years 1987-89 is used for the calculation of both the observed and expected ratio for Navarra. Due to the increasing incidence the background incidence will be higher at the moment the interval cancers arise and the second screening round takes place. Because of this, the sum of the screen-detected cancers of the first round and the interval cancers after the first round divided by the product of incidence and screening interval as calculated in Table 2.5 will be overestimated. Furthermore, the long pre-clinical duration in Navarra will cause a high prevalence at the first screening; therefore, the absolute number of cases prevalent but not screen detected due to a lack of sensitivity will be higher than in case of a shorter pre-clinical duration. These cases will become interval cancers or will be screen detected at the subsequent screening, resulting in a higher ratio (D/I+C)/S. Although the possibility remains that the expected detection rates of the subsequent round are slightly overestimated, the observed number of interval cancers and detection rates of the subsequent round are rather low considering the increasing incidence and long pre-clinical durations.

Trying to explain the observed results in Navarra, we changed model assumptions on the natural course and the characteristics of the screening test. Changes made were assuming a longer pre-clinical duration and a lower sensitivity than in the first instance. These alternative assumptions, however, cannot explain the relative high detection rates in the prevalent round together with the rather low detection rates in the subsequent round. Furthermore we have investigated the possibility of the existence of a high risk group in Navarra. A portion of the women were asked to

have a review within 1 year after the screening test because of radiological reasons. Due to the high detection rate found in the 9% of the women who were reviewed after the first round, we assumed them to comprise a higher risk group having a 3 times higher risk of having breast cancer during lifetime than other women. This assumption cannot explain the results observed in Navarra, again mainly due to the low detection rates observed in the second round.

For further explanations of the differences with screening results of other programmes discussed above, and the impossibility of reproducing satisfactorily the observed detection rates with the present assumptions on the natural course of breast cancer and screening characteristics, future studies on the following aspects should be undertaken. On the one hand, review of the breast cancers detected in the first round, including pathology, to possibly extract the cancers with borderline malignancy which are often more prevalent at early screens. Review of possible differences in the natural course of breast cancer in this (Southern) region compared to the other European ones is certainly also an aspect to be analysed.

The estimated reduction of T2+ cancers prevented by screening after 10 years of 36% is comparable to the estimates for the Swedish Two County Trial (STC) and Nationwide Screening Programme in the Netherlands (NSPN) (40% and 36%, respectively (8)). This rather favorable figure for Navarra is illustrated by the expected mortality reduction upto 2020 of 17% by MISCAN and the predicted maximum annual mortality reduction of 20 %, which are higher than predictions for The Netherlands, Germany and Florence (11,13,19). The predicted cost-effectiveness ratio of the Early Detection Programme, which ranges from 427,000 to 555,100 pesetas per life-year gained (5% discount), is also rather favorable. Such high estimated effectiveness and favorable cost-effectiveness ratio of the Early Detection Programme are due to the specific characteristics of Navarra, the increasing incidence and relatively unfavorable clinical stage distribution associated with a longer pre-clinical duration.

In conclusion, our results emphasize that the quality of a programme cannot be evaluated on results of a prevalent screening round only. Detection rates of subsequent rounds, the interval cancer rates and tumor size distributions, related to underlying epidemiology and natural course of disease in a specific situation, are also crucial indicators of the quality of a screening programme and of its effects for the future. In modelling analysis, we have tested different assumptions on epidemiological and screening characteristics of the specific Navarrian situation to explain the observed data.

## REFERENCES


(1) de Koning HJ, Fracheboud J, Boer R, Verbeek AL, Collette HJ, Hendriks JH, et al. Nation-wide breast cancer screening in The Netherlands: support for breast-cancer mortality reduction. National Evaluation Team for Breast Cancer Screening (NETB). Int J Cancer 1995;60:777-80.

(2) Moss SM, Michel M, Patnick J, Johns L, Blanks R, Chamberlain J. Results from the NHS breast screening programme 1990-1993. J Med Screening 1995;2:191-4.

(3) de Waard F, Kirckpatrick A, Perry NM, Tornberg S, Tubiana M, de Wolf C. Breast cancer screening in the framework of the Europe against Cancer Programme. Eur J Cancer Prev 1994;3:3-5.

(4) de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantative Interpretation of Age-Specific Mortality Reductions from the Swedish Breast Cancer-Screening Trials. J Natl Cancer Inst 1995;87:1217-23.
(5) Nystrom L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Ryden S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. Lancet 1993;341:973-8.
(6) Ascunce N, del Moral A, Murillo A, Alfaro C, Apesteguia L, Ros J, et al. Early detection programme for breast cancer in Navarra, Spain. Eur J Cancer Prev 1994;1:41-8.
(7) Tabar L, Fagerberg G, Duffy SW, Day NE, Gad A, Grontoft O. Update of the Swedish Two-County program of mammographic screening for breast cancer. Radiol Clin North Am 1992;30:187-210.
(8) National Evaluation Team for Breast Cancer Screening (NETB). National evaluation of screening for breast cancer in The Netherlands V (In Dutch). Rotterdam: Dept. of Public Health, Erasmus University, Rotterdam; 1996.
(9) Chen HH, Tabar L, Fagerberg G, Duffy SW. Effect of breast cancer screening after age 65. J Med Screen 1995;2:10-4.
(10) Tabar L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A, et al. Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. Cancer 1995;75:2507-17.
(11) Beemsterboer PM, de Koning HJ, Warmerdam PG, Boer R, Swart E, Dierks ML, et al. Prediction of the effects and costs of breast-cancer screening in Germany. Int J Cancer 1994;58:623-8.
(12) van Oortmarssen GJ, Habbema JD, van der Maas PJ, de Koning HJ, Collette HJ, Verbeek AL, et al. A model for breast cancer screening. Cancer 1990;66:1601-12.
(13) Paci E, Boer R, Zappa M, de Koning HJ, van Oortmarssen GJ, Crocetti E, et al. A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy. Eur J Cancer 1995;3:348-53.
(14) Muir C, Waterhouse J, Mack T, Powell J, Whelan S. Cancer Incidence in Five Continents, Volume V. Lyon: IARC Scientific Publication; 1987.
(15) Parkin D, Muir C, Whelan S, Gao Y-T, Ferlay J, Powell J. Cancer Incidence in Five Continents, Volume VI. Lyon: IARC Scientific Publication; 1992.
(16) Waterhouse J, Muir C, Shanmugaratnam K, Powell J. Cancer Incidence in Five Continents, Volume IV. Lyon: IARC Scientific Publications; 1982.
(17) López-Abante G, Pollán M, Ruiz Tovar M, Jiménez M, Vázquez F. Cancer Mortality in Spain, 1952-1986. Effect of age, birth cohort and period of death. Madrid: Centro Nacional de Epidemiologia, Instituto de Salud Carlos III, Ministerio de Sanidad y Consumo; 1992.
(18) Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H. Trends in cancer incidence and mortality. Iarc Sci Publ 1993;121:1-806.
(19) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JC, Collette HJ, Hendriks JH, 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.
(20) Mason J, Drummond M. The DH register of cost-effectiveness studies: A review of study content and quality. University of York, Centre for Health Economics; 1995.
(21) OECD (Organisation for Economic Co-operation and Development). Purchasing Power Parities and Real Expenditures in the OECD. Paris; 1992.
(22) OECD. The socio-economic environment & statistical references. Vol II. Paris; 1993.
(23) OECD. OECD Health Data, Electronic Version 1.5. Paris: Credes; 1993.
(24) Paci E, Ciatto S, Buiatti E, Cecchini S, Palli D, Rosselli del Turco M. Early indicators of efficacy of breast cancer screening programmes. Results of the Florence District Programme. Int J Cancer 1990;46:198-202.
(25) Boer R, Warmerdam P, de Koning H. Extra incidence caused by mammographic screening[letter]. Lancet 1994;343:979.
(26) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, Boer R, Collette HJA, Verbeek ALM, et al. The costs and effects of mass screening for breast cancer. Final report (In Dutch). Department of Public Health, Erasmus University, Rotterdam; 1990.

# 3

# REDUCTION IN BREAST CANCER MORTALITY DUE TO THE INTRODUCTION OF MASS SCREENING IN THE NETHERLANDS: COMPARISON WITH THE UNITED KINGDOM

## SUMMARY

*Objective* - To assess the impact of the national breast cancer screening programme on breast cancer mortality in the first years after its introduction.

*Setting* - The Netherlands and United Kingdom.

*Methods* - MISCAN models, incorporating demographic, epidemiological and screening characteristics of the region under study, were used to assess the mortality in the presence and absence of screening.

*Results* – Breast cancer mortality decreased in women aged 55-74 as the Dutch nationwide screening programme built up, and was 5% lower in 1996 than before the start of the programme. The mortality reduction due to screening in the age-group 55-74 is expected to increase gradually to 18% in 1999, 10 years after the introduction of screening, and to 29% in the long term. In the United Kingdom screening was expected to achieve a mortality reduction of 5% and 18% in the age-group 55-69 five and 10 years respectively after screening was started. A maximum mortality reduction of 24% in this age group is predicted.

*Conclusions* – The effects of screening will be small in the first years after the start of the programme. Accordingly, it was expected that the reduction in breast cancer mortality due to the Dutch nationwide breast screening programme, which started around 1989, would be statistically significant from 1997 onwards, the point at which the target population of women was completely covered; 70% of the reported 12% mortality reduction in England and Wales in 1994 is expected to be attributed to screening.

# INTRODUCTION

Nationwide breast screening programmes have been introduced since the late 1980s. Although their goal is reduction in breast cancer mortality, relatively few reports have yet been published on such reductions (*1-4*). Changes in breast cancer mortality related to early detection and treatment are expected to appear only several years after the introduction of a screening programme. In England and Wales, however, a decrease of 12% in the age-standardised mortality in the age group 55-69 was already observed within seven years after the introduction of the NHS breast screening programme. This fall in mortality was concluded to be only partially as a result of screening (*2*). In Finland (*4*) a reduction in mortality of 24% was found in women aged 50-59 in the period 1987-92. However, this result was based on small numbers and was not statistically significant.

In the Netherlands the nationwide screening programme carried out at two-yearly intervals started around 1989. By the end of 1997 the target population of women was completely covered: all women aged 50-69 had been invited at least once. Before its introduction we predicted that the programme would yield a maximum annual mortality reduction of 17% in the total female population from the year 2015 onwards (*5*).

In the present study we analysed the breast cancer mortality in the first years after introduction of a screening programme in the Netherlands and United Kingdom, and investigated whether there was a decrease that might be attributed to the increased use of screening mammography. We used a validated model on breast cancer screening (*5,6*). The expected figures for the Dutch situation were compared to the observed breast cancer mortality rates in the Netherlands in the period 1986-1996.

# MATERIAL AND METHODS

## The Microsimulation SCreening ANalysis (MISCAN) approach

In MISCAN a population is simulated according to the demographic characteristics of the population under study (births, life tables, deaths from other causes than breast cancer). The incidence and mortality of breast cancer of this population are reproduced. The natural history of the breast cancer is modelled as a progression through successive disease states. Four invasive stages are distinguished according to tumour diameter - T1a, T1b, T1c and T2+ (≤ 5, > 5-10, >10-20, and > 20 mm)- and a proportion of the lesions is assumed to be preceded by a screen detectable ductal carcinoma in situ (dCIS). Without screening a screen detectable preclinical cancer may be diagnosed clinically or progress to the next preclinical invasive state. The time spent in each preclinical state, which is assumed to be exponentially distributed, and the rate at which preclinical cancers are clinically diagnosed is inferred from data on the clinical incidence and stage distribution before screening started, and on the detection rates and interval cancer rates by stage

in the population concerned. The stage-specific survival after clinical diagnosis without screening is based on data from survival registries.

The effects of a screening programme are calculated by comparing the clinical incidence and mortality with and without screening. To simulate the clinical incidence and the mortality without screening demographic and epidemiological characteristics of the population under study are implemented in the model. After adding the screening characteristics (screening ages, interval and attendance) and the performance of the screening programme (determined by the sensitivity and specificity of the screening test and improvement in prognosis after screen detection) of the population under study to the model, the clinical incidence and mortality when screening is carried out are calculated. The sensitivity in the model is based on the results of the Dutch nationwide screening programme (*7*), and for women aged over 50 it was fixed as 0.4, 0.65, 0.8, 0.9, and 0.95 for dCIS, T1a, T1b, T1c and T2+ tumours respectively. Women with screen detected cancers are assumed to have a reduced risk of dying from breast cancer depending on the tumour size at detection. This improvement in prognosis (defined as 1 minus the ratio of the risk of dying of screen detected cancer divided by the risk when the cancer has been diagnosed in the absence of screening) was based on analyses of the five Swedish randomised trials (*5*) (see Appendix 1).

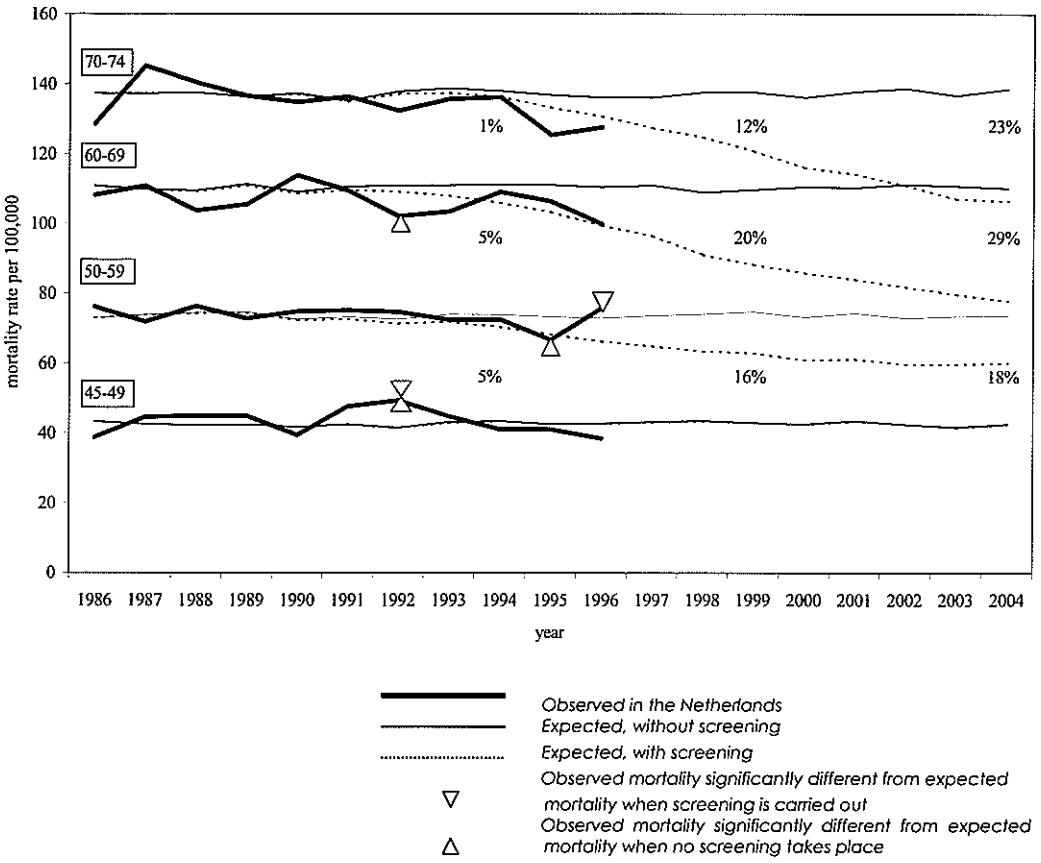
As the output of the microsimulation model MISCAN is subject to random fluctuation we simulated a multiple of the population under study to minimize the model fluctuation, resulting in a standard deviation that was five times smaller than in real life. A more detailed description of MISCAN is given elsewhere (*5,6*).

## Mortality reduction of breast cancer after the introduction of screening

The cancer mortality reduction after introduction of screening was assessed by comparing the mortality in a situation with and without screening. Mortality reduction was calculated for both a MISCAN model incorporating Dutch demographic, epidemiological and screen characteristics (*5,8*) and for a model adjusted for the demographic, epidemiological, and screen characteristics in the UK (9), assuming the North West health region to be representative of the UK (*10*). The age distribution of the population was based on the European standard population (*11*) in both models. Important differences in screening performance between the two models are the duration of the preclinical phase for small tumours and the survival after clinical diagnosis (9). The duration of the preclinical tumours up to 10 mm is assumed to be twice as long in the UK as in the Netherlands, which indicates a better detection of smaller tumours in the UK programme. Survival after clinical diagnosis is less favourable in the UK than in the Netherlands. In both models the gradual introduction of the screening programme was taken into account. In the Netherlands the programme was gradually introduced from 1989 to 1995, except for two pilot projects which started earlier. In the UK the separate screening programmes started between November 1987 and March 1993 [Wrench R, NHS Breast Screening Programme, personal communication].

*Figure 3.1*

*Age-standardised mortality rates, with and without screening observed in the Netherlands and predicted by MISCAN, for the age groups 45-49, 50-59, 60-69, and 70-74. The expected mortality reduction for these age groups after five, 10 and 15 years of screening is also shown.*




# RESULTS

## Mortality reduction after introduction of the Dutch programme

Figure 3.1 shows the breast cancer mortality, with and without screening, predicted by MISCAN for the age groups 45-49, 50-59, 60-69 and 70-74 in the period 1989-2004, the first 15 years after start of the screening programme, together with the observed breast cancer mortality in these age groups in the years 1989-1996 (12).

In the first years after the introduction of a screening programme little reduction in breast cancer mortality is expected: from 1989 to 1991 the expected mortality reduction was less than 1% for all

*Table 3.1*

*Mortality rates (SD) in the age-group 55-74 as observed in the Netherlands, their difference (95% confidence interval) from the observed mortality rate in the age-group 55-74 in 1986-88, and the expected mortality rates for screening by MISCAN*

| Year | Observed(SD) | | Observed compared with observed 1986-88 difference (95% CI) | | MISCAN |
|---|---|---|---|---|---|
| 1986-88 | 105.2 | (1.5) | | | 106.4 |
| 1989 | 102.9 | (2.8) | 2.3 | (- 4.0, 8.6) | 107.0 |
| 1990 | 107.5 | (2.8) | -2.4 | (- 8.7, 4.1) | 105.1 |
| 1991 | 105.0 | (2.8) | 0.2 | (- 6.1, 6.6) | 105.6 |
| 1992 | 102.1 | (2.7) | 3.2 | (- 3.1, 9.4) | 105.4 |
| 1993 | 101.1 | (2.7) | 4.1 | (- 2.1, 10.3) | 104.6 |
| 1994 | 103.9 | (2.7) | 1.3 | (- 4.9, 7.6) | 102.8 |
| 1995 | 100.2 | (2.7) | 5.0 | (- 1.2, 11.1) | 99.8 |
| 1996 | 100.0[a] | (2.7) | 5.2 | (- 1.0, 11.4) | 97.0 |
| 1997 | | | | | 94.3 |
| 1998 | | | | | 90.0 |
| 1999 | | | | | 87.3 |
| 2000 | | | | | 84.6 |

[a] preliminary data

ages. This is due to the time lag between early diagnosis and the effect of this on mortality, and due to the gradual introduction of screening. After that time the expected reduction in mortality increased continually. In 2004, the difference in the breast cancer mortality, with and without screening, is expected to increase to 18% in the 50-59 age-group, 29% in the 60-69 age-group and 23% in the 70-74 age-group.

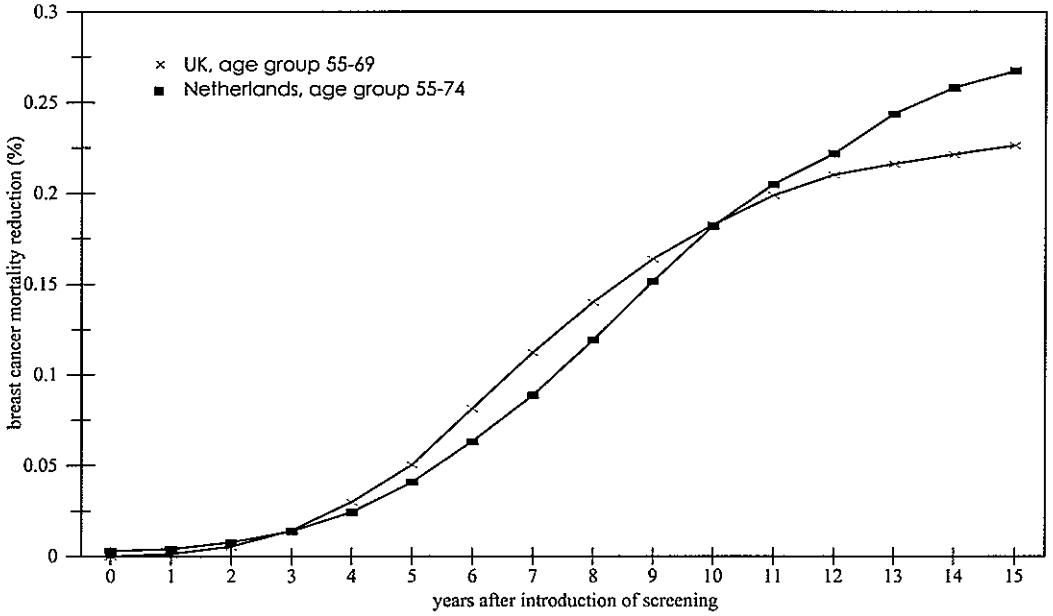
The observed mortality rates during 1986-1988, the period before screening started, compared well with expected mortality. After the introduction of screening in the years 1989-1996, the expected mortality rates, with screening, compared well with the observed mortality rates, but observed mortality rates were also not statistically different from the expected mortality, without screening, except for some peculiar fluctuations in the observed mortality rates.

Table 3.1 shows that the observed mortality in 1996 was still not significantly different from the observed mortality in 1986-88, before screening started. This result, however, was expected, because 1997 is the first year in which the expected probability of seeing a significant reduction is high enough to reach a power of 0.80.

The null hypothesis that the difference in the mortality in 1986-88 and a following year can be explained by random fluctuation will only be rejected with a probability of 80% if the mortality in that year is smaller than the mortality rate in 1986-88 minus 1.96 x (expected standard deviation of the difference in mortality under the null hypothesis) minus 0.84 x (expected standard deviation of the difference in mortality if the alternative hypothesis that mortality after 1988 is lower than the mortality before screening, is true). The mortality in a certain year must therefore be smaller than $105.2 - 1.96 \times 3.2 - 0.84 \times 3.1 = 96.3$. The first year in which the mortality in the age group 55-74 is

*Figure 3.2*

*Expected reduction in breast cancer mortality after the introduction of screening (UK 1988-2003, the Netherlands 1989-2004) in the age-group and 55-69 (UK) and 55-74 (The Netherlands).*




expected to be smaller than 96.3 is 1997. Therefore only from this year onwards is there a considerable probability that the reduction in mortality due to screening will be reflected in the mortality rates.

## Comparison of the Dutch and NHS breast screening programme

In the NHS breast screening programme reduction in breast cancer mortality in the first years after the start of screening was predicted to be low. Figure 3.2 compares the expected mortality reduction of the NHS and Dutch nationwide screening programme in the first 15 years of screening. We compared the mortality reduction in respectively the 20-year and 15-year age group (in accordance with the target age range of the Dutch and the NHS programmes respectively) with the highest mortality reduction, the age group 55-74 in the Netherlands to the age group 55-69 in the UK. After a short period in which the expected mortality reduction was higher for the Dutch screening programme because of two pilot projects which started earlier, the expected reduction in breast cancer mortality is higher for the UK until 10 years after the introduction of screening. This was due to the faster implementation of the screening programme in the UK. After that, the effects of the Dutch screening programme are expected to be higher, because of the shorter screening interval in the Netherlands (two years v three years). In the long term the Dutch nationwide

screening programme is predicted to produce a 29% mortality reduction for the age group 55-74, while the maximum mortality reduction of the NHS breast screening programme is predicted to be 24% in the age group 55-69.

## DISCUSSION

According to the MISCAN breast cancer model the reduction in breast cancer mortality due to the nationwide screening programme in The Netherlands is expected to be low in the first years after introduction of screening. These slight changes cannot be seen in the observed breast cancer mortality rates owing to the random fluctuation of mortality rates. It is expected that from 1997 onwards breast cancer mortality in The Netherlands will differ significantly from the mortality before screening in the age group 55-74; it will take longer to establish the effect of screening for the five year age groups. Swedish (*1*) and Australian (*13*) analyses have also shown the need for a long follow up to estimate the effects of screening on mortality.

One of the key parameters affecting the prediction of mortality reduction due to screening is the improvement in prognosis after screen detection, which is an extrapolation of the results of the Swedish randomised trials. This should be noted carefully, by comparing predicted and observed breast cancer mortality, if it is valid to assume that the Swedish trial results are representative of the Dutch situation. Until now, the observed mortality in the Netherlands has compared well with the expected mortality predicted by MISCAN if screening takes place, but this is not yet decisive as the expected mortality predicted by MISCAN when there is no screening does not differ significantly from the observed breast cancer mortality. Furthermore, other factors in addition to screening, such as improvement in treatment, changes in risk factors and performance of screening, may influence the breast cancer mortality and complicate the comparison between the observed and expected mortality. These factors also affect assessment of the impact of a screening programme on breast cancer mortality by comparing the observed mortality before and after the introduction of screening. This is illustrated by the mortality reduction reported by Quinn (*2*). The observed 12% reduction in breast cancer mortality after seven years of screening was assumed to have been influenced by the widespread adoption of tamoxifen during this period. It is, therefore, unknown to what extent the reduction in mortality was attributable to screening. We predicted the reduction in the breast cancer mortality in 1994 induced by the NHS breast screening programme in England and Wales would be 8% in the 55-69 age group. Any additional observed mortality reduction is then due to improvement in the treatment, but random fluctuation may also play a part. Chu *et al.* (*3*) also reported that their statistical modelling indicated that the recent drop in the breast cancer mortality rate for American white females (6.8% from 1989 to 1993) was too rapid to be explained only by the increased use of mammography. In 2000 the expected mortality reduction attributable to the NHS breast screening programme in women aged 50-69 is 17%. Accordingly, the current *Health of the Nation* target of a 25% reduction in the mortality by the year 2000 in the 50-69 age-group seems difficult to achieve with the current screening programme alone - other factors are needed. For the age group 50-64 we expect a maximum mortality reduction of 17% in the long

*Table 3.2*

*Improvement in prognosis after screen detection*

| State | Reduced risk of dying from breast cancer[a] |
|---|---|
| Screen-detected dCIS | 1.000 |
| Screen-detected T1a | 0.892 |
| Screen-detected T1b | 0.814 |
| Screen-detected T1c | 0.567 |
| Screen-detected T2+ | 0.395 |

[a] Defined as the proportion of women with cancer that have their risk of dying of breast cancer reduced. This proportion depends on cancer size at detection and is based on the results of analysis of the Swedish randomised trials

term, which is lower than the estimated mortality reduction of 21% for East Anglia based on interval cancer rates (*14*). This difference may reflect the difference between the results of the Two County study only, and the five Swedish randomised trials combined, used in the respective studies.

In the Finnish report (*4*) the effect on breast cancer mortality attributable to the nation-wide screening programme was estimated by comparing the observed breast cancer mortality of the women invited for screening with the breast cancer mortality in women not yet invited. When this method is used, improvement in treatment, changes in risk factors, cohort effects and performance of screening do not affect assessment of the effects of screening, as these factors can be expected to influence both groups evenly. But the reported reduction in breast cancer mortality was based on small numbers and was not statistically significant. If the comparison between the women screened first and those invited last is repeated after a longer follow up period, when the women invited last have also been screened, the difference in breast cancer mortality between these two groups will still represent the difference in breast cancer mortality, with and without screening, and the numbers may by then be sufficiently large to show a statistically significant reduction. However, the period in which this comparison is possible is also limited, as the difference in breast cancer mortality between the two groups will become relatively smaller by increasing numbers of breast cancer deaths in both groups, and, in the long term, will not be distinguishable from random fluctuation. This method of demonstrating mortality reduction attributable to screening can also be applied to the Dutch nationwide screening programme and the NHS breast screening programme as in these countries, also, the screening programmes were introduced gradually. For the NHS breast screening programme the phased introduction of screening has already been used to compare the survival of patients with cancer in an unscreened population with the survival of those with interval cancers (*15*).

In conclusion, the effects of screening on breast cancer mortality are low in the first years after the introduction of screening, and are expected to become apparent around 10 years later. Accordingly, we expected the effect of screening in the Netherlands to be visible only from 1997 onwards and that 70% of the reported 12% mortality reduction in England and Wales in 1994 can be attributed to screening.

# REFERENCES


(1) Tornberg S, Carstensen J, Hakulinen T, Lenner P, Hatschek T, Lundgren B. Evaluation of the effect on breast cancer mortality of population based mammography screening programmes. J Med Screen 1994;1:184-7.

(2) Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. BMJ 1995;311:1391-5.

(3) Chu KC, Tarone RE, Kessler LG, Ries LA, Hankey BF, Miller BA, et al. Recent trends in U.S. breast cancer incidence, survival, and mortality rates. J Natl Cancer Inst 1996;88:1571-9.

(4) Hakama M, Pukkala E, Heikkila M, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. BMJ 1997;314:864-7.

(5) de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. J Natl Cancer Inst 1995;87:1217-23.

(6) van Oortmarssen GJ, Habbema JD, van der Maas PJ, de Koning HJ, Collette HJ, Verbeek AL, et al. A Model For Breast Cancer Screening. Cancer 1990;66:1601-12.

(7) de Koning HJ, Fracheboud J, Boer R, Verbeek AL, Collette HJ, Hendriks JH, et al. Nation-wide breast cancer screening in The Netherlands: support for breast-cancer mortality reduction. National Evaluation Team for Breast Cancer Screening (NETB). Int J Cancer 1995;60:777-80.

(8) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JC, Collette HJ, Hendriks JH, 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.

(9) Boer R, de Koning HJ, Threlfall A, Warmerdam P, Street A, Friedman E, et al. Cost effectiveness of shortening screening interval or extending age range of NHS breast screening programme: computer simulation study. BMJ 1998;317:376-9.

(10) Chamberlain J, Moss S, Kirkpatrick A, Michell M, Johns L. National Health Service breast screening programme results for 1991-2. BMJ 1993;307:353-6.

(11) Waterhouse J, Muir C, Correa P, Powell J. Cancer Incidence in Five Continents, Volume III. Lyon: IARC Scienific Publications No. 15; 1976.

(12) Netherlands Central Bureau of Statistics. Death by causes of death, age and gender 1989-1996: Netherlands Central Bureau of Statistics, published annually.

(13) Smith C, Kricker A, Armstrong B. Breast cancer mortality trends in Australia: 1921 to 1994. Med J Aust 1998;168:11-4.

(14) Day N, McCann J, Camilleri-Ferrante C, Britton P, Hurst G, Cush S, et al. Monitoring interval cancers in breast screening programmes: the east Anglian experience. Quality Assurance Management Group of the East Anglian Breast Screening Programme. J Med Screen 1995;2:180-5.

(15) Collins S, Woodman C, Threlfall A, Prior P. Survival rates from interval cancer in NHS breast screening programme. BMJ 1998;316:832-3.

# 4

# Low risk of cervical cancer during a long period after negative screening in the Netherlands

## SUMMARY

*Introduction* – For cervical cancer screening to be considered effective, one condition is a low incidence of cervical cancer after negative screening compared to that in the absence of screening. This relative risk was studied for the period 1994-1997 in the Netherlands and compared with previous studies.

*Material and Methods* - All cases of invasive cervical cancer diagnosed from 1994 to 1997 in the Netherlands were related to woman-years at risk, stratified by age, number of preceding negative screenings and time since the preceding negative screening. These incidence rates were compared with that before screening started in the Netherlands.

*Results* - The relative risk increases from 0.13 in the first year after screening to 0.24 after more than 6 years after screening for women with one previous negative screening. These figures reduce to 0.06 and 0.18 respectively, for women with two or more previous screenings. However, these estimates are less favourable when account is taken of the likely decrease in risk for cervical cancer in the period studied.

*Conclusions* - Our data show a low relative risk of cervical cancer for several years following the last negative Pap smear. However, the denominator of the relative risk, that is, the incidence without screening, may have been overestimated. The same applies to the IARC multi-country study, and may have caused too optimistic expectations about the effectiveness of cervical cancer screening.

## INTRODUCTION

The effectiveness of cervical cancer screening has never been established by randomised controlled trials. Evidence for mortality reduction, the primary aim of cervical cancer screening, came from studies that compared regions or individuals with different screening intensities (*1-3*).

One indicator of such effectiveness is the incidence of cervical cancer after a negative screen related to that in the absence of screening. The smaller this relative risk, the better has screening succeeded in selecting women at low risk of getting cervical cancer in subsequent years. Combined with the improvement in prognosis for women with a true positive screening result, such a selective power warrants a reduction in incidence and mortality.

The present study estimates the relative risk for cervical cancer after a negative screen on the basis of nationwide Dutch data. This risk is determined by the duration of the screen-detectable preclinical stage and the sensitivity of the test for this stage. The predictive value of a negative screen for not developing cervical cancer increases with a longer preclinical duration and a higher sensitivity.

When data from large-scale screening programmes became available, a working group of the International Agency on Research for Cancer (IARC) estimated the incidence after a negative screen, compared to the estimated background incidence in eight countries, that is, Canada, Scotland, Iceland, Denmark, Norway, Sweden, Switzerland and Italy (*4,5*). Expectations about the effectiveness of cervical cancer screening are often based on the results of this 'classic' study, and important models for such screening have been validated using the IARC results (6-9). This was the rationale to compare the results of the present study with those of the IARC study: if the results correspond, expectations concerning the effects of cervical cancer screening are reinforced.

## MATERIALS AND METHODS

All cytological and histological examinations of the cervix in the Netherlands up to 31 December 1997 were retrieved from the Pathological National Automated Archive (PALGA). When this registry started in 1975 few laboratories participated, but within a decade a high level of national coverage was achieved.

Using the PALGA identification method (i.e. first four characters of the family name, date of birth and gender), different examinations of the same woman could be linked.

In this study, all cases of invasive cervical cancer occurring from 1994 to 1997 were identified by selecting histologically confirmed diagnoses of invasive cancer from the database. These include all malignant neoplasms of the cervix, most of which are squamous-cell carcinomas.

Woman-years were counted for each woman, from each negative screen until the next negative screen, until the histological diagnosis of a (precursor of) invasive carcinoma, or until 31 December 1997. A negative screen was defined as an episode consisting of a cytological or histological examination with a negative result, or a cytological examination with a positive result without a

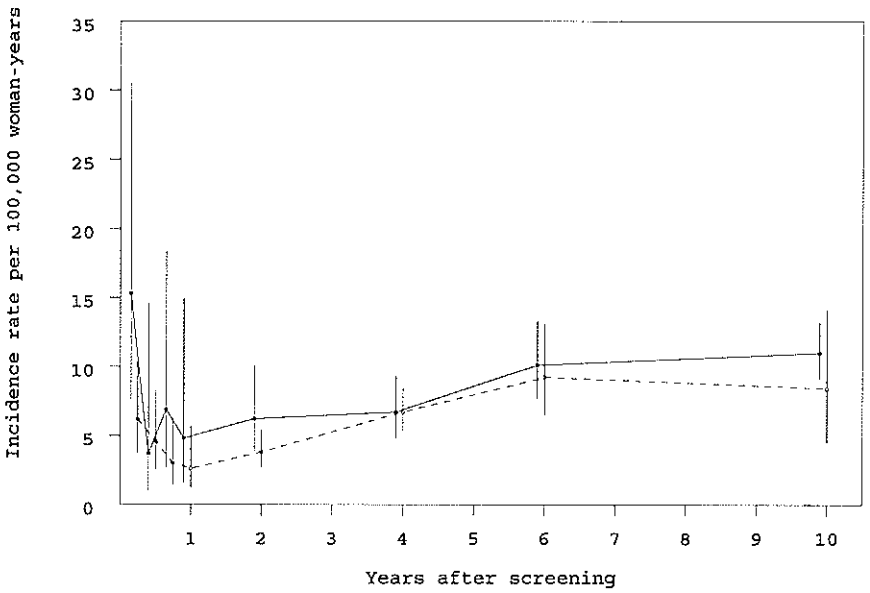
histological confirmation of a (precursor of) invasive cancer (see Appendix and Table A1 for definitions).
The invasive cases were related to woman-years at risk, and presented as the number of cases per 100,000 woman-years at risk. These incidence rates were stratified by age, number of preceding negative screenings, and the interval since the preceding negative screen.

The incidence of invasive cervical cancer was calculated for women aged 35-64 years with one previous negative screen and with two or more previous negative screenings. This method is comparable to that of the IARC study. Next, the relative risk for cervical cancer was calculated by dividing the incidence rate after screening by the incidence in the absence of screening.
Since this background incidence cannot be observed, it had to be estimated indirectly. For this we used the clinical incidence of invasive cervical cancer in the period 1965-69, this being the last period before screening was introduced in the Netherlands. National incidence figures were based on incidence data of three regions in the Netherlands (Friesland, The Hague and Rotterdam) covering together 8% of the women in the Netherlands (*10*). Regional differences in cervical cancer incidence have been accounted for by using the differences between the age-standardised mortality rate of cervical cancer for these three regions and for the entire land for the period 1968-1978 as a proxy.

The identification method used by PALGA (first four characters of the family name, etc.) is not 100% exclusive and will sometimes combine two or more women in one identification code. To investigate the influence of this lack of discriminative power of the identification key, we also calculated the incidence rates excluding the examinations of those women with 0.5% and 1% of the most frequently occurring first four characters of the family name. The corresponding percentages of women thus excluded from analysis are 31.7% and 43.5%, respectively.
The lack of discriminative power of the identification key leads to an upward bias in incidence after a negative screen, because negative screening results may be erroneously linked to a cancer. We indeed found that the incidence rate including all women is about 20% higher after one and two or more negative screenings than the rate after excluding 0.5% of the most frequent first four characters of the family name. As the difference in incidence between excluding 0.5% and 1% of the most frequent first four characters is very small, in our analyses, we chose to exclude only those women with 0.5% of the most frequent first four characters of family name in the corresponding table and figures (Table 4.1, Figures 4.1-4.2). This served to limit the lack of discriminative power of the identification key while maintaining sufficiently large numbers on which to base our analysis.

*Figure 4.1*

*Incidence of invasive cervical cancer over time since the last negative screen, for women with one and two or more preceding negative screenings. The solid line represents one negative screen and the dashed line represents two or more negative screens (95% CIs are shown).*




## RESULTS

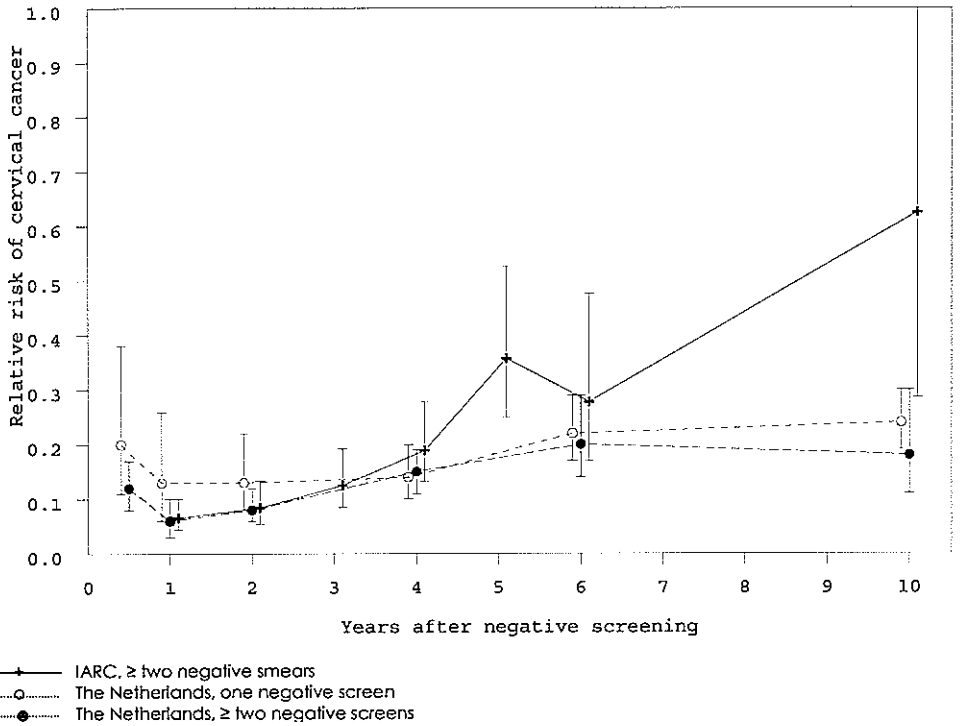
A total of 1648 invasive carcinomas in women aged 35-64 years were retrieved in the period 1994-1997 in the Netherlands. Of these carcinomas, 879 were diagnosed without a preceding negative screen, 376 after one negative screen, and 393 after two or more preceding negative screenings.

Figure 4.1 shows the incidence of invasive carcinoma per 100,000 woman-years by interval since the last negative screen for women with one and with two or more negative screenings. In the first months after negative screening, the incidence of cervical cancer is relatively high. This may be because the women and/or physicians were not reassured by the recent (false) negative Pap smear result (e.g. because of persisting signs or symptoms) and thus elected for additional diagnostic procedures. After this initial peak, the incidence is low and will mainly consist of cases of neoplasia missed at screening. Over time, the incidence increases because of new lesions that developed after the negative screening.

Figure 4.2 shows the incidence of invasive cervical cancer over time since the previous negative screening, compared with the incidence of invasive cervical cancer in the period 1965-1969 in the Netherlands, which was 46.1 per 100,000 women years for women aged 35-64 years. In the first years, the relative risk is lower after two or more negative screenings than after only one negative screening. Figure 4.2 also compares our incidence data with the IARC results. In the first years after a negative screen the relative risk for cervical cancer is comparable in the two studies, but from 4 years onwards the relative risk is higher in the IARC study than in the Netherlands.

*Figure 4.2*

*Relative risk of invasive cervical cancer after negative Pap smears as assessed in the eight countries contributing to the IARC study (see text), compared with the risk in the Netherlands after one and two or more negative screens (95% CIs are shown).*




## DISCUSSION

As estimated from an age period cohort (APC) analysis of prescreening mortality rates in the Netherlands, the risk of cervical cancer decreases sharply for cohorts of women born after 1927 (*11*). Based on these data, the projected incidence for the period 1994-1997 in the absence of screening on the basis of these figures was 20.5 per 100,000 woman-years. For the women born after 1950 this might be an underestimate because there was an increased risk for cervical cancer in the youngest cohorts (*12*), whereas we extrapolated the low risk of the latest cohort for which prescreening mortality rates were available (born 1940 - 1950) to the youngest cohorts. Table 4.1 gives the relative risk using the projected incidence for the period 1994-1997 (from the APC analysis) compared with the incidence just before screening started (1965-1969). Using the projected incidence for 1994-1997 results in a factor two higher risk.

Estimation of the background incidence has also proven problematic in other studies. In the IARC study, some centres used a case-control approach whereas (as in the present study) others used a cohort approach. Some cohort studies used the incidence in women who were never screened as background incidence, while others used the incidence before screening became widespread. Both estimates have their problems: that is women never attending screening are reported to be at

*Table 4.1*

*Relative risk of cervical cancer (95% confidence interval) after two or more negative screenings over time since the last negative screening for women aged 35-64 years and the number of actual cancer cases (1994-1997).*

| Interval | Relative risk | | Number of cancer cases |
|---|---|---|---|
| | Incidence in 1965-1969 | Projected incidence[a] in 1994-1997 in a situation without screening | |
| 0-6 months | 0.12 [0.08-0.17] | 0.26 [0.18-0.39] | 26 |
| 7-12 months | 0.06 [0.03-0.10] | 0.14 [0.08-0.23] | 13 |
| 1-2 years | 0.08 [0.06-0.12] | 0.19 [0.13-0.26] | 31 |
| 2-4 years | 0.15 [0.11-0.19] | 0.33 [0.26-0.42] | 65 |
| 4-6 years | 0.20 [0.14-0.29] | 0.45 [0.32-0.64] | 31 |
| 6-10 years | 0.18 [0.11-0.30] | 0.41 [0.25-0.68] | 15 |

Two different ways of calculating the background incidence (i.e. the denominator of the relative risk) were used (see text).

[a] Based on the APC analysis of pre-screening mortality (see Discussion).

higher risk for cervical cancer (*13-15*) and in several countries a nonscreening-related decrease in cervical cancer risk in the considered period is reported (*2,11,16*). The case-control studies contributing to the IARC results may also have resulted in an underestimation of the relative risk because of healthy screenee bias and frequency bias. Sasieni et al. found a factor two higher relative risk compared with the relative risk of the IARC and the present study. However, they used a case-control approach with carefully selected appropriate controls to cases, which may have reduced the bias (*17*). Viikki et al., who found a three times higher relative risk, used the incidence of the total population in the screening period (*18*). However, because of the incidence-reducing effect of screening this will be an underestimation of the background incidence, which may have led to the relatively high risks.

Other differences between studies are less important. The IARC results were presented for two or more previous negative screenings only, whereas other studies (*17-19*) also included a single previous screening. This latter case leads to a higher incidence and thus relative risk, especially in the period immediately following a negative screen (see Figure 4.2).

In contrast to the four studies discussed above, we considered a positive smear result that was not followed by a histological diagnosis as a negative screen. Calculating the relative risk after a negative Pap smear did not have a strong effect on the results.

As a result of the methodological differences, comparison of the performance of screening between different countries is difficult. Nevertheless, for example, the suspected suboptimal performance of screening in the UK in the 1990s (*20,21*) may have contributed to the high relative risks reported in that period (*17*).

In most of the prominent models on cervical cancer screening (6-9), the IARC results have been used to validate the assumptions of the model regarding the sensitivity for, and duration of, the preclinical disease stage. If, however, the IARC results are too favourable (e.g. because of overestimation of the background risk) then these models will also overestimate the effectiveness of cervical cancer screening.

The longer ago screening was started, the greater the uncertainty will be about the background incidence. As a result, assessment of the relative risk after negative screening will become increasingly difficult, that is more difficult than in the IARC study and the current study.

In conclusion, our data show that the relative risk for cervical cancer incidence is low for several years following a negative screening using the Pap smear. There are strong indications that relative risk estimates are too favourable, because of a too high estimate of the background incidence. However, even an underestimate of the background incidence shows a considerable reduction in the relative risk after negative screening. The overestimation also applies to the widely used IARC results, and may have caused too optimistic expectations about the effectiveness of cervical cancer screening.

## REFERENCES


(1) Clarke EA, Anderson TW. Does screening by "Pap" smears help prevent cervical cancer? A case-control study. Lancet 1979;2:1-4.

(2) Laara E, Day NE, Hakama M. Trends in Mortality From Cervical Cancer in the Nordic Countries: Association With Organised Screening Programmes. Lancet 1987;1:1247-9.

(3) van der Graaf Y, Zielhuis GA, Peer PG, Vooijs PG. The effectiveness of cervical screening: a population-based case-control study. J Clin Epidemiol 1988;41:21-6.

(4) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for cancer of the uterine cervix. Lyon; 1986.

(5) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J 1986;293:659-64.

(6) van Oortmarssen GJ, Habbema JD. Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative pap smears. Int J Epidemiol 1995;24:300-7.

(7) Eddy DM. The frequency of cervical cancer screening. Comparison of a mathematical model with empirical data. Cancer 1987;60:1117-22.

(8) Gustafsson L, Adami HO. Cytologic screening for cancer of the uterine cervix in Sweden evaluated by identification and simulation. Br J Cancer 1990;61:903-8.

(9) Gyrd-Hansen D, Holund B, Andersen P. A cost-effectiveness analysis of cervical cancer screening: health policy implications. Health Policy 1995;34:35-51.

(10) Central Cancer Registry. Overview of incidence of carcinoma of the cervix uteri in Friesland, Rotterdam, and The Hague in the years 1960-1970. Koningin Wilhelmina Fonds; 1993.

(11) van Ballegooijen M. Effects and costs of cervical cancer screening [Thesis]. Rotterdam: Department of Public Health, Erasmus University; 1998.

(12) Beral V, Hermon C, Munoz N, Devesa SS. Trends in cancer incidence and mortality. Cancer Surv 1994;19/20:265-85.

(13) 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.

(14) Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-6.

(15) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.

(16) Cuzick J, Boyle P. Trends in cervix cancer mortality. Cancer Surv 1988;7:417-39.

(17) Sasieni PD, Cuzick J, Lynch-Farmery E. Estimating the efficacy of screening by auditing smear histories of women with and without cervical cancer. The National Co-ordinating Network for Cervical Screening Working Group. Br J Cancer 1996;73:1001-5.

(18) Viikki M, Pukkala E, Hakama M. Risk of cervical cancer after a negative Pap smear. J Med Screen 1999;6:103-7.

(19) Mitchell HS, Giles GG. Cancer diagnosis after a report of negative cervical cytology. Med J Aust 1996;164:270-3.

(20) Raffle AE, Alden B, Mackenzie EF. Detection rates for abnormal cervical smears: what are we screening for? Lancet 1995;345:1469-73.

(21) Anonymous. The screening muddle. Lancet 1998;351:459.

# APPENDIX

## Episodes

After linking all examinations belonging to one woman (based on the first four characters of her family name and date of birth), the examinations were divided into primary and secondary ones. The definition of primary and secondary examinations is given in Table A.1. Briefly, an examination is considered to be secondary if in the 48 months preceding the examination there has been a non-negative examination result that could have given rise to the present examination. If not, the examination is a primary one. Next, series of examinations belonging to one woman were divided into episodes. An episode is defined as a time period consisting of a primary examination with (in case the primary examination is not negative) the accompanying follow-up examinations. If a primary examination is negative, the episode consists of that examination only.

***Table A.1***

*Definition of primary and secondary examinations*

An examination is secondary if in the preceding 48 months:
- there is a histological diagnosis of invasive cervical cancer
- there is a histological diagnosis of pre-invasive cancer after which there has not been three cytological examinations with a negative result
- there is a cytological diagnosis of severe dysplasia after which there has not been three cytological examinations with a negative result
- there are at least two cytological diagnoses of light-moderate dysplasia after which there has not been three cytological examinations with a negative result
- there is a histological examination without diagnosis after which there is no histological examination with a negative result or three cytological examinations with a negative result
- there is a cytological diagnosis of light-moderate dysplasia after which there is no cytological examination with a negative result
- there is an inadequate cytological examination after which there has not been a cytological examination with a negative result
- there is a cytological examination without endocervical cells after which there has not been a cytological examination with a negative result

otherwise an examination is primary

## Screen-detected and interval carcinomas

An invasive carcinoma was categorised as screen-detected when, in the episode in which the invasive carcinoma is diagnosed, the reason for the primary smear was coded as being for screening purposes. If the reason for the primary smear was not known, the invasive carcinoma was considered to be screen-detected if no biopsy was taken at the same time as the primary examination and if there was no unexpected (from a follow-up point of view) histological examination in the episode. Otherwise, the invasive carcinoma was considered to be diagnosed because of symptoms.

## Age

Age was determined at the time of the screening examination for screen-detected carcinomas and at the time of histological confirmation of invasive cancer for symptomatic cases.

## Negative examinations

All primary examinations, both cytological and histological, without a histological confirmation of a (precursor of) invasive cancer in their episode, are counted as a negative examination. This implies that if the primary examination had a positive result, this examination is counted as negative examination if no histological confirmation of the positive smear result takes place.
In case an episode contains a histological diagnosis of a cervical neoplasia, we assume that the woman will have the normal risk of getting cervical cancer after the treatment of the cervical abnormality, and the counting of the negative examinations is reset to zero.

## Interval since preceding negative examination

The interval since the preceding negative examination is defined as the time between the last negative primary examination and the screening examination for screen-detected carcinomas and the time between the last negative primary examination and the histological confirmation of invasive cancer for cancers detected because of symptoms.

# 5

# COST-EFFECTIVENESS OF CERVICAL CANCER SCREENING: COMPARISON OF SCREENING POLICIES

## SUMMARY

*Background* - Recommended screening policies for cervical cancer differ widely among countries with respect to targeted age range, screening interval, and total number of scheduled screening examinations (i.e., Pap smears). We compared the efficiency of cervical cancer-screening programs by performing a cost-effectiveness analysis of cervical cancer-screening policies from high-income countries.

*Methods* - We used the microsimulation screening analysis (MISCAN) program to model and determine the costs and effects of almost 500 screening policies, some fictitious and some actual (i.e., recommended by national guidelines). The costs (in U.S. dollars) and effects (in years of life gained) were compared for each policy to identify the most efficient policies.

*Results* – There were 15 efficient screening policies (i.e., no alternative policy exists that results in more life-years gained for lower costs). For these policies, which considered two to 40 total scheduled examinations, the age range expanded gradually from 40-52 years to 20-80 years as the screening interval decreased from 12 years to 1.5 years. For the efficient policies, the predicted gain in life-expectancy ranged from 11.6 to 32.4 days, compared with a gain of 46 days if cervical cancer mortality were eliminated entirely. The average cost-effectiveness ratios increased from $ 6700 (for the longest screening interval) to $ 23900 per life-year gained. For some countries, the recommended screening policies were close to efficient, but the cost-effectiveness could be improved by reducing the number of scheduled examinations, starting them at later ages, or lengthening the screening interval.

*Conclusions* - The basis for the diversity in the screening policies among high-income countries does not appear to relate to the screening policies' cost-effectiveness ratios, which are highly sensitive to the number of Pap smears offered during a lifetime.

# INTRODUCTION

The purpose of cervical cancer screening with the Pap smear test is to detect preinvasive cancers and to prevent subsequent death from the disease. Although no randomized, controlled trials on mortality reduction from cervical cancer screening have been performed, there is ample evidence that screening has led to a reduction in cancer-related mortality (*15-18*). In the high-income countries of Western Europe, North-America, and Australia, preventive Pap smear tests are performed on a large scale in organized, often invitation-based, programs and by the personal initiative of individual women and physicians or practitioners. The screening recommendations and official policies in different countries and regions show considerable variation. For example, in The Netherlands and Finland, the recommended number of Pap smears during a woman's lifetime is seven, whereas in Germany and Australia, the recommended number is more than 25. In addition, in the U.K., The Netherlands, and Finland, the recommended time interval between screening examinations is 5 years, whereas in Australia it is 2 years and in Germany it is 1 year. There are also differences in the target age range. For example, in The Netherlands, screening is offered between the ages of 30 and 60 years (before 1996, screening was offered between the ages of 35 and 53 years), whereas in Australia screening is recommended between the ages of 18 and 70 years (*21*).

How much the differences in the recommendations alter the cost effectiveness of the screening policies is unclear. The method of choice for the evaluation and comparison of different health care policies is cost-effectiveness analysis, which, for cervical cancer screening, involves a comparison of different screening policies that consider screening costs, possible savings in treatment, and potential health effects, such as life-years gained and cervical cancer deaths prevented. Such a comparison of policies would lead to the identification of efficient policies for which no alternative policies currently exist that result in more life-years gained for lower costs. In the rational decision-making process for making cervical cancer screening recommendations, a policy maker can compare the incremental and/or average costs per life-year gained of the efficient policies with the maximum allowed values or thresholds for the incremental and/or average costs per life-year gained and identify the most efficient screening policy given the available resources.

In this study, the microsimulation screening analysis (MISCAN) model (*7,8*) for cervical cancer screening was used to evaluate and compare almost 500 screening policies that differed with respect to the recommended number of screenings, screening intervals, and targeted age ranges. These screening policies consist of fictitious screening policies, policies used in countries with a cervical screening program or in which screening was recommended in national guidelines (*1,5,11,13,14,20,21*), policies recommended in the literature (*23*), and policies found to be cost-effective in other studies (*6,10,12,16,19,22,24,25*). We estimated the life-years gained and costs of the policies and identified efficient screening policies. We determined the best policy for different thresholds for the incremental costs per life-year gained. The results were compared with existing policies and recommendations, and with policies that have emerged from other cost-effectiveness

analyses (6,10,12,16,19,22,24,25). Our analysis uses demographic, epidemiologic, screening and treatment characteristics from the Netherlands. Because these characteristics may be different for other countries, we investigated the extent to which differences in demographic, epidemiologic, screening and treatment characteristics result in differences in screening recommendations.

# METHODS

## Policies

The fictitious screening policies considered in this cost-effectiveness analysis are listed in Table 5.1. We also included screening policies used in countries with cervical screening programs or recommended in national guidelines (1,5,11,13,14,20,21) and screening policies considered in other cost-effectiveness analyses (6,10,12,16,19,22,24,25).

## MISCAN

Costs and effects for the different screening policies were estimated using MISCAN (7,8). The MISCAN simulation program was developed at the Department of Public Health, Erasmus University Rotterdam, The Netherlands, and has been used to evaluate breast, cervical, colon and prostate cancer screening programs. In MISCAN, a comparison is made between the situation with and without screening. A large population (i.e., 40 million women) is generated. This population consists of fictitious individual life histories, in which some of the women may develop cancer and some may die of the disease. This results in an age-specific and time-specific output of cancer incidence and mortality. This fictitious population then undergoes simulated screening. Screening may change some of the life histories. For example, in some life histories, preclinical lesions will be detected by screening, which may prevent further development of the disease and subsequent cancer-related death. The aggregated changes in all of the life histories constitute the effectiveness of the screening program. The cervical model specifications used as input for the MISCAN program include demographic characteristics, the epidemiology and natural history of the disease, the screening characteristics, and the costs. A detailed description of the MISCAN program is given elsewhere (7,8).

***Table 5.1***

*Screening policies evaluated in the cost-effectiveness analysis.*

| No. of scheduled examinations | Screening interval, y | Starting age for screening, y |
|---|---|---|
| 1 | - | 32, 35, 37, 40, 42 |
| 2 | 9 | 37, 40, 42 |
| | 10 | 25, 30, 35, 37, 40, 42 |
| | 11 | 25, 30, 35, 37, 40, 42 |
| | 12 | 25, 30, 35, 37, 40, 42 |
| | 13 | 37,40, 42 |
| 3 | 8 | 32, 35, 37 |
| | 9 | 32, 35, 37 |
| | 10 | 32, 35, 37, 40, 42 |
| | 11 | 20, 25, 30, 32, 35, 37, 40, 42 |
| | 12 | 20, 30, 35, 37, 40, 42 |
| | 13 | 20, 25, 30, 32, 35, 37, 40 |
| | 14 | 20, 30, 32, 35, 37, 40 |
| | 15 | 20, 25, 30, 32, 35, 37, 40 |
| 4 | 7 | 32, 35, 37 |
| | 8 | 30, 32, 35, 37, 40 |
| | 9 | 20, 25, 30, 32, 35, 37, 40 |
| | 10 | 20, 30, 32, 35, 37, 40 |
| | 11 | 20, 25, 30, 32, 35, 37, 40 |
| | 12 | 20, 30, 40 |
| | 13 | 20, 25, 30, 35, 40 |
| 5 | 6 | 20, 22, 25, 27, 30, 32, 35, 37, 40 |
| | 7 | 20, 22, 25, 27, 30, 32, 35, 37, 40 |
| | 8 | 20, 25, 27, 30, 32, 35, 37, 40 |
| | 9 | 20, 25, 30, 32, 35, 37, 40 |
| | 10 | 20, 30, 32, 35, 40 |
| | 11 | 20, 25, 30, 35, 40 |
| 6 | 5 | 25, 27, 30, 32, 35 |
| | 6 | 20, 25, 27, 30, 32, 35, 40 |
| | 7 | 20, 25, 27, 30, 32, 35, 40 |
| | 8 | 20, 22, 25, 27, 30, 32, 35, 40 |
| | 9 | 20, 25, 27, 30, 40 |
| | 10 | 20, 25, 30, 35, 40 |
| 7 | 4 | 25, 27, 30, 32 |
| | 5 | 20, 25, 27, 30, 32, 35, 40 |
| | 6 | 20, 25, 27, 30, 32, 35, 40 |
| | 7 | 20, 25, 27, 30, 32, 35, 40 |
| | 8 | 20, 25, 27, 30, 40 |
| | 9 | 20, 22, 25, 27, 30 |
| 8 | 4 | 20, 25, 30, 35, 40 |
| | 5 | 20, 30, 40 |
| | 6 | 20, 22, 25, 27, 30, 35, 40 |
| | 7 | 20, 22, 25, 27, 30, 40 |
| | 8 | 20, 22, 25, 27, 30 |
| | 9 | 20, 22, 25, 27 |
| 9 | 3 | 20, 25, 30, 35, 40 |
| | 4 | 20, 30, 40 |
| | 5 | 20, 25, 27, 30, 35, 40 |
| | 6 | 20, 25, 27, 30, 40 |
| | 7 | 20, 22, 25, 27, 30 |
| | 8 | 20, 22, 25 |
| 10 | 2 | 20, 25, 30, 35, 40 |
| | 3 | 20, 30, 40 |
| | 4 | 20, 25, 27, 30, 35, 40 |
| | 5 | 20, 22, 25, 27, 30 |
| | 6 | 20, 22, 25, 27, 30 |
| | 7 | 20, 22, 25, 27 |

| | | |
|---|---|---|
| 12 | 1 | 20, 25, 30, 35, 40 |
| | 2 | 20, 30, 40 |
| | 3 | 20, 25, 27, 30, 35, 40 |
| | 4 | 20, 22, 25, 27, 30 |
| | 5 | 20, 22, 25, 27, 30 |
| | 6 | 17, 20, 22, 25, 27 |
| 15 | 1 | 20, 25, 30, 35, 40 |
| | 2 | 20, 25, 30, 35, 40 |
| | 3 | 17, 20, 22, 25, 27, 30, 35, 40 |
| | 4 | 17, 20, 22, 25, 27, 30 |
| | 5 | 17, 20, 22 |
| 20 | 1 | 20, 25, 30, 35, 40 |
| | 2 | 17, 20, 22, 25, 30, 35, 40 |
| | 3 | 17, 20, 22, 25, 27, 30 |
| | 4 | 20, 22 |
| 25 | 1 | 20, 25, 30, 35, 40 |
| | 1½ | 17, 20, 22, 25, 27 |
| | 2 | 20, 22, 25, 27, 30, 35, 40 |
| | 2½ | 20, 22, 25 |
| 30 | ½ | 20, 25, 30, 35, 40 |
| | 1 | 20, 25, 30, 35, 40 |
| | 1½ | 17, 20, 22, 25, 27 |
| | 2 | 17, 20, 22, 25, 27, 30 |
| 40 | ½ | 20, 25, 30, 35, 40 |
| | 1 | 17, 20, 22, 25, 30, 35, 40 |
| | 1½ | 17, 20, 22, 25, 27 |

## Model Specifications: Demography and Epidemiology

The Dutch population at risk for cervical cancer was simulated from demographic data (26) and hysterectomy (for reasons other than cervical cancer that were obtained from the National Hospital Admission Registration (27).

The background risk, i.e., the risk of dying of cervical cancer in a situation without screening, of cervical cancer-related mortality was derived from an age-period cohort analysis (28). For our analyses, we assumed that the lifetime background risk of developing cervical cancer (or its progressive precursors) was proportional to the estimated relative level of cervical cancer mortality for each birth cohort. Furthermore, we assumed that there was a fixed ratio between the lifetime risks of preinvasive disease that will spontaneously regress and preinvasive disease that will progress to cervical cancer. The cumulative incidence of progressive preinvasive cervical cancer by birth cohort was 0.0229 for those born from 1889 through 1918, 0.0235 for those born from 1919 through 1928, 0.0128 for those born from 1929 through 1938, 0.0106 for those born from 1939 through 1948, and 0.0148 for those born from 1949 through 2000.

For our analyses, we considered the reported negative association between attendance to the screening program and risk of cervical cancer (29-31) by subdividing the simulated population into two risk strata: 90% of the women were assumed to be potential attenders and were assumed to have a low risk of developing cervical cancer, and the remaining 10% of the population was assumed to be persistent nonattenders. On the basis of results from British Columbia in which the risk of attenders was estimated at 0.74 of the average risk (31), we assumed that the persistent nonattenders have a risk of cervical cancer three times higher than that of the attenders.

*Table 5.2*

*Input parameter: distribution of the total incidence of progressive and regressive preinvasive cervical neoplasia by age group.*

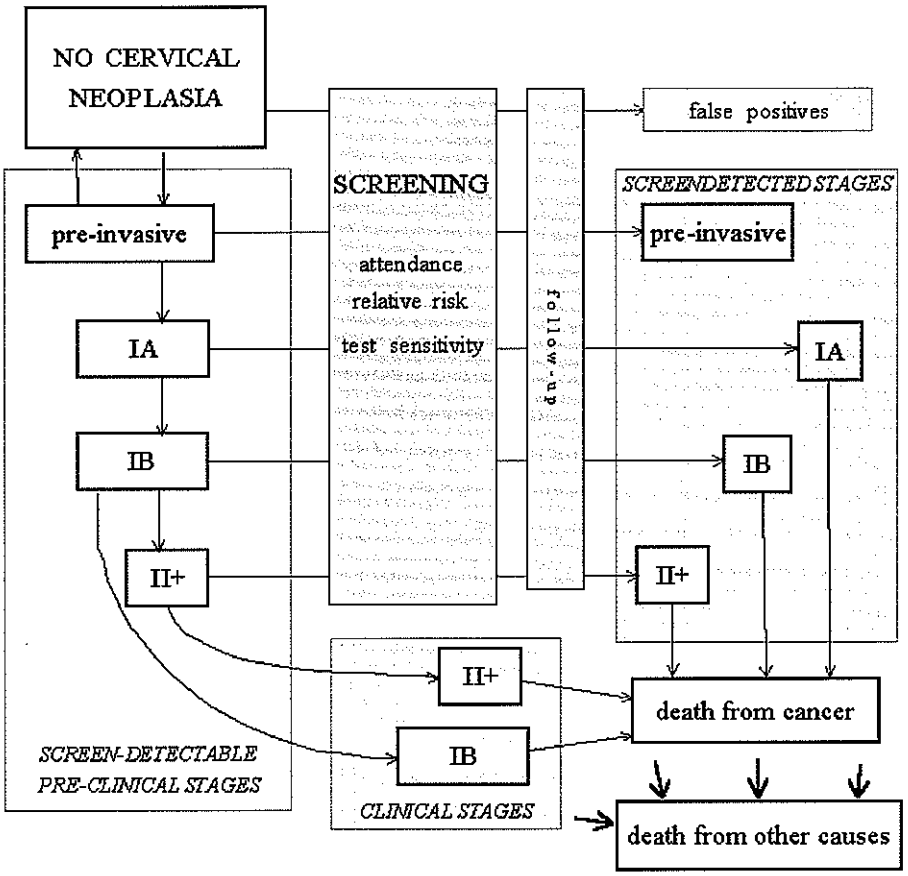
| Age group, y | Progressive | Regressive |
|---|---|---|
| 15-19 | 4.6% | 17.7% |
| 20-24 | 6.8% | 10.9% |
| 25-29 | 9.0% | 10.9% |
| 30-34 | 12.6% | 9.8% |
| 35-39 | 10.9% | 5.6% |
| 40-54 | 14.0% | 16.8% |
| 55-100 | 42.1% | 28.3% |
| | 100% | 100% |

The age distribution of the incidence of progressive preinvasive neoplasia was determined with the use of the age components of the mortality derived from the age-period cohort analysis (*28*), the distribution of the duration of the preclinical stages of the disease, and the duration between clinical diagnosis and death combined with the age-specific lethality from cervical cancer. The age-distribution of the incidence of regressive pre-invasive neoplasia was calibrated by calculating the difference between observed cervical intraepithelial neoplasia (CIN) detection rates in the Netherlands (derived from the Dutch Network and National Database for Pathology [PALGA] data for the year 1992) and the detection rates of progressive CIN predicted by MISCAN. The resulting age-distributions of pre-invasive incidence of regressive and progressive disease, respectively, are shown in Table 5.2.

Preclinical disease is subdivided into four sequential stages (Figure 5.1): preinvasive (corresponding to CIN; the stage in which the disease is not yet invasive and spontaneous regression may occur) and the three preclinical stages, i.e., the stages in which the disease has become invasive but is not yet detected (International Federation of Gynecology and Obstetrics definitions IA, IB and II+; (*32*)). A Weibull distribution was used to assume variation between women in the duration of the different preinvasive and preclinical disease stages (*31*). The mean duration and the standard deviation of the different stages were 11.8 ± 2.2 years for preinvasive stage CIN, 2.0 ± 0.9 years for preclinical invasive stage IA, 1.9 ± 0.9 years for preclinical invasive stage IB and II+ combined. Progressive and regressive preinvasive stages were assumed to have the same duration distribution. The mean duration and the standard deviation of the preinvasive stages were estimated from British Columbia screening data that used one combined CIN stage (*31*). Regressive lesions never become invasive and will return to normal (without evidence of cervical neoplasia) after the preinvasive stage. After progressive preinvasive lesions become macroinvasive (stage IB), some will be clinically diagnosed, whereas others will progress to stage II+ before any symptoms develop. The mean duration of the preclinical invasive stage was based on the ratio of the prevalence to clinical incidence before screening began in British Columbia (*31,33*) and in a Dutch pilot study (*34*). In the Dutch pilot study, 54% of the invasive cancers detected at the prevalent screen (i.e., the first

*Figure 5.1*

*Schematic representation of the disease model for cervical cancer illustrating preclinical stages and screen-detected stages.*




screening in a previously unscreened population) were diagnosed with stage IA disease; this indicates that the duration of preclinical stage IA compared with preclinical stages IB and II+ is about the same. For our analyses, we assumed (on the basis of data from Dutch hospital registries (35) and the Norwegian cancer registry (3)) that the proportion of clinical cancers that are diagnosed in stage IB decreased linearly from 58% at age 30 years to 26% at age 70 years.

## Model Specifications: Screening and treatment

The simulated screening policies were assumed to start in 1993 and to continue for 27 years until 2020. Screening practices before 1993, however, will influence the effectiveness of the screening program after 1993; therefore, this practice has been included in the simulation of the Dutch situation.

***Table 5.3***

*Input parameter: probability of long-term survival after a clinical diagnosis of cervical cancer stage IB or II+ by age group.*

| Age, y | Stage IB | Stage II+ |
|---|---|---|
| <25 | 0.699 | 0.200 |
| 30 | 0.812 | 0.500 |
| 50 | 0.812 | 0.500 |
| >65 | 0.624 | 0.000 |

Information on the screening activities before 1993 was obtained from survey data (*4,28,36*). The attendance rate from 1993 onwards was assumed to be 80% until age 50 years, and to decrease by 0.5% per year thereafter. Because we assumed that 10% of the population will never attend the screening program, we calculated a probability of 88.9% for the potential attenders, which constituted 90% of the population, to actually respond to a scheduled screening examination. After age 50 years, the attendance rate was assumed to decrease by 0.5% per year. This is in accordance with the percentage of women in The Netherlands who had a Pap smear from 1990 through 1994.

The sensitivity of the Pap smear for different disease stages is 80% for preinvasive CIN (*31*), 85% for preclinical invasive stage IA and IB, and 90% for preclinical invasive stage II+.

False-positive test results indicate the specificity of the Pap smear. We assumed that 0.06% of screening attenders were referred for a colposcopy and a biopsy after which no cervical neoplasia was found and that 6.2% of screening attenders will, on average, have 1.8 repeat smears because of borderline test results after their primary smear before they return to the regular screening schedule (PALGA 1992).

For the simulation model, the percentage of women surviving after a clinical diagnosis of cancer was assumed to be age dependent and stage dependent on the basis of Dutch incidence and mortality figures from the prescreening period in The Netherlands (*28*). Cancers clinically detected in stage IB have a more favorable prognosis than cancers detected in stage II+, and women aged 30 - 50 years who are diagnosed with stage II+ disease have a higher probability of surviving than women diagnosed with the same disease when younger than 30 years or older than 50 years (see Table 5.3).

Screen-detected pre-invasive lesions were assumed to lead to a 100% cure rate. For screen-detected invasive cancers, the survival was modelled as a reduction in the risk of dying of cervical cancer compared with that of dying of clinically diagnosed cancer. This reduction was assumed to be 80% for screen-detected stage IA disease, a percentage that was found to reproduce the reported 97% 5-year relative survival for this stage (*37*). For screen-detected stage II+ disease, the reduction was fixed at 20%, resulting from a comparison of the stage distribution within stage II+ in a

*Table 5.4*

*Input parameter: costs of screening, diagnosis, treatment, and palliative care for cervical cancer.*

| Cost item | Costs (US $) |
|---|---|
| Fixed costs of a screening program (annually) | 3,300,000 |
| Variable costs of a screening program | |
| - Invitation costs (per invitation) | 1.0 |
| - Screening costs (per screening) | 26-32[a] |
| Repeat smear | 49 |
| Diagnosis and treatment (2-4) | |
| - no neoplasia found | 485[b] |
| - CIN | 1,950 |
| - IA | 5,315 |
| - IB | 11,265 |
| - Screendetected II+ | 10,620[c] |
| - Clinically diagnosed II+ | 9,705[c] |
| Treatment and palliative care for advanced cervical cancer (9) | |
| - < 50 years | 30,800 |
| - 50-70 years | 21,955 |
| - > 70 years | 9,345 |

[a] The screening costs depend on the number of smears taken in the national program, and decrease from US$ 32 per smear if 80,000 smears were taken annually to US$ 26 for per smear if 2,500,000 smears were taken annually.

[b] Diagnostic and treatment procedures include cytology, colposcopy and biopsy modalities (cryocoagulation or LETZ and 'local treatment'). Also, 10% of the women referred without subsequent diagnosis of cervical abnormalities will receive a conization (9).

[c] The difference in costs between screen-detected and clinically diagnosed stage II+ cervical cancer relates to the less favorable stage-distribution in clinically diagnosed cases, which generate a lower number of radical hysterectomies and a higher number of radiotherapeutic treatments. The costs of the latter treatment are about 35% lower.

period with little screening (1970-1975) to the stagedistribution within screen detected II+ cases (*34*). For screen-detected stage IB disease, the reduction was assumed to be 40%, an intermediate value.

## Model Specifications: Costs

The costs of a screening program (see Table 5.4) were divided into fixed costs and variable costs. Fixed costs are associated with coordinating and evaluating a cervical cancer-screening program. Variable costs are divided into invitation costs and screening costs. Screening costs include time and travel costs for the woman, costs of smear taking, costs of cytologic evaluation, costs of registration in the PALGA, and the costs for 5.3% (estimated from PALGA data for the year 1992) of the smears that are repeated because of inadequate smears or smears without endocervical cells (9,38).

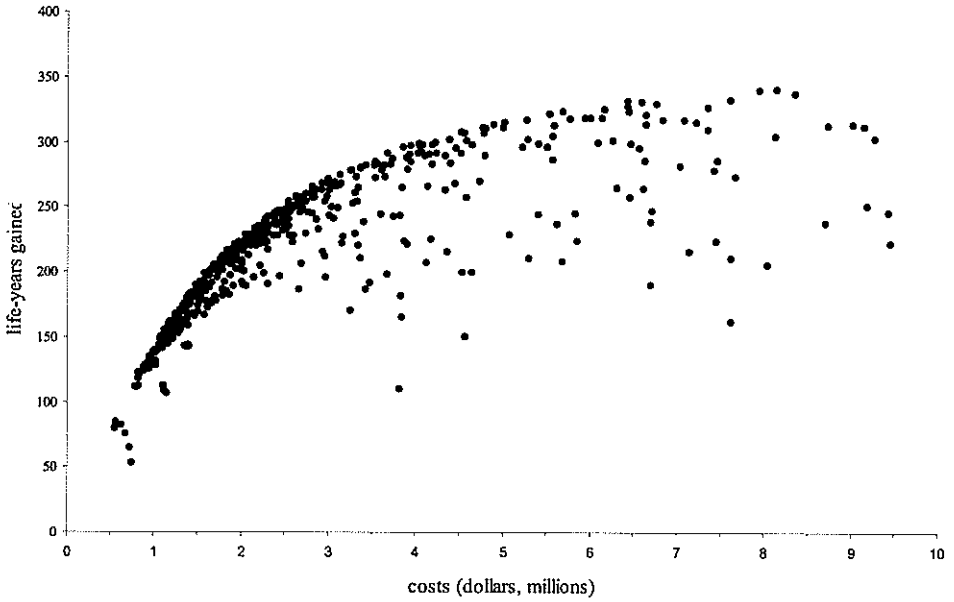
The costs of diagnostic and treatment procedures for the different disease stages, and the costs of treatment and palliative care for advanced cervical cancer in the last phase before dying of cervical cancer were derived from cost studies in The Netherlands (*4,9,28,39,40*).

## Cost-effectiveness analysis

In this study, MISCAN was used to predict costs and effects for organized screening programs for a 27-year period. We assumed a hypothetic situation in which the organized program is the only

*Figure 5.2*

*Costs (in millions of U.S. dollars) and effects (life-years gained) per 1,000,000 women in a simulated general population per year for nearly 500 simulated screening policies (each represented by a dot) differing with respect to age-range, screening interval and number of scheduled screening examinations in a lifetime. The analysis considered a 3% discount rate to convert future costs and health effects to their present values.*



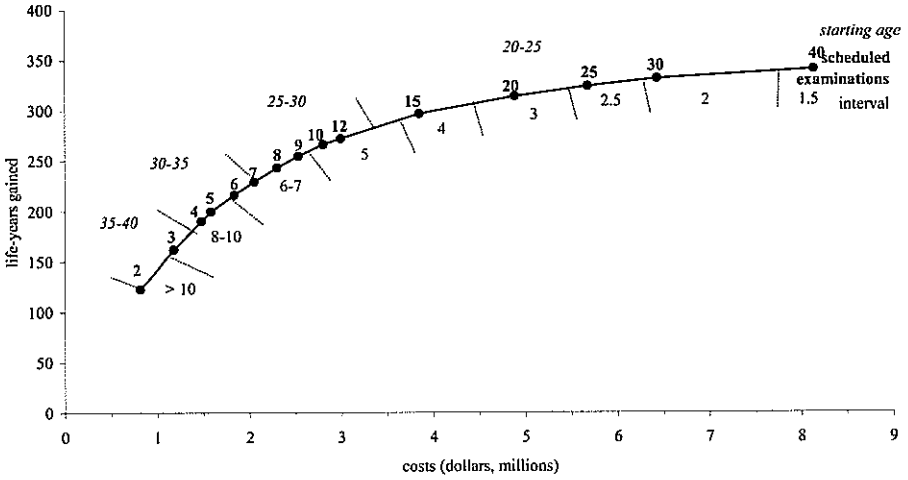

screening program and in which no opportunistic screening (i.e., spontaneous screening for other than medical reasons) occurs. The simulated effects are accounted for until all simulated women who could have benefited from the program have died. The costs are presented in U.S. dollars (1 U.S. dollar = 2 Dutch florins). The effects are presented in days of life gained per woman per year of the screening program. The cost-effectiveness calculations are conducted from the societal perspective.

To identify efficient screening policies, we compared the simulated costs and life-years gained from each policy. A policy was considered to be efficient when there was no alternative policy resulting in more life-years gained for the same or lower costs (simple dominance) and when there was no combination of two other screening policies that gained more life-years for the same costs (extended dominance) (*41,42*).

The effects per woman during her lifetime were derived by multiplying the number of days gained per woman per year of the screening program by the average life expectancy of a woman in the Netherlands, which is 80 years (*43*). In the calculation of incremental and/or average cost-effectiveness ratios, both costs and effects were discounted at a rate of 3% to convert future costs and health effects to their present value (i.e., dollars expended or health effects experienced *n*

*Figure 5.3*

*Schematic representation of the simulated efficient frontier showing the location of optimal starting ages, number of scheduled examinations, and screening intervals. Costs (in millions of U.S. dollars) and effects (life-years gained) are per 1,000,000 women in the simulated general population per year of the screening program. The analysis considered a 3% discount rate to convert future costs and health effects to their present values. The starting ages are shown as a range in years. The interval between scheduled Pap smears is in years. Broken lines represent the boundaries between the age ranges and scheduled intervals.*




years in the future are discounted a factor of $1/(1.03)^n$ *(42)*, as recommended by the Panel on Cost Effectiveness in Health and Medicine.

Because of the nature of microsimulations, estimates for costs and effects are affected by random fluctuation. We calculated this fluctuation to be less than 2% of the estimated value of the cost-effectiveness ratio and up to 35% for the incremental cost-effectiveness ratio. Therefore, to reduce the influence of random fluctuation, the incremental cost-effectiveness ratio was estimated by enlarging the simulated population 10 times to 400 million women.

## Sensitivity analysis

A one-way sensitivity analysis was performed on background incidence, attendance, sensitivity (proportion of false-negative tests) and specificity of the screening test, and costs (fixed costs, screening costs, and assessment and treatment costs). The background incidence and fixed costs, screening costs, and treatment costs were halved and doubled to obtain the low and high values, respectively, used in the sensitivity analysis. To determine attendance, the lack of attendance values were halved and doubled to obtain the high estimates and low estimates, respectively. To determine the sensitivity of the screening test, we halved and doubled the proportion of false-negatives for all stages to obtain the high estimates and low estimates, respectively. To determine the specificity of the screening test, we obtained the high values by halving the percentage of

*Table 5.5*

*Efficient policies, estimated by MISCAN, characterized by the number of scheduled examinations, screening interval and age range and expressed as the cost and effects per woman per year of the screening program, the effects of the program per woman during her lifetime, the CER, and the ICER[a]*

| No. of scheduled exams | Interval between exams, y | Age range, y | Costs U.S. dollars per woman Per year of screening (no discounting)[c] | Effects Days gained per woman per year of screening (no discounting) | Effects Days gained per woman During lifetime (no discounting) | CER Costs (U.S. dollars) per life-year gained (3% discounting) | ICER[b,d] Costs (U.S. dollars) per life-year gained (3% discounting) |
|---|---|---|---|---|---|---|---|
| 2 | 12 | 40-52 | 0.99 | 0.14 | 11.6 | 6,700 | 6,700 |
| 3 | 9 | 35-53 | 1.42 | 0.19 | 15.5 | 7,300 | 9,100 |
| 4 | 8 | 32-56 | 1.79 | 0.23 | 18.3 | 7,700 | 10,200 |
| 5 | 9 | 32-68 | 1.94 | 0.23 | 18.7 | 7,900 | 11,800 |
| 6 | 7 | 32-67 | 2.27 | 0.25 | 20.3 | 8,500 | 15,200 |
| 7 | 6 | 32-68 | 2.55 | 0.27 | 21.4 | 8,900 | 17,300 |
| 8 | 7 | 27-76 | 2.86 | 0.29 | 22.9 | 9,400 | 16,800 |
| 9 | 6 | 27-75 | 3.17 | 0.30 | 23.9 | 9,900 | 20,800 |
| 10 | 5 | 27-72 | 3.54 | 0.31 | 25.2 | 10,600 | 26,300 |
| 12 | 5 | 27-82 | 3.79 | 0.32 | 25.4 | 11,000 | 32,300 |
| 15 | 4 | 22-78 | 4.93 | 0.35 | 28.1 | 13,000 | 35,000 |
| 20 | 3 | 22-79 | 6.35 | 0.37 | 29.6 | 15,500 | 55,700 |
| 25 | 2½ | 20-80 | 7.43 | 0.38 | 30.8 | 17,500 | 79,500 |
| 30 | 2 | 22-80 | 8.49 | 0.39 | 31.4 | 19,400 | 119,000 |
| 40 | 1½ | 20-79 | 10.88 | 0.41 | 32.4 | 23,900 | 173,700 |

[a] CER = cost-effectiveness ratio; ICER = incremental cost-effectiveness ratio; MISCAN = Microsimulation screening analysis (7,8).

[b] CER was estimated using a fictitious population of 40 million women, ICER was estimated using an enlarged simulated population to reduce random fluctuation

[c] Discounting refers to converting future costs and health effects to their present values.

[d] The incremental cost-effectiveness ratio is calculated by dividing the difference in costs between the next less intensive efficient screening policy with the current screening policy by the difference in effects between these screening policies. To calculate the ICER for the screening policy with two scheduled examinations the costs and effects of this screening policy are compared with a situation without screening.

repeat smears because of borderline test results and the proportion of referrals for biopsy after which no cervical neoplasia was found, and the low values were obtained by doubling the baseline estimates for these parameters.

# RESULTS

## Costs and effects of screening policies

From nearly 500 screening policies, there was a broad range of combinations of predicted costs and effects as measured by life-years gained (Figure 5.2). Per 1,000,000 women of the general population the costs varied between 0.5 million and 9.5 million U.S. dollars, and the effects ranged from 50 life-years to 350 life-years gained per year of the screening program.

Next, after deleting those that were not efficient, we obtained the efficient screening policies for which no alternative policy exists that result in more life-years gained for lower costs. There were 15 efficient screening policies, and together they represented the efficient frontier (Figure 5.3). The age range of efficient screening policies increased from age 40-52 years for policies that

*Table 5.6.*

*Detailed overview of the predicted costs and effects per 1,000,000 women of the general population per year of the screening program for the efficient screening policies as estimated by MISCAN, with 5, 10, 20 and 40 scheduled examinations[a].*

| No. of scheduled examinations | 5 | 10 | 20 | 40 |
|---|---|---|---|---|
| Interval between examinations, y | 9 | 5 | 3 | 1½ |
| Age range, y | 32-68 | 27-72 | 20-77 | 20-79 |
| No. of Pap smears per year of screening | 40,000 | 78,000 | 152,000 | 300,000 |
| Costs[b] (U.S. dollars, millions) | | | | |
| Screening | 1.74 | 2.94 | 5.05 | 8.66 |
| Repeat smears | 0.22 | 0.42 | 0.83 | 1.62 |
| Referred, no cervical intra-epithelial neoplasia (CIN)[c] | 0.01 | 0.02 | 0.04 | 0.09 |
| CIN[d] | 1.12 | 1.70 | 2.28 | 2.54 |
| Invasive carcinoma[cd] | -0.58 | -0.80 | -0.98 | -1.07 |
| Advanced disease[e] | -0.56 | -0.75 | -0.88 | -0.96 |
| Total costs | 1.94 | 3.53 | 6.35 | 10.87 |
| Effects | | | | |
| Prevented deaths | 26 | 35 | 41 | 45 |
| Life-years gained[f] | 640 | 862 | 1,022 | 1,110 |

[a] MISCAN = Microsimulation screening analysis (7,8). The values are not discounted because no time preference is established.
[b] Costs determined relative to a situation without screening.
[c] Costs of follow-up and treatment.
[d] If all invasive cases could be prevented a total U.S. dollar 1.54 million saved, as predicted by MISCAN.
[e] Costs of treatment for recurrence and palliative care in women who die from cervical cancer. Maximum savings on costs of advanced disease total U.S. dollar 1.38 million.
[f] The days gained per woman per year of screening can be calculated by multiplying the number of life years gained by 365/1,000,000.

recommend two examinations during a woman's lifetime to age 20-80 years for those that recommend more than 20 examinations. In general, a more intensive screening policy was one that recommended that screening start at a younger age, end at an older age, and have a shorter interval between examinations (Table 5.5). For the efficient policies, the effects of the total screening program on life expectancy ranged from 11.6 days for those that recommend two scheduled examinations during a woman's lifetime to 32.4 days for those that recommend 40 scheduled examinations. We estimated that total elimination of cervical cancer would yield a gain in life expectancy of 46 days.

According to the law of diminishing returns, if the number of scheduled examinations in a screening program were increased, the increase in the number of life-years gained would slow down (Table 5.5). When a detailed assessment of the costs of a screening program was compared with that of the costs of no screening, screening costs were in excess of the costs of diagnostic testing and treatment of preinvasive disease detected by screening combined and were only partially compensated for by the savings incurred from preventing invasive carcinoma and advanced disease (Table 5.6). Because the costs of coordinating and evaluating the screening program were assumed to be independent of the number of scheduled examinations, and because a scale effect (i.e., the costs per smear are lower if more smears are performed) was assumed for the costs per smear, the screening costs increased less than proportionally with the number of scheduled examinations. Moreover, when moving toward more intensive policies, the incremental cost-effectiveness ratio increased because the incremental effects rapidly diminished (Table 5.5). For a

*Table 5.7.*

*Sensitivity analysis: Incremental cost-effectiveness ratios, expressed as U.S. dollars per life-year gained, with 3% discounting to convert future costs and health effects to present values[a].*

| No. of scheduled Pap smears | 5 | | 10 | | 20 | | 40 | |
|---|---|---|---|---|---|---|---|---|
| Interval, y | 9 | | 5 | | 3 | | 1½ | |
| Age range, y | 32-68 | | 27-72 | | 20-77 | | 20-79 | |
| | low[b] | high | low | high | low | high | low | high |
| Background incidence | 24,400 | 5,100 | 49,800 | 11,400 | 124,300 | 26,800 | 331,700 | 83,600 |
| Attendance | 14,500 | 10,800 | 16,200 | 33,300 | 33,300 | 81,500 | 80,100 | 315,400 |
| Sensitivity of screening test | 19,200 | 9,300 | 19,200 | 32,000 | 41,700 | 66,500 | 117,300 | 191,000 |
| Specificity of screening test | 13,500 | 11,000 | 30,100 | 24,400 | 64,300 | 51,400 | 206,300 | 157,400 |
| Screening costs | 7,000 | 21,400 | 16,100 | 46,700 | 34,800 | 97,600 | 107,200 | 306,600 |
| Treatment costs CIN& false positives[c] | 10,200 | 15,100 | 23,700 | 31,500 | 51,400 | 64,400 | 167,300 | 186,600 |
| Treatment costs invasive cancers | 13,100 | 9,200 | 27,700 | 23,400 | 57,200 | 52,700 | 175,200 | 170,600 |
| Baseline[d] | 11,800 | | 26,300 | | 55,700 | | 173,700 | |

[a] Sensitivity analysis is defined as investigations that isolate key parameters involved in the cost-effectiveness analysis that indicate the degree of influence each key parameter has on the outcome of the analyses. Incremental cost-effectiveness ratios are calculated by comparing the costs and effects of a screening policy with five scheduled examinations to one with four scheduled examinations, 10 scheduled examinations to one with nine scheduled examinations, 20 scheduled examinations to one with 15 scheduled examinations, 40 scheduled examinations to one with 30 scheduled examinations.

[b] The background incidence (defined as the incidence in a situation without screening) and fixed costs, screening costs and treatment costs were halved and doubled to obtain the low values and high values, respectively. To determine attendance, the lack of attendance values were halved and doubled to obtain the high estimates and low estimates, respectively. To determine the sensitivity of the screening test, the proportion of false-negative results for all stages was halved and doubled to obtain the high estimates and low estimates, respectively. The high values for the specificity of the screening test were obtained by halving the percentage of repeat smears because of borderline test results and the proportion of attending women that are referred for a colposcopy and a biopsy after which no cervical neoplasia is found. The low values for the specificity of the screening test were obtained by doubling the baseline estimates for these parameters.

[c] CIN = cervical interepithelial neoplasia

[d] The baseline estimates for the background incidence for women willing to participate in screening were 0.0229 for women born 1889-1918; 0.0235 for those born 1919-1928; 0.128 for those born 1929-1938; 0.0106 for those born 1939-1948; and 0.0148 for those born 1949-2000. For attendance the baseline estimate was 80%. The sensitivity was assumed to be 80% for preinvasive CIN, 85% for preclinical invasive stage IA and IB, and 90% for preclinical invasive stage II+ at baseline. The specificity assumed that 0.06% of attending women are referred for a colposcopy and a biopsy after which no cervical neoplasia is found, and that 6.2% of attending women will, on average, have 1.8 repeat smears because of borderline test results after their primary smear before they return to the regular screening schedule. The baseline estimates for the screening costs decrease from US$32 per smear if 80,000 smears were taken annually to US$26 per smear if 2,500,000 smears were taken anually. The invitation costs were US$1 per invitation at baseline. The baseline estimates for the costs of diagnosis and treatment of CIN and false positive test results are US$1950 and US$485, respectively. The costs for diagnosis and treatment of invasive cancers at baseline are US$5315, US$11265, US$10620 and US$9705 for diagnosis and treatment of invasive cancers stage IA, IB, screen detected II+ and clinically diagnosed II+, respectively. The costs for treatment (including palliative care) for advanced cervical cancer are age-dependent; US$30800 for women aged less than 50 years, US$21955 for women aged between 50 and 70 years, and US$9345 for women aged more than 70 years at baseline.

policy maker, if the decision regarding a policy depends only on a maximal allowed value or threshold value for the incremental costs per life-year gained, then for reference values of $15,000, $30,000 and $60,000, screening policies with five, 10 and 20 scheduled examinations, respectively, and screening intervals of 9, 5 and 3 years, respectively, are optimal.

## Sensitivity analysis

Differences in demographic, epidemiologic, and screening characteristics, such as background incidence, attendance, sensitivity, and specificity of the screening test, and cost, may lead to

different choices in efficient screening policies. The influences of these differences were investigated in a sensitivity analysis (Table 5.7). A higher background incidence, i.e., the incidence of invasive cancer in the hypothetic situation where there has never been screening, led to higher effects of a screening policy because the effects were proportional to the incidence of cervical cancer (see Table 5.7). This resulted in a more favorable incremental cost-effectiveness ratio, and consequently more intensive screening policies were feasible given a threshold value for the incremental cost-effectiveness ratio.

Differences in either the percentage of women who will attend a recommend screening examination (screening attendance) or sensitivity of the screening test will not only affect the choice of the number of screening examinations to be offered per woman but will also affect the choice of the age range and time interval between the scheduled examinations (*44*). Higher attendance and/or sensitivity will make longer intervals between screenings and, simultaneously, broader age ranges more favorable in terms of cost-effectiveness because the role for a subsequent screening to detect abnormalities previously missed would be less important (Table 5.7).

A lower specificity will increase the incremental costs per life-year gained and, therefore, lead to a lower number of scheduled examinations that would be offered per woman to achieve the same incremental cost per life-year gained compared with the baseline situation in which Dutch characteristics are incorporated. The choice of the number of scheduled examinations depends on the costs of medical procedures (including screening itself) that are generated or prevented by screening. If the costs of Pap smears, assessment, and treatment of false-positive results and CIN generated by screening are higher than those assumed in the baseline situation, then the cost-effectiveness ratio of screening will be unfavorably influenced. In contrast, higher costs for treatment of invasive cancers and advanced disease, some of which are prevented by screening, will lower the incremental and/or average cost-effectiveness ratio. The fixed costs for coordinating and evaluating a cervical screening program do not influence the incremental costs per life-year gained. However, the average costs per life-year gained will increase if the fixed costs are higher.

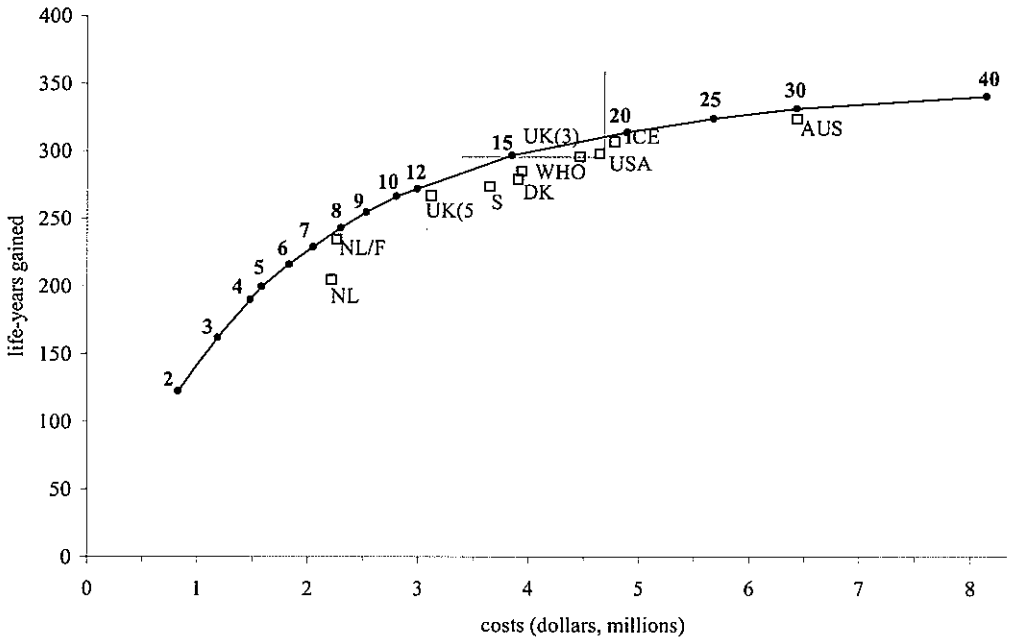
## International comparison

We next compared the screening policies from countries with cervical cancer-screening programs or national guidelines, with the assumption that their demographic, epidemiologic, screening and treatment characteristics were similar to those in The Netherlands. As shown in Figure 5.4, several of the screening policies are remarkably close to the efficient frontier. However, for several screening policies, such as those from Sweden, Denmark, the U.K. (16 scheduled examinations), the United States, and Australia, alternative policies could be recommended to reduce costs for the same amount of life-years gained or to improve effectiveness while keeping the costs the same. These alternative policies are situated in the upper-left quadrant of the marking for a screening policy of a country in Figure 5.4. For example, the area in which more cost-effective screening policies are situated for the US is identified in Figure 5.4 by a broken line. It can be seen in Figure 5.4 that the

***Figure 5.4.***

*Comparison of the costs (in millions of U.S. dollars) and effects (life-years gained) per 1,000,000 women in the general population per year of screening for screening policies used in countries with a cervical screening program or program recommended in national guidelines and the simulated efficient frontier, with 3% discounting. Discounting refers to converting future costs and health effects to their present values. The solid line represents the simulated efficient frontier. The number of scheduled lifetime Pap smears are identified with solid filled circles.*

*NL/F = The Netherlands from 1996/ Finland (1), (7/5/30-60, where 7 is the number of scheduled Pap smears, 5 is the screening interval in years, and 30-60 is the age range); NL = The Netherlands before 1996 (7/3/35-53); DK = Denmark (5), (13/3/23-59); S = Sweden (11), (12/25/25-58); ICE = Iceland (13), (19/2½/25-70); UK(3) = United Kingdom (14), (16/3/20-65); UK(5) = United Kingdom (14), (10/5/20-65); US = United States (20), (17/3/18-66); AUS = Australia (21), (27/2/18-70); EUR = WHO/Eurogin (23), (14/3/25-64).*

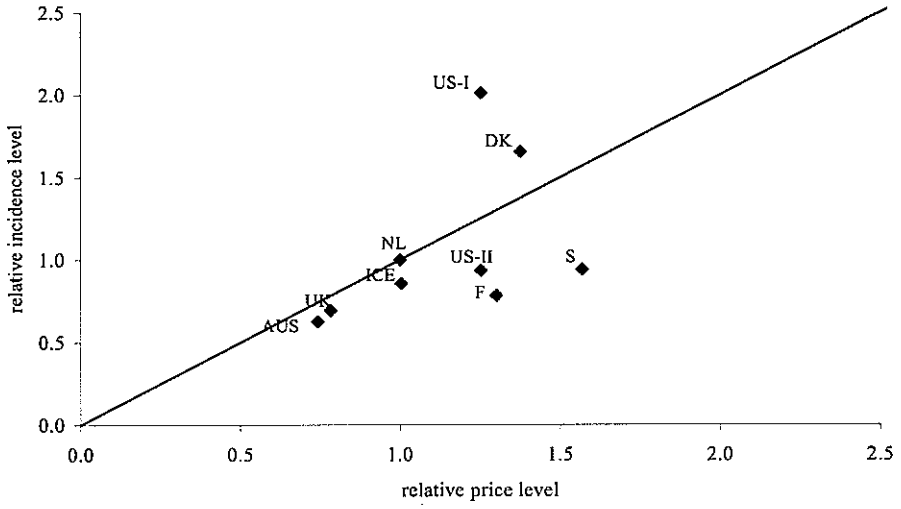



policy for screening every 4 years between age 22 and 78 years with 15 examinations has the same effects for much lower costs (yearly almost $1 million less) than the recommendations issued by the U.S. Preventive Task Force (20) for screening every 3 years between ages 18 and 66 years with 17 scheduled examinations. If, however, a more intensive policy is committed to (the U.S. Preventive Task Force policy is conservative compared with recommendations from other U.S. authorities and current U.S. practice that recommends annual screening), the efficient screening policies with 20-30 examinations and an interval of 2-3 years starting after age 20 years are more cost effective than current practice.

To investigate whether the wide diversity in screening recommendations and official policies among countries originate from differences in the epidemiology of cervical cancer or price level, we compared the incidence of cervical cancer and price levels among countries. If the background incidence is higher in a country than it is in The Netherlands, then the effects of a

*Figure 5.5a*

*Comparison of the incidence of cervical cancer and health care-specific price levels of seven countries (Australia, Denmark, Finland, Iceland, Sweden, the United States, and the United Kingdom) relative to the Netherlands. The relative incidence level is calculated by dividing the incidence in a situation before screening by the Dutch incidence in a situation without screening. The relative price level is calculated by dividing the price level of a country by the Dutch price level. Price levels are calculated by dividing health-care specific price levels by the exchange rates.*

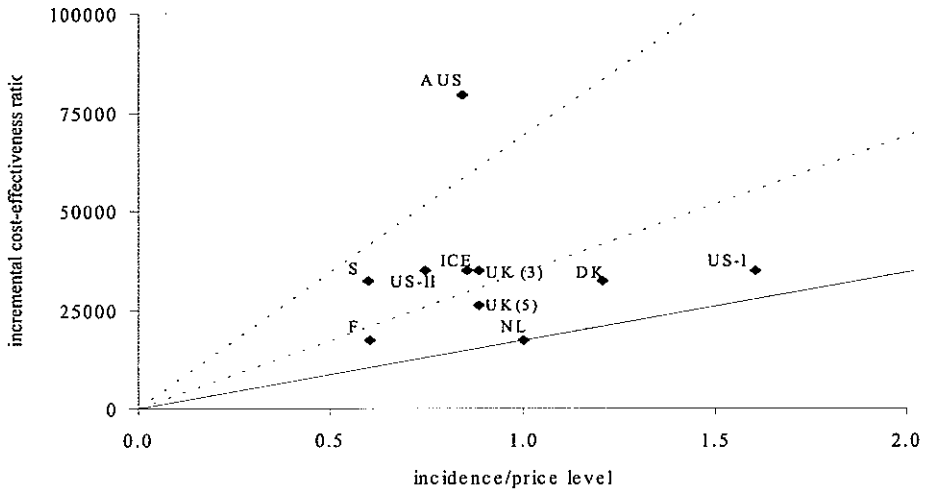



screening policy would be proportionally higher than those calculated in this analysis, whereas the total costs would stay at the same level. If the price level in a country is lower than it is in The Netherlands, then the total costs of a screening policy would be proportionaly lower than those calculated in this analysis. The differences in background incidence of cervical cancer and/or price level will result in more favorable cost-effectiveness estimates than those calculated for the Netherlands (and vice versa, if the background incidence is lower and/or the costs are higher), which may lead to the choice of an efficient screening policy with a higher number of scheduled examinations despite having the same threshold value of the incremental and/or average cost-effectiveness ratio. Differences in incidence and/or price level may, therefore, explain the diversity in the number of scheduled examinations among different countries.

We obtained the background incidence of cervical cancer for each country (*2,45,46*). We calculated the price levels for each country by dividing the health-care specific purchasing power parities (*47*) which adjusts the exchange rates for different countries to the health care specific price levels, by the current exchange rates. By comparing the incidence and the price levels to those of the Netherlands, we obtained relative incidence and relative price levels, respectively. These values were then plotted (Figure 5.5, A). The solid line represents the situation in which the relative incidence is equal to the relative price level. For the countries that fell above the line (Denmark and the United States, assuming a high background incidence to be representative), the cost-effectiveness estimates of a policy will be more favorable than those estimates in The

*Figure 5.5b*

*Ratios of incidence to price level relative to The Netherlands and incremental cost-effectiveness ratios of cervical screening policies used or recommended in Australia, Denmark, Finland, Iceland, Sweden, the United States and the United Kingdom. The solid line represents all situations with the same incremental cost-effectiveness ratio as The Netherlands, after correction for incidence and price level. The dotted lines indicate all situations with two (lower line) or four (upper line) times the incremental cost-effectiveness ratio of The Netherlands. For the incremental cost-effectiveness ratios of screening policies used or recommended, comparisons are made between the efficient screening policies with the same number of scheduled examinations.*
*AUS = Australia; DK= Denmark; F = Finland; ICE = Iceland; S = Sweden; US-I = United States, Second National Cancer Survey 1947; US-II = US- Connecticut Tumor Registry; UK = United Kingdom; UK (3) = UK screening every 3 years; UK (5) = UK screening every 5 years.*




Netherlands. Consequently, more intensive screening policies will stay below a certain threshold value of the incremental and/or average cost-effectiveness ratio. For the countries that fell below the line (Australia, the U.K., Iceland, Finland, Sweden, and the United States, assuming a low background incidence), having a relatively low incidence and/or a higher price level, the cost-effectiveness estimates will be less favorable than those in The Netherlands.

We next plotted the combination of incidence and price level for the different countries against the incremental cost-effectiveness ratios for each of the respective screening policies (Figure 5.5, B). If the differences in intensity of screening among countries can be explained by differences in incidence and/or price level, all screening policies would be situated on a straight line through the origin when plotted. However, it can be concluded from Figure 5.5, B, that the diversity in screening policies, which range from recommending seven to 27 scheduled examinations, cannot be explained by differences in incidence level or price level.

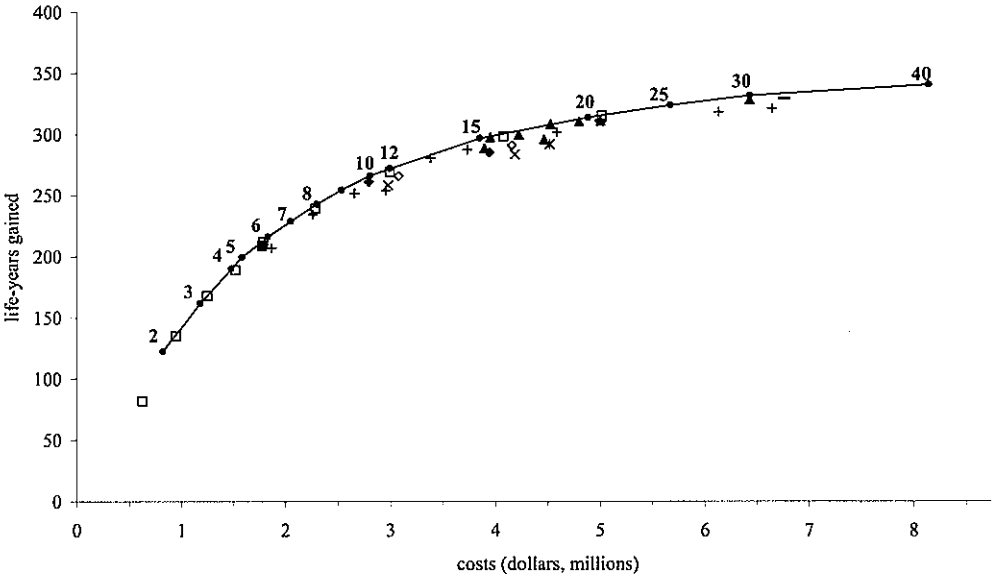
Finally, we calculated the costs and effects of screening policies evaluated in other cost-effectiveness analyses (6,10,12,16,19,22,24,25) with our MISCAN model (Figure 5.6). Most policies appeared to be close to our efficient frontier (Figure 5.6). The screening policies with intervals between screening examinations varying by age that were found to be efficient by Gustafsson and Adami (24) are close to our efficient frontier, as were those with fixed intervals (screening every

***Figure 5.6.***

*Comparison of the costs (in millions of U.S.dollars) and effects (life-years gained) per 1,000,000 million women in a simulated general population per year of screening for screening policies considered in other cost-effectiveness analyses. Comparison is made the simulated efficient frontier with 3% discounting. Discounting refers to converting future costs and health effects to their present values.*

*- = Hristova and Hakama (6)(7/5/30-60, where 7 is the number of scheduled Pap smears, 5 is the screening interval in years, and 30-60 is the age range); + = Gyrd-Hansen et al. (10) (5/5/30-50; 7/5/30-60; 8/5/25-60; 8/4/30-58; 9/4/25-57; 12/4/25-69; 13/4/20-68; 13/4/20-68; 18/3/20-68; 25/2/20-68; 28/2/15-69);▲ = Eddy (12) (16/3/29-74; 15/4/20-76; 17/3/26-74; 16/3/20-65; 18/3/23-74; 19/3/20-74; 20/3/17-74; 28/2/20-74); x = Waugh and Robinson (19) (14/3/20-59; 9/5/20-60);ж = Sherlaw-Johnson (22) (15/3/18-63); ♦ = International Agency for Research on Cancer (IARC) (16) (9/5/25-65; 14/3/25-64); ◊ = IARC (16) varying intervals indicated by * and all screening ages (10/*/25,26,30,36,40,45,50,55,60,65; 15/*/25,26,29,32,35,38,41,44,47,50,53,56,59,62,65); ■ = Gustafsson and Adami (24) (5/7/30-58); □ = Gustafsson and Adami (24) varying intervals indicated by * (1/*/37; 2/*/34,45; 3:32,40; 3/*/32,40,51; 4/*/31,37,44,57; 5/*/30,35,43,50,60; 7/*/28,33,37,43,49,57,66; 10/*/26,30,33,37,42,46,51,57,63,70; 15/*/23,27,30,32,35,38,41,44,48,51,55, 60,64,69,74;*

*20/*/22,25,28,30,32,34,36,38,40,43,46,49,52,55,58,62,65,69,73,78; — = McCrory (25) (20/3/18-75;30/2/18-76) An upper age for screening of, respectively 76 and 75 years was assumed, although no upper age limit was mentioned in the study.*




7 years between ages 30 and 58 years). Eddy (12) investigated screening every 3 years between age 20 and 74 years, alternative ages to start screening (17, 23 or 26 years), and alternative screening intervals (every 2 or 4 years), and concluded that a minimal screening policy of every 3 years between ages 20 and 65 years was cost efficient. However, this minimal policy and screening every 3 years between ages 29 and 74 years are less than efficient according to our model (Figure 5.6). For this number of scheduled examinations (16 in both cases) an interval of screening every 4 years would be more efficient (see Figure 5.3). McCrory et al. (25) calculated the costs and effects for three screening policies based on conventional Pap smears that started at age 18 years and had a screening interval of every 1, 2, or 3 years. The screening policies with a screening interval of every 2 or 3 years, which we included in our analysis, appeared to be close to the efficient frontier. The screening policy with a 1-year interval was omitted, as no screening policies with more than 40

scheduled examinations were included in our analyses. Although the screening policy also may have appeared to be quite close to the efficient frontier, the incremental and/or average cost-effectiveness will be far outside the range that we considered to be acceptable. The screening policies considered by Waugh (19) (screening every 3 years between age 20 and 59 years and screening every 5 years between age 20 and 60 years) and Sherlaw-Johnson (22) (screening every 3 years between age 18 and 64 years), and some of the screening policies considered optimal by Gyrd-Hansen (10) (varying from five scheduled examinations between age 30 and 50 years to 28 scheduled examinations between age 25 and 69 years) were not efficient according to our model.

## DISCUSSION

The results show that efficient screening policies for cervical cancer can be characterized by an average screening age of about 50 years. This means that an intensive screening policy would begin at a younger age, end at an older age, and have a shorter interval between the scheduled examinations.

With the use of the MISCAN program, we determined a predicted gain in life expectancy of 46 days if cervical cancer is eliminated. This gain is small compared with the predicted increase in life expectancy of other diseases, such as the approximately 1.5 years if coronary heart disease were eliminated in women (48), and is directly related to the relatively low mortality rate from cervical cancer. However, it is more relevant to compare the gain in life expectancies with different health interventions because most health interventions can only partly eliminate the disease. Also, elimination or near elimination of cervical cancer through screening does not seem possible considering the persistent level of nonattendance of women at high risk for cervical cancer. Because we based our model on the Dutch cervical cancer-screening figures, we assumed an attendance of about 80%. Half of the remaining 20% are persistent nonattenders. The nonattenders were assumed to be at high risk for cervical cancer and accounted for 25% of the cervical cancer mortality, putting the upper limit of attainable gain in life expectancy by cervical cancer screening at 75% of 46 days or a total of 34 days.

Our predictions show that the efficient screening policies that range from two to 40 scheduled examinations result in a gain in life expectancy from 12 to 32 days. Wright and Weinstein (49) reviewed gains in life expectancy from a variety of health interventions and found estimates on a gain in life expectancy of 0.8 months for women aged 50-60 years who are offered biennial mammography, and of 8 months for women aged 35 years who quit cigarette smoking. Our estimates for the effects of cervical cancer screening are at the lower side of this range. However, in addition to the effects, costs also must be considered when evaluating diverse health interventions. Cost-effectiveness ratios as estimated in this study express the trade-off between costs and effects of interventions (50).

There are several limitations associated with cost-effectiveness analyses, including random fluctuation and outcome uncertainty (42). Random fluctuation complicates the determination of the efficient screening policies because repeat estimations of costs and effects may yield different

estimates for the costs and effects that result in small differences in screening policies determined to be efficient. This was illustrated by the screening policy including seven scheduled examinations between ages 27 and 68 years, which was found to be efficient in our initial predictions but not after enlarging the simulated population.

Outcome uncertainty is related to both parameter and model uncertainty. Parameter uncertainty is the uncertainty about the true values of the input parameters, whereas model uncertainty involves the way these parameters are modeled. An example of model uncertainty is that we made the assumption that costs for coordinating and evaluating a cervical screening program were fixed and thus that these costs were independent of the number of scheduled examinations. Increasing the coordinating and evaluating costs with the number of scheduled examinations will decrease the incremental cost-effectiveness ratio for screening policies with a small number of examinations but will increase the incremental cost-effectiveness ratio for more intensive screening policies.

In addition to the study design limitations, our results would be influenced if quality-adjusted life-years gained were used instead of life-years gained to include any side-effects of the intervention. The negative side effects of screening, including those on quality of life, are largely proportional to the number of screening examinations. By contrast, the favorable effects of screening follow the law of diminishing returns. Combining the negative side effects and the favorable effects of screening in terms of quality-adjusted life-years will result in a rapid decrease in the number of incremental quality-adjusted life-years gained for screening policies with an increase in the number of examinations (*28*), and eventually any additional intensifying screening will decrease the net health effects. Uncertainty analysis and quality-of-life considerations are both subjects of ongoing research.

The present cost-effectiveness estimates are obtained for a model that aimed to be representative of cervical cancer screening in The Netherlands. Different demographic, epidemiologic, and screening characteristics led to changes in the choice of the number of Pap smears offered per woman, the choice of the age range to be screened, and the time period between the scheduled number of Pap smears. However, we found (Figure 5.5, B) that the diversity in the number of scheduled examinations in currently used or recommended screening policies, which varied from seven to 27 examinations, cannot be explained by differences in the incidence of or price level in the countries involved. A factor that may influence the age range is the age-specific incidence of invasive cervical cancer, which reflects the age-specific incidence of progressive CIN. By comparing the age-specific incidence among different populations, Gustafsson et al. (*45*) found that, in addition to differences in the level of cervical cancer incidence, there were two patterns of age-specific incidence. In the first pattern, illustrated by some European countries, including the Netherlands, the peak age-specific incidence of invasive cervical cancer occurs at a younger age and declines rapidly thereafter. In the second pattern, illustrated by the United States, New Zealand, and Asian and African countries, the peak age-specific incidence of invasive cervical cancer occurs at an older age and declines slowly thereafter. Therefore, in countries where the

initial peak in age-specific incidence occurs at an older age, there will be a shift in the estimated optimal screening starting age, moving upward, to an older age and/or to lengthening the screening interval. Thus, when considering the incidence of invasive cervical cancer in the United States and Australia, it is unclear why screening policies that have short screening intervals and that start at a young age are recommended. Possible differences among countries in the implicit threshold values of the acceptable incremental and/or average cost-effectiveness provide no plausible explanation for the diversity in screening policies. The diversity, therefore, originates from other sources, including, for example, the rationality of the recommendation process, the data and evidence used in choosing among policies, or the methods used in evaluating policies. The latter is illustrated by the fact that even though policies evaluated in other cost-effectiveness studies (*28*) were close to our efficient frontier (Figure 5.6), the estimated incremental and/or average cost-effectiveness ratios differed considerably among studies (*28*) and may, subsequently, have led to different elected screening policies.

Moreover, our model considers features that were not considered in other cost-effectiveness analyses (*28*); this may have contributed to the different cost-effectiveness estimates. First, in our model both costs and effects were discounted to the start of screening at a rate of 3 % (*42*). Second, we assumed that nonattendance is associated with an increased risk of cervical cancer (*29-31*). Third, we accounted for the fact that the population is already screened to a certain extent. Our assumption leads to a lower prevalence of preclinical disease and, consequently, to a lower base-line risk for cervical cancer at the start of the screening program. Therefore, our cost-effectiveness estimates will be less favourable.

The current cost-effectiveness analyses concern high-income countries. However, in low-income countries in Southern America, Africa and Asia, the incidence and cancer-related death from cervical cancer is much greater than the "high" incidence selected in our sensitivity analyses. Although the incidence of cervical cancer can be reduced by a Pap smear-based screening program, such a program is often not feasible in low-income countries because it requires a high degree of organization with cytologic laboratories and personnel. An alternative for Pap smear screening in developing countries may be aided visual inspection of the cervix, which has a sensitivity similar to Pap smears but a lower specificity (*51,52*). A specific cost-effectiveness analysis to investigate the possibility of a screening program based on aided visual inspection in low-income countries is warranted.

Although the present analyses are based on Pap smear screening, which is the conventional method for detection of cervical lesions in large-scale settings, there are new methods for the detection of cervical cancer; for example, screening for the presence of oncogenic variants of the human papillomavirus. The cost-effectiveness of screening for human papillomavirus is not yet known (*53*). Other developments involve new diagnostic technologies in cytopathology, such as liquid-based cytology and computer-aided imaging. In the future, these or other new developments may lead to improvements in test characteristics and/or changes in costs, which would require reconsidering the optimal screening policies. Because women who are regularly

screened at the appropriate ages already have a reduced risk of cervical cancer, the gain in cost-effectiveness of cervical cancer screening must arise from reducing the overall costs and simplifying the screening process by reducing the number of false-positive results. The great breakthrough in the latter has to come from methods that are able to distinguish progressive and regressive disease.

## REFERENCES


(1) Hakama M, Miller AB, Day NE. Screening for Cancer of the Uterine Cervix. Lyon: International Agency for Research on Cancer; 1986.
(2) Doll R, Muir C, Waterhouse J, editors. Cancer incidence in five continents. Vol I. Berlin: Springer; 1966.
(3) CRN (Cancer Registry Norway). Survival of cancer patients. Cases diagnosed in Norway 1968-1975. Oslo; 1980.
(4) de Bruin A, de Koning H, van Ballegooijen M. Pap smears and mammographies. Dutch National Health Interview Surveys 1991. Monthly Bulletin of Health Statistics 1993:12-5.
(5) Sundhedsstyrelsen. Screening for cervical cancer (in Danish). Copenhagen: Sundhedsstyrelsen; 1986.
(6) Hristova L, Hakama M. Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. Acta Oncol 1997;36:1-60.
(7) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
(8) Loeve F, Boer R, van Oortmarssen G, van Ballegooijen M, Habbema J. The MISCAN-COLON simulation model for the evaluation of colorectal cancer. Comput Biomed Res 1999;32:13-33.
(9) van Ballegooijen M, Boer R, van Oortmarssen GJ, Koopmanschap MA, Lubbe JTN, Habbema JDF. Cervical cancer screening: age- ranges and intervals. An updated cost-effectiveness analysis (in Dutch). Rotterdam: Erasmus University Rotterdam; 1993.
(10) Gyrd-Hansen D, Holund B, Andersen P. A cost-effectiveness analysis of cervical cancer screening: health policy implications. Health Policy 1995;34:35-51.
(11) Mahlck CG, Jonsson H, Lenner P. Pap smear screening and changes in cervical cancer mortality in Sweden. Int J Gynaecol Obstet 1994;44:267-72.
(12) Eddy DM. Screening For Cervical Cancer. Ann Intern Med 1990;113:214-26.
(13) Sigurdsson K, Hrafnkelsson J, Geirsson G, Gudmundsson J, Salvarsdottir A. Screening As a Prognostic Factor In Cervical Cancer: Analysis Of Survival and Prognostic Factors Based On Icelandic Population Data, 1964-1988. Gynecol Oncol 1991;43:64-70.
(14) NHS cervical screening programme. Cervical screening. A pocket guide. NHS cervial screening programme. United Kingdom; 1996.
(15) Laara E, Day NE, Hakama M. Trends In Mortality From Cervical Cancer In the Nordic Countries: Association With Organised Screening Programmes. Lancet 1987;1:1247-9.
(16) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J 1986;293:659-64.
(17) van der Graaf Y, Zielhuis G, Peer P, Vooijs G. The effectiveness of cervical cancer screening. A population based case-control study. J Clin Epidemiol 1988;41:21-6.
(18) Clarke E, Anderson T. Does screening by 'Pap' smears help prevent cervical cancer? A case-control study. Lancet 1979;2:1-4.
(19) Waugh N, Robertson A. Costs and benefits of cervical screening. II. Is it worthwhile reducing the screening interval from 5 to 3 years? Cytopathology 1996;7:241-8.
(20) US Preventive Task Force. Guide to Clinical Preventive Services. 2nd ed. Baltimore: Williams and Williams; 1996.
(21) Department of Health Housing and Community Services. Screening for the prevention of cervical cancer. Camberra: AGPS; 1991.
(22) Sherlaw-Johnson C, Gallivan S, Jenkins D, Jones MH. Cytological screening and management of abnormalities in prevention of cervical cancer: an overview with stochastic modelling. J Clin Pathol 1994;47:430-5.
(23) Monsenego J, Franco E. Cervical cancer control. General statements and guidelines: World Health Organisation/European Research Organisation on Genital Infection and Neoplasia; 1997.
(24) Gustafsson L, Adami HO. Optimization Of Cervical Cancer Screening. Cancer Causes Control 1992;3:125-36.
(25) McCrory D, Matchar D, Bastian L, Datta S, Hasselblad V, Hickey J, et al. Evaluation of Cervical Cytology. Rockville, MD: Agency for Health Care Research; 1999.

(26) CBS (Netherlands Central Bureau of Statistics). Death by cause of death, age and sex. Series A1, 1950-1992. Voorburg; 1994.
(27) SIG (Information Centre for Health Care). Hospital Diagnosis Statistics 1963-1985 (in Dutch). Utrecht; 1985.
(28) van Ballegooijen M. Effects and costs of cervical cancer screening [Thesis]. Rotterdam: Department of Public Health, Erasmus University; 1998.
(29) Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-6.
(30) 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.
(31) van Oortmarssen GJ, Habbema JD. Epidemiological Evidence For Age-Dependent Regression Of Pre-Invasive Cervical Cancer. Br J Cancer 1991;64:559-65.
(32) Changes in definitions of clinical staging for carcinoma of the cervix and ovary. Am J Obstet Gynecol 1987;156:263-4.
(33) 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.
(34) EVAC (Evaluation Committee). Population Screening For Cervical Cancer In the Netherlands. a Report By the Evaluation Committee. Int J Epidemiol 1989;18:775-81.
(35) Ketting B. Surgical treatment of invasive carcinoma of the uterine cervix. Amsterdam: University of Amsterdam; 1981.
(36) EVAC (Evaluation Committee). Population screening for cervical cancer in the pilot regions Nijmegen, Rotterdam and Utrecht. A report by the Evaluation Committee. First and second interim report (in Dutch). Leidschendam: Ministry of Welfare, Public Health and Cultural Affairs; 1980.
(37) van der Graaf Y. Screening for cervical cancer. The Nijmegen project. University of Nijmegen. Nijmegen: University of Nijmegen; 1987.
(38) Koopmanschap MA, Lubbe KT, van Oortmarssen GJ, van Agt HM, van Ballegooijen M, Habbema JDF. Economic aspects of cervical cancer screening. Soc Sci Med 1990;30:1081-7.
(39) van Ballegooijen M, Koopmanschap M, Habbema J. The management of cervical intraepithelial neoplasia: extensiveness and costs. Eur J Cancer 1995;28A:1703-8.
(40) van Ballegooijen M, Koopmanschap MA, Tjokrowardojo AJ, van Oortmarssen GJ. Care and Costs For Advanced Cervical Cancer. Eur J Cancer 1992;28A:1703-8.
(41) Cantor S. Cost-effectiveness analysis, extended dominance, and ethics. Med Decis Making 1994;14:259-65.
(42) Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in Health and Medicine. New York: Oxford University Press, Inc.; 1996.
(43) CBS (Netherlands Central Bureau of Statistics). Life expectancies by age and sex, 1993. Voorburg; 1994.
(44) Koopmanschap MA, van Oortmarssen GJ, van Agt HMA, van Ballegooijen M, Habbema JDF, Lubbe JTN. Cervical-cancer screening: attendance and cost-effectiveness. Int J Cancer 1990;45:410-5.
(45) Gustafsson L, Ponten J, Bergstrom R, Adami H. International incidence rates of invasive cervical cancer before cytological screening. Int J Cancer 1997;71:159-65.
(46) Waterhouse J, Muir C, Shanmugaratnam K, Powell J, editors. Cancer incidence in five continents. Vol IV. Lyon: International Agency for Research on Cancer; 1982.
(47) OECD (Organization for Economic Cooperation and Development). Health data, electronic version 1.5. 1993.
(48) Bonneux L, Barendregt J, Nusselder W, Van der Maas P. Preventing fatal diseases increases healthcare costs: cause elimination of life table approach. BMJ 1998;316:26-9.
(49) Wright J, Weinstein M. Gains in life expectancy from medical interventions -standardizing data on outcomes. N Engl J Med 1998;339:380-6.
(50) Detsky A, Redelmeier D. Measuring health outcomes-putting gains into perspective [Editorial]. N Engl J Med 1998;339:402-4.
(51) Sankaranarayanan R, Wesley R, Somanathan T, Dhakad N, Shyamalakumary B, Amma NS, et al. Visual inspection of the uterine cervix after the application of acetic acid in the detection of cervical carcinoma and its precursors. Cancer 1998;83:2150-6.
(52) Sankaranarayanan R, Shyamalakumary B, Wesley R, Sreedevi Amma N, Parkin DM, Nair MK. Visual inspection with acetic acid in the early detection of cervical cancer and precursors. Int J Cancer 1999;80:161-3.
(53) van Ballegooijen M, van den Akker-van Marle M, Warnerdam P, Meijer C, Walboomers J, Habbema J. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. Br J Cancer 1997;76:651-7.

# 6

# OVERVIEW OF IMPORTANT CERVICAL CANCER SCREENING PROCESS VALUES IN EU-COUNTRIES, AND TENTATIVE PREDICTIONS OF THE CORRESPONDING EFFECTIVENESS AND COST-EFFECTIVENESS

# SUMMARY

The objective was the evaluation of the (cost-)effectiveness of cervical cancer screening in the European Union (EU) countries. Data were collected on recommended screening age ranges and intervals, coverage, proportion of non-negative smears and smear use. Estimates reported by representatives of each participating Member State were compared, and used as input for model based (using the MISCAN simulation model for cancer screening) effectiveness and cost-effectiveness calculations.

Differences in coverage from below 50 to 82% resulted in more or less proportional differences in expected percentage life-years lost reduction, almost regardless of differences in 7 to 50+ smears recommended in a lifetime. Differences in screening intensity (resulting from the recommended number of smears per lifetime and the number of excess smears on top of these recommendations) resulted in more than 2-fold difference in the expected number of smears per percentage life-years lost reduction. (Cost-)effectiveness predictions would have greatly improved if estimates of long-term coverage had also been available.

To conclude, estimates for a restricted set of well defined parameters – a few for short and long-time coverage and one for the total number of smears - are quite useful for country-specific (cost-)effectiveness evaluations. The main, and to some extent, unsolvable problem for further improvement of the analysis is the lack of reliable country-specific estimates for the background risk of cervical cancer in women eligible for screening in the near future.

# INTRODUCTION

The usefulness of cervical cancer screening depends on the positive and negative health effects and the costs. Therefore, we collected estimates for the values of a set of key process parameters for each participating country or region of the European Union (EU). This restricted set - coverage, proportion of non-negative smears, and total number of smears for the (excess) smear use - were chosen because of their impact in predicting the effectiveness and cost-effectiveness of screening in the countries. The first objective was to describe the screening activities in EU countries quantitatively, following the publication of the "European Guidelines for Quality Assurance in Cervical Cancer Screening" in 1993 (1). The second objective was to try to use the data for assessing the effectiveness and cost-effectiveness of screening in the respective situations. The overall aim was to provide data for rational decision making concerning cervical cancer screening on a national, regional or local level.

Coverage, proportion of positive smears and excess smear use are closely related to each other and to the effects and costs of Papanicolaou (Pap) smear screening. Smears either contribute to coverage, or they are excess smears. Those contributing to coverage help to reach the potential effectiveness of the screening programme. Excess smears consist of smears outside the target age-range or those taken after too short an interval. They add little to the effectiveness of the regular programme smears and therefore decrease the cost-effectiveness of the screening activities. Follow-up smears, which can be regarded as diagnostic excess smears, depend on the proportion of positive (or at least non-negative) smears. Positive smears are the key to positive health effects, but they also generate negative health effects.

# MATERIALS AND METHODS

## The working group

The Epidemiology Working Group of the European Cervical Cancer Screening Network consists of representatives of 13 of the 15 EU Member States (there was no representative from Austria and Luxembourg). The working group came together twice in 1998 in Rotterdam. A set of quantitative data to be collected from each country or region was discussed and agreed upon. The data were chosen to describe important aspects of the (cost-)effectiveness of screening. In addition, the definitions of the concepts behind the data were decided upon. It was agreed that the most recent data available should be used. Where the data sources are not mentioned in the presentation here, they can be found in the country specific contributions to the Special Issue: Cervical Cancer Screening in de Eurpean Union (2).

*Table 6.1*

*Policies or recommendations for cervical cancer screening by EU country*

| Country | Screening age-range | Screening interval in years | Smears per woman in a lifetime |
|---|---|---|---|
| Belgium (B) | 25-64 | 3 | 14 |
| Denmark (DK) | 23-59 | 3 | 13 |
| Finland (FIN) | 30-60 | 5 | 7 |
| France (F) | 25-65 | 3 | 14 |
| Germany (G) | 20+ | 1 | 50+ |
| Greece[a] (Ormylia) (GR) | 25-64 | 3 | 14 |
| Ireland (IRL) | 25-60 | 5 | 8 |
| Italy (I) | 25-64 | 3 | 14 |
| Netherlands (NL) | 30-60 | 5 | 7 |
| Portugal[a] (Midregion) (P) | 20-64 | 3 | 16 |
| Spain[a] (C. y. León) (E) | 25-65 | 3 | 14 |
| Sweden (S) | 20-59 | 3 | 14 |
| UK (England) (UK) | 20-65 | 3 or 5 | 16-10 |

[a] Of Greece, Portugal and Spain, no national recommendations were available

## The age-range and screening interval

Evaluation of the screening process values studied requires knowledge of the recommended age range and screening interval. These recommendations differ between countries and regions (Table 6.1).

## Process values: definitions

In order to collect comparable and coherent estimates for the respective process parameters, the following definitions have been agreed upon.

*Coverage*

We will consider the *3-year coverage* for direct comparison between countries, and the so-called *interval coverage* for the policy-specific model predictions. In both cases, the denominator is the number of women in the target age group in the population of the area in question. The numerator is the number of women in the target age group that had at least one smear in the period preceding the moment of evaluation. For the 3-year coverage this period is fixed at 3. For the interval coverage the last *i* years are considered, where *i* is the length in years of the recommended screening interval.

*Positive screening results*

The *percentage screen positives* only concerns the primary (as opposed to follow-up) programme smears. It is the percentage of non-negative adequate smears. Positive is defined as having a more stringent follow-up recommendation than the normal screening policy. Thus, results that require a

repeat smear, e.g. after 6 months (e.g. atypical squamous cells of undetermined significance (ASCUS)) are included in the percentage.

*Excess smear use*

We will call all smears not contributing to the coverage *excess smears*. Some of these smears are follow-up smears after non-negative screening results or are taken because of signs or symptoms. Smears following signs or symptoms would also occur without screening (although their number can be influenced by screening). The reason we had to add them to preventive excess smears is that they can often not be discerned from each other in registrations. In order to obtain comparable figures for all countries, we therefore chose to count all smears for calculating the excess smear use.

We considered the excess smears per year related to the 3-year interval for direct comparison between countries, and the excess smear related to the recommended screening interval for the policy-specific model predictions. The former is expressed in number of excess smears per year per thousand women, the latter as a percentage of excess smears considering all smears. The formulas used are as follows (for an example, see the footnotes of Table A2 in the Appendix).

*The number of excess smears per year per 1000 women [related to the 3-year interval]=*

(total yearly number of smears – number of smears needed yearly to reach the observed 3-year coverage) * 1000 / number of women in the target population where the yearly number of smears needed to reach the observed coverage is: the population in the target age range * 3-year coverage / 3.

*The percentage excess smear use [related to the recommended screening interval]=*

the percentage of excess smears of the total number of smears = (total number of smears * 100/number of smears needed to reach the observed coverage) – 100 % where the number of smears needed to reach the observed coverage is: (the population in the target age range * *i*-year coverage) / *i* , where *i* = the length in years of the recommended screening *interval*

*Target population*

The number of women in the target population is defined as the number of eligible women in the country or region in a screening round (a period of screening that lasts the recommended screening interval). For the last previous round, this is the number of women in the age range beginning with the recommended starting age and ending with the recommended ending age plus the number of years in the screening interval minus one year (e.g. 25 up to but not including 67 years of age for the '25 to 64 every 3 year' policy).

*Cumulative risk*

The cumulative risk is defined as the cumulative (background) incidence to age 100 in the hypothetical situation without screening.

## Model-calculations: predicted effects and costs

We used the MISCAN cervical cancer screening simulation model to exploratively predict effects and costs of screening in EU member states. MISCAN is a microsimulation model described extensively elsewhere (3,4). The principal predicted effect measure presented is percentage life-years lost reduction. The number of smears is used as an approximate proportionality factor for the costs. Accordingly, the number of smears per percentage life-years lost reduction is the cost-effectiveness measure presented. The number of life-years gained per 1000 women and the number of smears per life-year gained are also discussed. Calculations were made for different screening policies (age range and interval combinations) and different coverage and excess smear rates. More precisely, the *interval* coverage rates and the *interval related* percentages of excess smears are used as input for the predictive calculations. The impact of different risk levels is also discussed.

Fixed parameters in this exploration, and thus parameters for which eventual differences between countries and regions are not accounted for, are:

1. The natural history of cervical cancer, especially the mean and variance of the duration of pre-clinical (pre-invasive and invasive) detectable disease.
2. The sensitivity (of the combination of screening test and follow-up).
3. The stage-specific prognosis after treatment.

These fixed parameters determine the incidence and mortality reducing potential of Pap smear screening. Compared with other models in the literature (the one of Eddy (5) and of Gustafsson and Adami(6)) the mortality reduction predicted by the MISCAN model in women participating in screening is at the same level (approximately 75% for a 30 to 60 every 5 year policy and around 90% for e.g. 16 smears between age 20 and 70 years). We used the MISCAN model because it can be easily tuned to different screening situations in different countries.

The age-distribution of the incidence of cervical cancer was also fixed. As Gustafsson and colleagues showed by studying age-specific incidence rates from different countries in periods before screening started, these distributions follow very much the same pattern for many Western European countries (among others Denmark, Germany, The Netherlands and Sweden with a peak-age at 44-47 years), although some countries seem to have a slightly different distribution (Finland with a peak age at 53.5 years; the UK with somewhat later onset, a peak-age of 48 years and a slower decline after the peak) (7).

The calculations presented in this paper concern a complete screening of a birth cohort of women following the recommended policy. Therefore, the results represent a steady-state situation in which screening has and will run forever, and in which all birth cohorts have the same cumulative risk of cervical cancer. What cumulative risk to consider when in several Western countries an increased risk is observed for cohorts born after, e.g. 1940 or thereabouts is a subject of discussion. Hysterectomies for reasons other than the management of (precursors of) cervical cancer are not taken into account.

*Table 6.2*

*Estimates for outcome parameter values of cervical cancer screening by EU country. For definitions of the outcome parameters, see the text.*

| Country | 3-year or [5-year][b] coverage (%) | Screen-positives (%) | 3-year excess smears (per 1000 women) | Population subjected to formal programme (%) |
|---|---|---|---|---|
| Belgium (B) | 78 | 3 | 167 | 58 |
| Denmark (DK) | 75 | 5 | 205 | 90 |
| Finland (FIN) | [93][b] | 5 | 121[b] | 100 |
| France (F) | N.r.e. | 5 | N.r.e. | <5 |
| Germany (G) | 80 | 7 | 248 | 90 |
| Greece[a] (Ormylia) (GR) | 71 | 5 | 117 | 88 |
| Ireland (IRL) | N.r.e. | 3 | N.r.e. | 0 |
| Italy (I) | 50 | N.r.e. | 77 | 13 |
| Netherlands (NL) | [77][b] | 5 | 24[b] | 100 |
| Portugal[a] (Midregion) (P) | 37 | 5 | 86 | 100 |
| Spain[a] (C. y. León) (E) | 27 | 15 | 14 | 86 |
| Sweden (S) | 82 | 1.5 | 140 | 100 |
| UK (England) (UK) | 61 | 8 | 90 | 100 |
| Average[c] | 75 | 5 | 134 | |

N.r.e. No reliable estimate. For France and Ireland the coverage ever was estimated at 60% and 65% respectively. If these rates were used as if it were 3-years coverage rates, the calculated number of 3-yearly excess smears would be 133 and -10 per year per 1000 women respectively.

[a] Of Greece, Portugal and Spain, no national data were available.

[b] For Finland and the Netherlands, only 5-year coverage rates were available. Therefore, the number of excess smears was calculated with the 5-year coverage.

[c] Unweighted average

For each combination of age range and recommended screening interval, the influence on (cost-) effectiveness was computed for various coverage and excess smear rates.

# RESULTS

## Estimates for screening process values in countries and regions

In Table 6.1 we describe the screening policies in the EU countries or regions, and in Table 6.2 we present the estimates for the screening process parameters resulting from the collected data. For information on other regional pilot projects and further details see Tables A1 and A2 in the Appendix. Although the definitions to be used in this paper were well set, the available data did not always make it possible to exactly meet these definitions. Therefore, the tables should not be interpreted without studying the country- and region-specific remarks in the Appendix. (Differences with figures reported in the country specific papers in the Special Issue: Cervical Cancer Screening in the European Union (2) are due to differences in definitions).

Although all policies are mainly in line with the European recommendations (screening women every 3-5 years), there is a large variation in screening intensity that is a consequence of these policies. This intensity varies from 7-16 smears per woman in a lifetime, with the exception of Germany where there is a 1-year screening interval and over 50 smears taken in a lifetime. The 3-

*Table 6.3*

*Percentage life-years lost reduction by policy, coverage and two assumptions on long-term distribution of participation in the population: the same women participate in all rounds (i.e. participants-participation is systematic), or participation is independent of previous participation (i.e. participants-participation is random)*

| Policy[a] | NL/FIN 30(5)60[7] | | B/F[b]/GR/I/E 25(3)64[14] | | G 20(1)72[53] | |
|---|---|---|---|---|---|---|
| Participants participation | Systematic | Random | Systematic | Random | Systematic | Random |
| Interval-coverage | % Life-years lost reduction | | | | | |
| 25% | 21 | 36 | 24 | 56 | 25 | 90 |
| 50% | 42 | 60 | 47 | 80 | 50 | 98 |
| 75% | 63 | 75 | 71 | 90 | 75 | 99 |
| 100% | 84 | 84 | 94 | 94 | 99.9 | 99.9 |

[a] Starting age (interval) ending age [number of smears in lifetime], policies ranked by increasing number of smears in a lifetime.

[b] For France, the stopping age is 65 years.

year coverage in the participating countries varies from 50 to 82%. The figures for the regions of Portugal and Spain involved in the screening programmes are lower, possibly because in these cases coverage by other than programme smears is not accounted for. For France and Ireland, no data on 3-year (or 5-year) coverage were available. The excess smear use varies strongly. However, as has been explained for each country in the Appendix, there are many reasons why these figures are not always comparable. The percentage screen positives varies from 3 to 8% of the screened women. This may reflect differences in the prevalence of neoplastic lesions or differences in cut-off point between negative and positive smears (between 'no follow-up required' and 'at least a repeat smear recommended'). In any case, the percentage of screened women that undergo some kind of negative effect of screening due to follow-up varies accordingly.

Altogether, the figures summarised in Table 6.2 plus the details in the Appendix, show how far we have got in estimating the respective parameters in EU Member States, and how much work remains to be done.

## Model-based predictions

On basis of the data collected, effectiveness and cost-effectiveness were predicted for a variety of screening situations, differing in screening strategy, coverage and smear use.

*Life-years lost reduction*

We focused on the reduction in life-years lost from cervical cancer as the effect measure of screening. (In Table A3 in the Appendix, the predicted percentage in incidence and mortality reduction is also presented.) First, we predicted the effectiveness at 100% coverage (and no excess smear use), see the bottom line of Table 6.3. We did so for three screening policies, including the least and most intensive policy and the intermediate EU recommended policy (see Table A3 for all policies). Because we assume an identical age distribution of incidence across countries, these numbers are applicable to any country with the policy under consideration. The number of life-

*Table 6.4*

*Predicted percentage life years lost reduction by policy and coverage, assuming systematic participants-participation*[a]

| Policy[b] | NL/FIN 30(5)60 [7] | IRL 25(5)60 [8] | UK(5) 20(5)65 [10] | DK 23(3)59 [13] | S 20(3)59 [14] | B/F[c]/GR/I/E 25(3)64 [14] | UK(3)/P 20(3)65 [16] | G 20(1)72 [53] |
|---|---|---|---|---|---|---|---|---|
| Interval-coverage | % Life-years lost reduction | | | | | | | |
| 20% | 17 | 18 | 19 | 18 | 18 | 19 | 19 | 20 |
| 25% | 21 | 22 | 23 | 23 | 23 | *24 E* | 24 | 25 |
| 30% | 25 | 27 | 28 | 28 | 28 | 28 | 29 | 30 |
| 35% | 29 | 31 | 33 | 32 | 32 | 33 | *34 P* | 35 |
| 40% | 34 | 36 | 37 | 37 | 37 | 38 | 38 | 40 |
| 45% | 38 | 40 | 42 | 41 | 42 | 42 | 43 | 45 |
| 50% | 42 | 44 | 47 | 46 | 46 | *47 I* | 48 | *50 G*[a] |
| 55% | 46 | 49 | 51 | 51 | 51 | 52 | 53 | 55 |
| 60% | 50 | 53 | 56 | 55 | 55 | *56 F(max)* | *58 UK(3)* | 60 |
| 65% | 55 | *58 IRL(max)* | 61 | 60 | 60 | 61 | 62 | 65 |
| 70% | 59 | 62 | 65 | 64 | 65 | *66 GR* | 67 | 70 |
| 75% | *63 NL* | 67 | *70 UK (5)* | *69 DK* | 69 | *71 B* | 72 | 75 |
| 80% | 67 | 71 | 75 | 74 | *74 S* | 75 | 77 | 80 |
| 85% | 71 | 75 | 79 | 78 | 79 | 80 | 82 | 85 |
| 90% | 76 | 80 | 84 | 83 | 83 | 85 | 87 | 90 |
| 95% | *80 FIN* | 84 | 88 | 87 | 88 | 89 | 91 | 95 |
| 100% | 84 | 89 | 93 | 92 | 92 | 94 | 96 | 100 |

For each country, the prediction resulting from using the estimates on coverage presented in Table A2 are indicated. The values for Greece, Portugal and Spain concern only part of these countries, see Table A2. (max), For Ireland and France, the indicated value is a maximum since it is based on the 'coverage ever' as if it was the interval coverage. Italicized values are those were estimates of interval coverage are applied. For country abbreviations, see Table 6.1.

[a] Assuming systematic participants participation is relatively unfavorable for policies with frequent screening. This is especially the case for Germany (see Table 6.3 and text).

[b] Starting age (interval) ending age [number of smears per women in a lifetime], policies ranked by increasing number of smears in a lifetime.

[c] For France the stopping age is 65 years.

years lost because of cervical cancer is reduced by between 84% and 94% when screening women between 7 and 14 times, respectively. The German policy with over 50 smears a lifetime is predicted to result in an almost 100% reduction (99,9%). The actual percentage life-years lost reduction depends on the coverage. Therefore, the predicted percentages are given by the coverage rate. This was done for two assumptions on the long-term coverage: (1) it is always the same women who participate at screening in successive rounds (participants-participation is 100% systematic), and (2) participation is independent of previous participation (participants-participation is random). In the first assumption, long-term coverage is equal to coverage within one screening round. With random participation, the long-term coverage increases every screening round. Therefore the mortality reduction is higher with random participants-participation. Assumption (1) is extremely unfavourable and (2) extremely favourable for screening, especially for frequent screening and a low (short-term) interval coverage (see the predictions for the German policy). In fact, with random participants-participation and a low interval coverage, the average screening interval becomes much longer than the recommended one, so that results are in some sense no longer representative for the screening interval under consideration. In Table 6.4, the results are given for all policies for the conservative assumption (1). Differences in coverage resulted in more or less proportional differences in expected percentage life-years lost reduction, with much

*Table 6.5*

*Predicted number of smears (x mln) per % life-years lost reduction by policy and excess smear use, assuming systematic participants-participation[a].*

| Policy[b] | NL/FIN 30(5)60 [7] | IRL 25(5)60 [8] | UK(5) 20(5)65 [10] | DK 23(3)59 [13] | S 20(3)59 [14] | B/F[c]/GR/I/E 25(3)64 [14] | UK(3)/P 20(3)65 [16] | G 20(1)72 [53] |
|---|---|---|---|---|---|---|---|---|
| Excess smear use | Number of smears per % life-years lost reduction | | | | | | | |
| 0% | 1.9 | 2.1 | 2.4 | 3.2 | 3.5 | 3.3 | 3.7 | *11.3 G[a]* |
| 10% | 2.1 | 2.3 | 2.7 | 3.5 | 3.8 | 3.7 | 4.1 | 12.5 |
| 20% | 2.3 | 2.5 | 2.9 | 3.9 | 4.2 | *4.0 E* | 4.5 | 13.6 |
| 30% | 2.5 | 2.7 | 3.2 | 4.2 | 4.5 | 4.3 | 4.9 | 14.7 |
| 40% | 2.7 | 2.9 | 3.4 | 4.5 | 4.8 | 4.7 | *5.2 UK (3)* | 15.9 |
| 50% | 2.9 | 3.1 | 3.7 | 4.8 | *5.2 S* | *5.0 GR/I* | 5.6 | 17.0 |
| 60% | 3.1 | *3.3IRL(min)* | 3.9 | 5.2 | 5.5 | *5.3 B* | 6.0 | 18.1 |
| 70% | 3.3 | 3.5 | 4.1 | 5.5 | 5.9 | *5.7 F(min)* | *6.4 P* | 19.3 |
| 80% | *3.5 NL* | 3.7 | 4.4 | *5.8 DK* | 6.2 | 6.0 | 6.7 | 20.4 |
| 90% | 3.7 | 3.9 | *4.6 UK(5)* | 6.1 | 6.6 | 6.3 | 7.1 | 21.5 |
| 100% | 3.9 | 4.1 | 4.9 | 6.5 | 6.9 | 6.7 | 7.5 | 22.7 |
| 110% | 4.1 | 4.4 | 5.1 | 6.8 | 7.3 | 7.0 | 7.9 | 23.8 |
| 120% | 4.2 | 4.6 | 5.4 | 7.1 | 7.6 | 7.4 | 8.2 | 24.9 |
| 130% | *4.4 FIN* | 4.8 | 5.6 | 7.4 | 8.0 | 7.7 | 8.6 | 26.1 |
| 140% | 4.6 | 5.0 | 5.9 | 7.7 | 8.3 | 8.0 | 9.0 | 27.2 |

For each country, the prediction resulting from using the estimates on coverage presented in Table A2 are indicated. The values for Greece, Portugal and Spain concern only part of these countries, see Table A2. (min), For Ireland and France, the indicated value is a minimum since it is based on the excess smear use calculated with the 'coverage ever' as if it is the interval coverage. Italicized values are those where estimates of interval coverage are applied. For country abbreviations, see Table 6.1.

[a] Assuming systematic participants participation is relatively unfavorable for policies with frequent screening. This is especially the case for Germany (see Table 3 and text)

[b] Starting age (interval) ending age (number of smears per women in a lifetime), policies ranked by increasing number of smears in a lifetime

[c] For France the stopping age is 65 years.

less impact for the number of smears recommended in a lifetime, which varies from 7 to over 50 smears.

For example, for a 7 smears per lifetime policy, increasing the coverage from 50 to 75% (which will increase the number of smears with approximately 50%) will add (63 - 42 =) 21% extra life-years lost reduction, while intensifying screening to 14 smears in a lifetime (twice as many smears) will only add (47 – 42 =) 5% life-years lost reduction. In the table, we italicised the predictions per EU Member State if the estimates for the interval coverage (see Table A2 in the Appendix) are applied.

*Numbers of smears per percentage life-years lost reduction*

The number of smears per percentage life-years lost reduction depends on the policy and the excess smear use. Predictions are given in Table 6.5 for each policy, with the results per EU Member State if the data collected on excess smear use are applied. Policies with a low smear-taking intensity (fewer smears recommended a lifetime and fewer excess smears in addition to the recommended smears) have a more favourable cost-effectiveness ratio compared with policies with many smears in a lifetime. Differences in + 40-130% excess smears and in 7-16 of smears recommended in a lifetime resulted in approximately 2-fold differences in the expected number of smears per percentage life-years lost reduction. If always the same women participate, as was

assumed here (i.e. participators-participation is 100% systematic), the predicted cost-effectiveness is independent of the coverage. Especially for Germany, where we know that the coverage after three 1-year screening rounds is considerably higher (80%) than after one round (50%), these predictions are underestimating the life-years lost reduction from screening (Table 6.4) and thus overestimating the number of smears per percentage life-years lost reduction (Table 6.5). Obviously, many participating women are not screened every year as recommended, but at a longer interval, which improves cost-effectiveness.
It should be noted that the extra mortality reduction resulting from preventive smears outside the target age range and recommended screening interval is neglected in these predictions. The expected influence of this simplification will be limited (see Table 6.3).

*Negative side effects*
In order to produce a measure for the negative side-effects, one could calculate the predicted number of screen-positives per cent life-years lost reduction (by multiplying the number of smears per percentage life-years lost reduction in Table 6.4 by the proportion of screen-positives in Table 6.2). However, the percentage of screen-positives is no more than an initial very approximate approach to quantify negative side-effects. The next necessary step would be to divide this percentage in women who are and women who are not referred to colposcopy/biopsy before the end of the follow-up episode. One would also have to describe how long (how many years) follow-up occurs in the respective groups, accounting for the period until a woman resumes regular screening, and what medical procedures (from repeat smears to conisation and hysterectomy) take place during this period.

## DISCUSSION

The EU member states have implemented a variety of screening policies. The screening interval varies from 3 to 5 years and the number of smears offered in a lifetime from 7 to 16 (Table 6.1). Germany uses recommendations that strongly differ from those from other countries, with a 1-year screening interval and over 50 smears per woman in a lifetime. Screening process values also differ between the countries (Table 6.2). The 3-year coverage varies from below 50% to over 82%, the excess smear use from less than 100 to over 200 per year per 1000 women, and the percentage screen-positives from 1.5% to 8%.
The predicted (cost-)effectiveness varied accordingly. Differences in coverage of 50-90% resulted in more or less proportional differences in the expected percentage life-years lost reduction, with a much smaller impact for differences from 7 to 16 in the number of smears recommended in a lifetime. Differences from 40 to 130% in excess smear use combined with the already mentioned differences in number of smears in a lifetime, resulted in at least 2-fold difference in the expected number of smears per percentage life-years lost reduction.
The fact that an association between high risk and non-attendance has been repeatedly observed (*8-10*) makes the predictions too favourable.

Hysterectomy rates can be substantial. In the UK the rate is 25% under age 55 years (data not shown). By including women without a cervix uteri in the denominator of the coverage rate, this rate is more seriously underestimated in countries with high hysterectomy rates than elsewhere.

The values for the parameters presented in Tables 6.2 and Table A.2 represent the best available estimates at the time they were collected by the working group. Some are based on (almost) nationwide registrations, others on more or less thorough surveys, or are the best guesses of experts. Sometimes, as explained in the text, the figures are minimum or maximum estimates. The potential biases are notified in the country-specific remarks in the Appendix. The lack of data on opportunistic screening is often a problem. Programme screening aims for early detection and treatment of cervical cancer. However, other smears have the same aim. The total performance of early detection in a country is the aggregate of programme screening and other smears taken. Our analysis shows that the high rate of excess screening in most countries or regions (well over 60% in most cases), caused by opportunistic screening either running alone or along with programme screening, results in cost-ineffective situations. In order to improve these situations, it is necessary to monitor the opportunistic screening activities together with the organised screening activity.

A next step in the cost-effectiveness analysis of cervical cancer screening is to go from the percentage life-year lost reduction to the number of life-years gained as the effect measure, and consequently also to the number of smears per life-year gained as the cost-effectiveness measure. To this end, country-specific knowledge is required on the cumulative risk: the cumulative incidence in situations without screening. Taking into account this cumulative risk is conditional for a judgement on how many smears in a lifetime is acceptable for a given country or region. We presented predicted numbers of life-years gained and number of smears per life-year gained for a background cumulative risk of 1% in Tables A4 and A5 in the Appendix. The results in these tables can easily be adjusted to any specific estimate for the cumulative risk (see footnotes to the respective tables). The percentage for women currently eligible for screening in the EU probably varies between 1 and 4%, depending on the birth cohort and geographic area under consideration. However, cervical cancer is not a very stable disease as far as risk over time is concerned. In most countries, screening started decades ago. This means there is a lot of uncertainty about the cumulative risk, especially for young birth cohorts, and that it will be more and more difficult (if not impossible) to estimate the cumulative risk in the future, even with new high quality data. This has consequences as to how proceed with the work presented. The predictions could be improved in their accuracy in many ways, as has been pointed out earlier in this paper. For instance, one could account for age dependency of coverage or for the hysterectomy rates. But is this worthwhile if these uncertainties and simplifications in the analysis are dominated by the uncertainty about the cumulative risk? This will be subject to further discussion among evaluators of cervical cancer screening.

## CONCLUSIONS

The results stress the impact of cervical cancer screening evaluation to provide reliable estimates for a restricted set of parameters: the short- and long-term coverage of screening (including opportunistic screening) and the total amount of smears (including opportunistic smears). The results also show the importance of a high coverage for the effectiveness of screening, and of a restricted intensity of smear taking for cost-effectiveness. Intensity of screening is derived from the combination of the recommended number of smears in a lifetime and the number of (excess) smears taken on top of these recommendations. In some countries, one might consider de-intensifying the recommended cervical cancer screening policy (i.e. fewer smears in a lifetime). These latter conclusions are not new. The presentation in this paper is highly individualised to participating countries and will therefore hopefully have its own impact on the improvement of cervical cancer screening in the respective countries and regions.

## REFERENCES


(1) Coleman D, Day N, Douglas G, Farmery E, Lynge E, Philip J, et al. European Guidelines for Quality Assurance in Cervical Cancer Screening. Europe against cancer programme. Eur J Cancer 1993;29A:S1-38.

(2) Linos A, Riza E, van Ballegooijen M. Cervical cancer screening in the European Union. Eur J Cancer 2000;36:2175-275.

(3) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.

(4) Koopmanschap MA, Lubbe KT, van Oortmarssen GJ, van Agt HM, van Ballegooijen M, Habbema JK. Economic aspects of cervical cancer screening. Soc Sci Med 1990;30:1081-7.

(5) Eddy DM. Screening for cervical cancer. Ann Intern Med 1990;113:214-26.

(6) Gustafsson L, Adami HO. Optimization of cervical cancer screening. Cancer Causes Control 1992;3:125-36.

(7) Gustafsson L, Ponten J, Bergstrom R, Adami HO. International incidence rates of invasive cervical cancer before cytological screening. Int J Cancer 1997;71:159-65.

(8) 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.

(9) 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.

(10) Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-6.

(11) Arbyn M, van Oyen H, Mulkens A. Implementation of central registration of cervical cancer screening in Flanders. International Symposium on Cervical Cancer Screening. Antwerp; 1998.

(12) Ronco G, Iossa A, Naldoni C, Pilutti S, Anghinoni E, Zappa M, et al. A first survey of organized cervical cancer screening programs in Italy. GISCi working group on organization and evaluation. Gruppo Italiano Screening Citologico. Tumori 1998;84:624-30.

# APPENDIX

## Details on estimates for screening process values in countries and regions

*Belgium*

In Belgium, there is a nation-wide consensus about the age range and screening interval, but a formal screening programme has only been implemented in Flandres (covering 58% of the Belgium population). In a telephone interview the 3-year coverage in Flandres was estimated at 82%. According to a Health Interview Survey the 3-year coverage in Flandres was almost 10% higher than in the other part of the country. Therefore, the coverage on a national level was estimated to be 0.58*82%+0.42*72%=78%. The difference between Flandres and the other part of the country in the percentage of screen-positives and in the number of excess smears is not known. The percentage of screen-positives (3%) is only known for the Flemish region (for programme and opportunistic screening, excluding smears with a clinical indication and follow-up smears (*11*)). This percentage was also used for the national estimate (Table 6.2). The total number of Pap smears (opportunistic and organised) is known only for the whole country. The data presented for the Flemish Region are based on estimation.

*Denmark*

In 1997, the screening programme with personal invitations covered 90 % of the 23-59-year-olds and 46% of the 60-74 year age group of women. According to the national guidelines, the latter age group had to be invited once. The 75% 3-year coverage and the 5% screen-positives were estimated on basis of data from 1994 to 1996 from the Copenhagen and Frederiksberg municipalities (see Table 6.2, 9% of Danish population) (data not shown).

*Finland*

Programme screening covers over 100% of the country, although ages targeted since the late 1980s (55, 60) makes that overall only 87% of the target population is covered. The percentage invited women among 30-year-old women has been approximately 70%. The 5-year coverage on basis of an annual population survey was estimated to be 93% for any Pap smear. There is no direct estimate of the 3-year coverage, but 18% of the women had a smear once every 5 years, and 27% every 3-4 years, so that the 3-year coverage could be (93-18-27/2=) 62%. The estimated annual number for all smears (including opportunistic and diagnostic smears and all age groups) is 500,000 – 600,000 (data not shown).

*France*

In France in 1998, except for three pilot cervical cancer programmes covering less than 5% of the country, screening was opportunistic. It is estimated that 60% of the 20-69 year age group of

women in France had a cervical smear (no separate data are available for the target age group of 25-64 years). In the region of Bas-Rhin (with programme screening on basis of public announcements), the 3-year coverage in the target age group was 69% and the 3.5-year coverage 75%. For the percentage of positive smears national data were not available. In Bas-Rhin it was 5% (follow-up smears excluded, clinical smears e.g. because of symptoms included).

*Germany*

In Germany, statutory health insurers, issue yearly a voucher (computer-readable plastic card) to all persons and thus make free annual cervical screening available to ≥ 20-year old women, covering 90% of the population. In this 90% (and presumably also in the other 10%) the 1-year coverage is approximately 50%. The 3-year coverage is over 80% if only programme smears are accounted for (accounting for all smears it will be higher still). The total annual number of smears is estimated at 15 million programme smears plus at least 1.5 million of 'private' smears. The percentage of screen-positives is estimated at 7%, including roughly 5% ASCUS (data not shown).

*Greece*

In Greece, two regional programmes are running, one in Ormylia and one in Messina and Ilia. Data on the process parameters needed for this paper were only available for the Ormylia programme. Both the coverage and the total number of smears only take programme smears (including follow-up smears) into account. Data on other smears are not available.

*Ireland*

To date, opportunistic screening is occurring in Ireland. The age-range of 25-60 years and an interval of 5 years or shorter was recommended in national guidelines in 1996. This results in a minimum of eight smears per woman per lifetime. The percentage of women (aged 25-60 years) who ever had a smear on the basis of a survey was estimated to be 65%. The estimated 3% of screen-positives concerns one large laboratory. All smears are included (also follow-up smears), so 3% is an overestimate (data not shown).

*Italy*

National guidelines were decided in 1996. In 1997, 13% of female target population was covered by programme screening, but this is rapidly increasing to probably around 50% in the year 2000. Local surveys on 3-year coverage conducted in the late 1980s provided estimates of less than 50% in the absence of organised screening. In areas that had screening programmes in 1997 (covering 13% of the female target population) approximately two-thirds could report coverage. In these regions, the 3-year coverage rate was estimated at 66% of the invited women (not all the programmes have run for 3 years) (12). In Turin, 3-year coverage was 43% before programme screening started in 1992, and was 74% in invited women in 1997. The total number of smears in Turin is estimated and subject to uncertainty. In Florence, where women without a smear in the last 3

years have been invited since 1980, the 3-year coverage in 1997 was 39% and the 4-year coverage 49%. In this latter measurement, 'private smears' were not included. These smears are also not included in the total number of smears (data not shown).
The percentage of screen-positives (including those requiring repeat smears) was not available for the programmes running in 1997. The average colposcopy referral rate was 2%. The percentage of positives (including those requiring repeat smears) in Florence was 5% (follow-up smears excluded, but clinical smears included) and 10% in Turin (programme smears only). The large majority of positive smears in the Turin programme only imply a single additional smear.

*The Netherlands*
In the Netherlands, the previous national 3-yearly screening policy between ages 35 and 53 years was changed into a 5-yearly policy between the ages 30 and 60 year in 1996. Coverage and the total number of smears are based on a nationwide registry including all smears in the country, irrespective of the reason for which they were taken. Not all 30-34-year olds and 55-64-year olds had at least one invitation in the last 5 years, because the first 5-year round with the extended age-range was not completed by the end of 1997. The percentage of screen-positives accounted for primary programme smears only, and has decreased from over 10% in 1994 to 5% in 1997.

*Portugal*
In 1990, programme screening was launched in the Central Region of Portugal. Screening data are only available from this programme. Initially a 1-year interval is used before proceeding with a 3-year interval. The 5% of screen-positives probably refers to secondary (follow-up) smears as well, and thus might be too high.

*Spain*
No national cervical cancer screening data are available from Spain. The data presented concern a (pilot) programme in Castilla Y Léon. In this programme, a 1-year interval is recommended before proceeding with a 3-year interval. The 3-year coverage as defined here is estimated at 27%, not including women covered by 'private' smears. The estimated annual number of 65,000 smears also does not include 'private' smears. It does include follow-up smears after programme-smears. Fifteen per cent of the smears result in at least the recommendation of a cytological follow-up, of which 14% is classified as 'with infections, including viruses', and 0,8% as 'with morphological alterations'.

*Sweden*
Sweden has programme screening nationwide. According to the national guidelines a 3-year screening interval is recommended, but almost half of the counties use a 4-year interval. Women with a recent smear (within 18 months) are sorted out and not invited. The coverage is based on data for the city of Malmö: 76% of the women had a recent smear, and a quarter of the other 24%

attended the screening programme. In some rural areas coverage is lower, therefore the 82% may be too high an estimate for the total country.

*United Kingdom*

In the United Kingdom a national screening programme is running. The data refer to screening in England. The target of the national programme is to screen women aged 20-64 years at least every 5 years. However, more than half of the health authorities invites women every 3 years. The 5-year coverage for the whole country is 76%, the 3-year coverage is 61%. Ideally, those parts of the country with a 3-yearly screening programme should be evaluated separately from those with a 5-yearly programme. The 8% screen-positives concerns all smears (including follow-up smears) instead of only primary programme smears, and may therefore be an overestimate (data not shown).

***Table A1***

*Policies for cervical cancer screening by EU country or region*

*National data*[a]

| Country | Policy/recommendations | | | Period described | Number of women in target population (x1000) | Population subjected to formal programme |
|---|---|---|---|---|---|---|
| | Age-range | Interval | Smears per woman | | | |
| Belgium | 25-64 | 3 | 14 | 1995/6/7 | 2,712 | 58% |
| Denmark | 23-59 | 3 | 13 | 1997/98 | 1,429 | 90% |
| Finland | 30-60 | 5 | 7 | 1996 | 1,275 | 100% |
| France | 25-65 | 3 | 14 | 1998 | 18,000 | <5% |
| Germany | 20+ | 1 | 50+ | 1996 | 33,000 | 90% |
| Ireland | 25-60 | 5 | 8 | 1996/7 | 792 | 0% |
| Italy | 25-64 | 3 | 14 | 1994/5/6 | 15,369 | 13% |
| Netherlands | 30-60 | 5 | 7 | 1997 | 3,692 | 100% |
| Sweden | 20-59 | 3 | 14 | 1994 | 2,300 | 100% |
| UK (England) | 20-64 | 3 or 5 | 16-10 | 1996/7 | 15,049 | 100% |

*Regional data (regions with programme screening/ pilot projects)*

| Region | Policy/recommendations | | | Period described | Number of women in target population (x1000) | Population subjected to programme |
|---|---|---|---|---|---|---|
| | Age-range | Interval | Smears per woman | | | |
| Flandres (B) | 25-64 | 3 | 14 | 1995/6/7 | 1,573 | 100% |
| Copenh.+ (DK) | 23-59 | 3 | 13 | 1998 | 165 | 100% |
| Bas-Rhin (F) | 25-64 | 3 | 14 | 1998 | 255 | 100% |
| Ormylia (GR) | 25-64 | 3 | 14 | 1997 | 13 | 88% |
| Florence (I) | 25-64 | 3 | 14 | 1997 | 206 | 100% |
| Turin (I) | 25-64 | 3 | 14 | 1996/97 | 271 | 80% |
| Midregion (P) | 20-64 | 3 | 16 | 1995/97 | 292 | 100% |
| C. y. León (E) | 25-65 | 3 | 14 | 1990+ | 628 | 86% |

B, Belgium; DK, Denmark; F, France; GR, Greece; I, Italy; P, Portugal; E, Spain

[a] Of Greece, Portugal and Spain, no national data are available

*Table A2*

*Estimates for outcome parameter values of cervical cancer screening by EU country or region. For definitions of the outcome parameters, see the text.*

*National data*[a]

| Country | Coverage | | % Screen-positive | Annual number of smears (x1000) | Excess smears | | Women in target population[d] (x1000) |
|---|---|---|---|---|---|---|---|
| | 3-year | Recomm. Interval | | | 3-year (per 1000 women)[b] | Recomm. Interval (%)[c] | |
| Belgium | 78% | 78% | 3 | 1,158 | 167 | 64% | 2,712 |
| Denmark | 75% | 75% | 5 | 650 | 205 | 82% | 1,429 |
| Finland | n.r.e. | 93% | 5 | 550 | 121[f] | 132% | 1,275 |
| France | n.r.e. | n.r.e. | 5 | 6,000 | 133[g] | 67%[g] | 18,000 |
| Germany | 80% | 50% | 7 | 17,000 | 248 | 3% | 33,000 |
| Ireland | n.r.e. | n.r.e. | 3 | 164 | -10[g] | 59%[g] | 792 |
| Italy | 50% | 50% | n.r.e. | 3,750 | 77 | 46% | 15,369 |
| Netherlands | n.r.e. | 77% | 5 | 1,037 | 24[f] | 82% | 3,692 |
| Sweden | 82% | 82% | 1.5 | 950 | 140 | 51% | 2,300 |
| UK (England) | 61% | 76%[h] | 8 | 4,408 | 90 | 93%[f] | 15,049 |
| Average[e] | 75% | | 5 | | 134 | | |

*Regional data (regions with programme screening/ pilot projects)*

| Region | Coverage | | % Screen-positive | Annual number of smears (x1000) | Excess smears | | Women in target population[d] (x1000) |
|---|---|---|---|---|---|---|---|
| | 3-year | Recomm. Interval | | | 3-year (per 1000 women)[b] | Recomm. Interval (%)[c] | |
| Flandres (B) | 82% | 82% | 3 | 750 | 203 | 74% | 1,573 |
| Copenhagen (DK) | 75% | 75% | 5 | 70 | 174 | 70% | 165 |
| Bas-Rhin (F) | 69% | 69% | 5 | 104 | 178 | 77% | 255 |
| Ormylia (GR) | 71% | 71% | 5 | 4.6 | 117 | 50% | 13 |
| Florence (I) | 39% | 39% | 5 | 41 | 69 | 53% | 206 |
| Turin (I) | 70% | 70% | 10 | 90 | 99 | 42% | 271 |
| Midregion (P) | 37% | 37% | 5 | 61 | 86 | 69% | 292 |
| C. y. León (E) | 27% | 27% | 15 | 65 | 14 | 15% | 628 |

n.r.e. No reliable estimate. For France and Ireland, estimates are available for the coverage ever, 60% and 65% respectively. For definitions of the outcome parameters, see the text.

a Of Greece, Portugal and Spain, no national data are available

b E.g. for Germany: [17,000,000 - (33,000,000 * 80%(3-year coverage) / 3 years)] * 1000 / 33,000,000 = 248

c E.g. for Germany: 100 * [17,000,000/(33,000,000*50%(interval coverage)/1 year (interval))] - 100% = 3%

d For all countries or regions except Denmark, Finland and The Netherlands, the figure concerns the women in the age group corresponding with the target age-range (e.g. 25-64 years), instead of the upper age increased with the length of the recommended screening interval minus 1 year (e.g. 25-66, see Materials and Methods). The resulting underestimation of the target population is probably less than 5%.

e Unweighted average.

f Calculated with 5-year coverage. For Finland, using the rough estimate for the 3-year coverage of 61% would result in 228 smears per 1000 women per year. For the UK, using the 3 year coverage of 61% would result in 44% relative excess smear use

g Using the 60% (France) and 65% (Ireland) estimates for the ever screened women as if they are 3-year coverages.

h 5-year coverage.

## Model-based predictions: more detailed tabels

***Table A3***

*Predicted percentage reduction in incidence, mortality and life years lost by policy (100% coverage)*

| Country policy[a] | NL/FIN 30(5)60[7] | IRL 25(5)60[8] | UK(5) 20(5)65[10] | DK 23(3)59[13] |
|---|---|---|---|---|
| % incidence reduction | 75 | 80 | 85 | 84 |
| % mortality reduction | 76 | 78 | 86 | 80 |
| % life years lost reduction | 84 | 89 | 93 | 92 |

| Country policy[a] | S[b] 20(3)59[14] | B/F[c]/GR/I/E 25(3)64[14] | UK(3)/P 20(3)65[16] | G 20(1)72[53] |
|---|---|---|---|---|
| % incidence reduction | 84 | 87 | 90 | 96 |
| % mortality reduction | 80 | 86 | 88 | 95 |
| % life years lost reduction | 92 | 94 | 96 | 99.9 |

NL, The Netherlands; FIN, Finland; IRL, Ireland; DK, Denmark; S, Sweden; B, Belgium; F, France; GR, Greece; I, Italy; E, Spain; P, Potugal; G, Germany; UK, United Kingdom

[a] Starting age (interval) ending age [number of smears per women in a lifetime], policies ranked by increasing number of smears in a lifetime.

[b] For Sweden, the new guidelines issued 1998 recommend 23-60 years, with 3-year intervals 23-49 years and with 5-year intervals in women 50-60 years. For this policy, the predicted figures are 84, 80 and 92% respectively.

[c] For France the stopping age is 65 years.

***Table A4***

*Life-years gained per 1000 women assuming a 1% cumulative risk[a], by policy and coverage.*

| Country policy | NL/FIN 30(5)60[7] | | B/F[b]/GR/I/E 25(3)64[14] | | G 20(1)72[53] | |
|---|---|---|---|---|---|---|
| Participants participation | Systematic | Random | Systematic | Random | Systematic | Random |
| Interval-coverage | Life-years gained per 1000 women | | | | | |
| 25% | 18 | 31 | 20 | 47 | 21 | 77 |
| 50% | 36 | 51 | 40 | 68 | 43 | 83 |
| 75% | 53 | 64 | 60 | 76 | 64 | 84 |
| 100% | 71 | 71 | 80 | 80 | 85 | 85 |

This risk is the cumulative incidence in the situation without any (previous or current) screening. For explication of the participation pattern see Table A3. For abbreviations of countries see Table A3.

[a] The number of life-years gained are proportional to the cumulative risk for incidence: a two times higher risk results in two times higher number of life-years gained. Therefore, the results can be adjusted to any specific percentage cumulative risk by multiplification.

[b] For France the stopping age is 65 years.

***Table A5***

*Cost-effectiveness ratio (CER), expressed in number of smears per life year gained, assuming a 1% cumulative risk[a] and no excess smear use[a]. The CER is given by policy and coverage.*

| Country policy | NL/FIN 30(5)60[7] | | B/F[b]/GR/I/E 25(3)64[14] | | G 20(1)72[53] | |
|---|---|---|---|---|---|---|
| Participants participation | Systematic | Random | Systematic | Random | Systematic | Random |
| Interval-coverage | | | | | | |
| 25% | 88 | 53 | 152 | 68 | 516 | 146 |
| 50% | 88 | 63 | 152 | 92 | 516 | 266 |
| 75% | 88 | 74 | 152 | 120 | 516 | 392 |
| 100% | 88 | 88 | 152 | 152 | 516 | 516 |

For explication of the participation pattern see Table A3. For abbreviations of countries see Table A3.

[a] The number of smears per life-years gained are proportional to the inverse of the cumulative risk for incidence and proportinal to 1 + the interval related excess smear use (see table B): a two times higher risk results in two times lower number of smears per life-years gained, and a 100% excess smear use results in a 2 times higher number of smears per life-year gained. Therefore, the results can be adjusted to any specific cumulative risk and excess smear use.

[b] For France the stopping age is 65 years.

# 7

# PRESENT EVIDENCE ON THE VALUE OF HPV TESTING FOR CERVICAL CANCER SCREENING: A MODEL-BASED EXPLORATION OF THE (COST-)EFFECTIVENESS

## SUMMARY

Human papillomavirus (HPV) is the main risk factor for invasive cervical cancer.

High risks ratios are found in cross-sectional data on HPV prevalence. The question raised is whether this present evidence is sufficient for making firm recommendations on HPV-screening.

A validated cervical cancer screening model was extended by adding HPV infection as a possible precursor of cervical intraepithelial neoplasia (CIN). Two widely different model quantifications were constructed so that both were compatible with the observed HPV risk ratios. One model assumed a much longer duration of HPV infection before progressing to CIN and a higher sensitivity of the HPV test than the other. In one version of the model, the calculated mortality reduction from HPV-screening was higher and the (cost-)effectiveness was much better than for Pap smear screening. In the other version, outcomes were the opposite, although the cost-effectiveness of the combined HPV+ cytology test was close to that of Pap smear screening.

Although small follow-up studies and studies with limited strength of design suggest that HPV testing may well improve cervical cancer screening, only large longitudinal screening studies on the association between HPV infection and the development of neoplasias can give outcomes that would enable a firm conclusion to be made on the (cost-)effectiveness of HPV screening. Prospective studies should address women aged 30–60 years.

## INTRODUCTION

Molecular and epidemiological studies have clearly demonstrated that HPV is the main risk factor for cervical cancer (3,4). These epidemiological studies are case-control studies that consistently show a very high-risk ratio for HPV in women with (precursors of) cervical cancer compared with controls with negative cytology (6-9). The association between CIN and high-risk HPV infection is stronger in high-grade than in low-grade abnormalities (10-14) and is well over 90% in invasive cancers (10,15) . A few small follow-up studies also corroborate the crucial role of HPV infections: progression is found almost only in women with (persistent) high-risk HPV genotypes both in normal (16) and in dysplastic cases (17,18) . In a small retrospective study on archived false-negative smears from women with subsequent invasive cervical cancer, the high-risk HPV types found in the cancers were detected in nearly 100% of the preceding smears (19).

On the other hand, test-positive rates for high-risk HPV types in women over 30 years of age with normal cytology in North American and western European countries vary from 3% to 6% (7,16,20,21). This is much higher than can be explained by the life-time risk of developing cervical cancer in these countries.

For example, in the Netherlands, the rate for high-risk HPV types in women aged 30+ with normal cytology is around 4%, while the cumulative risk for invasive cervical cancer is around 1.5%; the risk in women aged 30+ with normal cytology is again much smaller. Therefore, only a fraction of the infections with high-risk HPV types will progress to cervical cancer.

The goal of this study was to incorporate the very high observed HPV-associated risks ratios in a cervical cancer screening model and to investigate the consequences for HPV screening as expressed in predicted mortality reduction, negative side-effects and costs. The outcome of the follow-up studies carried out to date have been incorporated in the model in so far that HPV infections were assumed to preceed HPV-infected neoplasias. They were not used for the quantification of the model as these studies were small or interpretation in quantitative epidemiological terms was limited by their design. The possible impact, however, will be discussed. The present study focusses on the question of whether recommendations about HPV screening can already be made on the basis of the available data and, if not, what type of data will be required to decrease uncertainty.

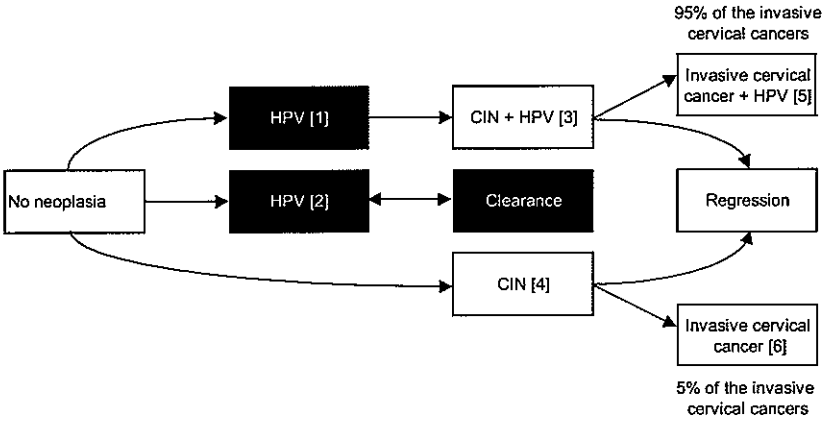
## MATERIALS AND METHODS

### The data

Test-positive rates for high-risk HPV types in women between the ages of 30 and 60 years were estimated on the basis of empirical data. Polymerase chain reaction (PCR-)based HPV-positive rates on cytological material of the cervix from women with negative cytology are 4% in the Netherlands (16), 5.7% in Portland, Oregon, USA (20), and 4.6% in Spain (7). PCR-based HPV-positive

*Figure 7.1*

*The stages and possible transitions in the HPV to CIN to invasive cervical cancer model. The disease stages that describe non-neoplastic conditions, and that have been added to the validated CIN to cervical cancer model, have been shaded*




rates on cytological material of women with a histologically confirmed diagnosis of CIN are 71% in Spain and 54% in Colombia (9), 75% in the USA (6), 72% in the UK (22) and 59% in the Netherlands (13). HPV rates are higher in high-grade than in low-grade lesions. Noting that the reported results are of the same order of magnitude, we summarized them by assuming 4% HPV positiveness in cytologically negative women and 67% in women with CIN. On the basis of the worldwide study on histological material of Bosch et al (15), we assumed that 95% of the invasive carcinomas were HPV infected, i.e. only 5% of invasive cervical cancers developed without being preceded by an HPV infection. In accordance with the results of the Dutch study (23), HPV-positive rates are assumed to be constant between 30 and 60 years of age.

## The model

Here, the relationship between HPV and cervical cancer in a stochastic microsimulation screening model is described. HPV in the model represents high-risk HPV types (16, 18, 31, 33, 35, 39, 45, 51, 52, 54, 56, 58, 59, 66, 68). As shown in Figure 7.1, the model is based on the hypothesis that the onset of HPV infections found in invasive cervical cancer and in CIN has preceded these neoplastic stages. Women who go through an HPV infection either become clear from the infection or develop HPV-infected CIN, which either regresses or progresses into HPV-positive invasive cervical cancer. Women can also develop CIN without an HPV infection, and this CIN again can regress or progress (only sometimes, see later) into invasive cancer. Allowing for the possibility that women can develop CIN (with or without HPV) after having become clear from HPV infection would cause a shift between the several arms in the model, without affecting the model outcomes presented in

*Table 7.1*

*Parameter values in model version A and B on the duration of detectable preclinical stages and the sensitivity of the HPV test for these stages*

| | Model version A | Model version B |
|---|---|---|
| Duration of stages (years) | | |
| HPV[1] that will develop into CIN+HPV | 10 | 1 |
| HPV[2] that will be cleared | 1 | 10 |
| CIN (with or without HPV)[3][4] | 11.8 | 11.8 |
| Invasive cancer (with or without HPV)[5] [6] | 3.9 | 3.9 |
| Sensitivity of HPV test (%) | | |
| HPV[1][2] | 100 | 50 |
| CIN+ HPV[3] | 100 | 80 |
| Invasive cancer+ HPV[5] | 100 | 87.5 |

[ ] Refers to the numbering of the disease stages in Figure 7.1

this article; therefore we did not complicate the model in this manner. This model is an extension of a validated cervical cancer screening Pap smear model (*5,24,25*).

According to this model, the average duration of CIN is 11.8 years and pre-clinical invasive cancer is 3.9 years (see Table 7.1). The sensitivity of the Pap smear is 80% in CIN and 87.5% in preclinical invasive carcinoma. These estimates on duration and sensitivity were derived from the British Columbia (Canada) screening data (*26*) and were compatible with data on interval cancers collected by the IARC (*27,28*). The incidence of progressive CIN was chosen to reproduce cervical cancer incidence and mortality in the Netherlands between 1965 and 1992. The regression rate was 72% of disease onset under 35 years, 40% between the age of 35 and 54 and very low in women aged 54 and over. These estimates resulted from subtracting progressive CIN from the age-specific CIN detection rates observed in the Dutch population (*29*). When adding HPV infection to the model, the part describing CIN and invasive cervical cancer was kept unchanged; the predicted CIN and cervical cancer incidences and prevalences were not affected. Consequently, previous validations are still valid. The incidence in the Dutch population accounted for is lower than incidences, for example, in the UK and the USA (7.8 and 12.9 per 100 000 for the Netherlands and the UK respectively, in 1978–1982 (*30*) and 9.9 in the USA in 1985 (*31*) ). The incidence level however did not influence the comparison of screening strategies.

## Two model versions

Because only cross-sectional HPV data were available for the quantification of the model, there was an identification problem for the parameters describing HPV infections. Test-positive rates in women screened for the first time are a result of incidence × duration × sensitivity. In view of this non-identifiability, we decided to construct two model quantifications that were contrasting in HPV-screening outcomes.

We varied duration and sensitivity and adjusted the incidence level to the observed test-positive rates for HPV. The longer the duration of progressive (to CIN) HPV infections (stage HPV [1] in Figure 7.1) and the higher the sensitivity of the HPV test, the more effective HPV screening will be in

*Table 7.2*

*Sensitivity by test (or combination of tests), stage and model version resulting from the values for sensitivity of the HPV test given in Table 7.1*

| Stages | Any model version | Model version A | | Model version B | |
|---|---|---|---|---|---|
| | Cytology only | Cytology + HPV | HPV only | Cytology + HPV | HPV only |
| HPV[1][2] | 0 | 100 | 100 | 50 | 50 |
| CIN + HPV[3] | 80 | 100 | 100 | 96 | 80 |
| CIN[4] | 80 | 80 | 0 | 80 | 0 |
| Invasive cancer + HPV[5] | 87.5 | 100 | 100 | 98.4 | 87.5 |
| Invasive cancer [6] | 87.5 | 87.5 | 0 | 87.5 | 0 |

[ ] Refers to the numbering of the disease stages in Figure 7.1

reducing cervical cancer mortality. In order to minimize the negative side-effects (i.e. follow-up of HPV-positive women who will not develop cervical neoplasia), it is favourable to assume a short duration of harmless (non-progressive) HPV infections (stage HPV [2] in Figure 7.1).

In model quantification A (see Table 7.1), the extra duration of the detectable preclinical phase because of HPV detection was assumed to be 10 years. The assumed sensitivity for HPV was 100% at all stages. Long duration and high sensitivity made model version A very favourable for HPV screening. In version B of the model, the detectable preclinical phase was only 1 year longer than in Pap smear screening, and sensitivity for high-risk HPV types was considerably lower than in version A. In HPV-infected neoplasia stages, sensitivity of the HPV test was equal to the sensitivity of the Pap smear (80% in HPV-positive CIN and 87.5% in HPV-positive invasive cancer), and sensitivity was only 50% in HPV infections without neoplasia. Compared with model A, model B was very unfavourable for HPV screening. The consequences of the two sets of assumptions for the sensitivity of the test (or combination of tests) are given in Table 7.2.

As a result of differences in sensitivity of the HPV test, the HPV test-positive rate of scrapes in invasive cervical cancer cases was (100% sensitivity × 95% invasive cervical cancers with preceding HPV infections =) 95% in model A and (87.5% × 95% =) 83% in model B. A high rate is in accordance with some PCR studies on cytological material of women with invasive cervical cancer (up to 100% (*10*)), but a lower rate has been found in other studies (e.g. 84% (*8*)).

## Simulated compared with observed HPV test-positive rates

In both model versions, predicted HPV test-positive rates in the age group 30–60 years was 4.01% in women with negative cytology and 67% in women with CIN.

## Consequences of true-positive test results

In the simulation, women with only negative tests at screening had a future screening after the regular screening interval. Women with positive cytology were followed up and in true-positive cases this led to the detection of neoplasia.

Women with a negative Pap smear and a positive HPV test were assumed to be followed up with HPV tests and Pap smears every six months. This follow-up stopped either when the HPV infection was cleared (after which women go back to screening) or when there was a transition of the HPV infection to HPV infected CIN (the neoplasia is detected). Detected CIN was assumed to be managed so that no invasive cancer would develop. For the management of CIN (diagnosis, treatment and after treatment check-ups), we accounted for 4 years of follow-up. This was in accordance with current practice in the management of CIN, at least in the Netherlands (*1*).

## Consequences of false-positive test results

Women with borderline (ASCUS) or low-grade abnormalities in their Pap smears in The Netherlands, and also in many other countries, are followed up with repeat smears. Some of these women have negative repeat smears and are referred back for routine screening. Women with high-grade abnormalities in their Pap smears are referred to the gynaecologist. In a proportion of these women, no neoplasia is found. As the model was adjusted for histologically confirmed detection rates, these so called 'false-positive' cytological outcomes have to be accounted for separately. We made the following assumptions:

- Five percent of the screening smears generated two repeat smears in women that did not have neoplasia.
- Five per 10,000 screened woman without CIN were referred to the gynacologist (32).

## The costs of screening

In order to account for the costs and savings of early detection, the costs of screening, follow-up, diagnosis and treatment were considered (see Table 7.3). The true resource costs were assessed for the screening Pap smear, the HPV test, colposcopy and radiotherapy. Costs charged in the Netherlands for the other medical procedures were used. The costs are presented in Dutch Guilders, for which the US$ exchange rate during 1995 was, on average 1.61.

## Screening strategies

In both model versions, the effects and costs have been calculated for several screening strategies for women between the ages of 30 and 60 years. We made predictive calculations for 3-yearly cytology and for six alternative strategies. Within these alternative strategies, we considered two screening test (or combination of tests) and three screening schedules. The screening tests were: cytology plus HPV test and HPV test only. In the three screening schedules, women were screened between 30 and 60 years of age: every 3 years (11 screenings per woman), every 5 years (seven screenings per woman) and every 10 years (four screenings per woman).

*Table 7.3*

*Assumptions on the costs by type of procedure, in Dfl*

| Procedure | Costs | Costs in the sensitivity analyses |
|---|---|---|
| Screening PAP smear[a] | 70 | |
| Repeat PAP smear[a] | 100 | |
| HPV test[a] | 90 | 45/155 |
| PAP smear and HPV test in one screening session[a] | 135 | 90/200 |
| Follow-up session in HPV-positive women with negative cytology | 140 | 280 |
| Diagnostic work-up of the referral whem no neoplasia is found | 800 | |
| Management of CIN[b] [1] | 3100[c] | |
| Curative primary treatment | | |
| microinvasive carcinoma | 9500 | |
| IB invasive carcinoma | 20200 | |
| II+ invasive carcinoma | 19100 | |
| Care for advanced disease[2] | 30700 | |

[a] Including Dfl 25 in total for costs for carrying out the smear/scrape and the costs for the women (time and transport)[5]
[b] CIN with or without HPV infection
[c] Including the costs of 15% recurrence of disease after primary treatment of CIN

### The cost-effectiveness calculations

Calculations were made for a cohort of women who attended all screenings.

Effects, costs and savings of the screenings were accounted for from birth to death. Outcomes were presented per 1000 women and have not been discounted.

## RESULTS

### Mortality reduction, years in follow-up and cost-effectiveness

The model predictions of the main effects and costs of the different combinations of frequency and types of screening tests are summarized in Table 7.4. For each of the two model versions and for each of the two alternative screening tests (cytology plus HPV test and HPV test alone), only the policy with the lowest screening frequency that had the same or higher mortality reduction compared with 3-yearly Pap smear screening is presented.

According to the model version A, which was favourable for HPV-screening, the combined test (cytology plus HPV test), even if performed only once every 10 years, reduced mortality more (91% vs 79%) than 3-yearly Pap smears. Costs were 37% lower, mainly because of the less frequent screening, and costs per life-year gained decreased by 41%. The number of years in follow-up was 26% lower, and the years in follow-up per life-year gained decreased by 27%. For 10-yearly screening with the HPV test only, mortality reduction was also higher than for 3-yearly cytology and only a little lower (89% vs 91%) than for the combined test. The costs for HPV only were very low, only 31% of the costs of 3-yearly Pap smear screening. Costs per life-year gained were 69% lower. The number of life-years spent in follow-up was less than half (because the repeat smears of the

*Table 7.4*

*Model outcomes: effects and costs of different screening policies in women between 30 and 60 years of age, two model versions. Only the least frequent HPV screening strategies with the same or higher mortality reduction compared to 3-yearly Pap smear screening are presented. All figures are per 1000 women screened, except for percentages (in brackets)*

| | Any model version | A Model version | | B Model version | |
|---|---|---|---|---|---|
| | Cytology only 3- yearly[a] | Cytology + HPV 10-yearly[a] | HPV only 10-yearly[a] | Cytology + HPV 5-yearly[a] | HPV only 3-yearly[a] |
| Favourable effects | | | | | |
| Morality reduction (%) | (79) | (91) | (89) | (80) | (76)[b] |
| Life-years gained [n(%)] | 65 (88) | 68 (93) | 66 (90) | 66 (89) | 62 (85) |
| Unfavourable effects | | | | | |
| Years in follow up | 700 | 520 | 290 | 1760 | 1790 |
| Costs (in Dfl x 1000) | | | | | |
| Screening | 650 | 460 | 300 | 800 | 830 |
| Follow up of HPV-positive cases | - | 60 | 60 | 361 | 470 |
| Follow up of false positive cytology[c] | 95 | 35 | 0.2 | 65 | 1.5 |
| Diagnosis and treatment | | | | | |
| CIN | 180 | 120 | 80 | 170 | 140 |
| Invasive and advanced cancer | -190 | -220 | -210 | -195 | -185 |
| Total costs | 740 | 460 | 230 | 1200 | 1250 |
| Ratios (per life-year gained) | | | | | |
| Years in follow-up | 11 | 8 | 4 | 27 | 29 |
| Costs | 11400 | 6800 | 3500 | 18300 | 20100 |

[a] At primary screening interval
[b] Using the HPV test only, according to model B, one would have to screen more frequently than 3-yearly to result in at least the same mortality reduction as 3-yearly cytology
[c] At screening and during follow-up of HPV-positive cases

borderline cytology do not occur in screening for HPV), and this also counts for the number of life-years in follow-up per life-year gained.

The results of model version B, which was unfavourable for HPV screening, were quite different. Combined screening performed every 5 years yielded a slightly higher mortality reduction (80% vs 79%, it was predicted at 77% with 10-yearly combined screening) than screening with cytology every 3 years, and was 63% more costly, resulting in 60% higher costs per life-year gained. The number of years in follow-up were 2.5 times higher, as were the number of years in follow-up per life-year gained. In the predictions for screening with the HPV test alone, even a 3-yearly interval did not result in a mortality reduction as high as with 3-yearly Pap smear screening (the 1 year extra detectable phase for which sensitivity is 50% is outbalanced by the 5% progressive lesions that are not detectable because they are HPV negative). Costs per life-year gained and years in follow-up per life-year gained were 1.8 and 2.6 times as high respectively.

Based on the model version A calculations, a decision might be made to replace Pap smear screening with HPV screening with a longer interval. This would lead to a greater mortality reduction at lower costs in terms of resources and negative side-effects. However, the model version B calculations suggest that Pap smear screening should not be replaced by any of the studied HPV screening strategies; costs and negative side-effects increased, while prevention of mortality did not improve.

*Table 7.5*

*Sensitivity analysis: costs per life-year gained with alternative cost assumptions, as percentage difference with the costs per life-year gained of 3-yearly cytology*

| | Any Model version | A Model version | | B Model version | |
|---|---|---|---|---|---|
| | Cytology only 3-yearly | Cytology + HPV 10-yearly | HPV only 10-yearly | Cytology + HPV 5-yearly | HPV only 3-yearly |
| Baseline cost assumptions[a] | 11400 | 6800 | 3500 | 18300 | 20100 |
| | | -40 | -70 | +60 | +80 |
| Alternative cost assumptions[b] | | | | | |
| HPV test, Dfl 45 | | -60 | -90 | +25 | +20 |
| HPV test, Dfl 155 | | -10 | -40 | +110 | +160 |
| HPV follow up, Dfl 280 | | -30 | -60 | +110 | +140 |

[a] HPV test Dfl 90; HPV follow-up Dfl 140

[b] These changes in assumptions do not affect the costs per life-year gained of 11400 of 3-yearly cytology

### Sensitivity analyses

We also calculated the costs of HPV screening assuming that HPV-positive women with negative cytology would be followed up every 3 years instead of every 6 months. The resulting total costs of HPV screening were lower, in particular according to model B in which cost-effectiveness of the combined test was close to the cost-effectiveness of Pap smear screening. However, less intensive follow-up in HPV-positive women would, with current knowledge, not be an acceptable option.

Economies of scale play an important role in the costs of an HPV test. Our estimate was based on a situation with, on average, 12 000 PCRs per year per laboratory. If the testing was concentrated in fewer laboratories, the test would become cheaper. Moreover, new developments can cause an increase or decrease in the costs of routine HPV tests. Therefore, calculations were repeated under the assumption that the laboratory costs per HPV test of Dfl 65 were less than one-third, i.e. Dfl 20, or doubled to Dfl 130. The total costs per test, including the Dfl 25 for carrying out the smear/scrape consequently will be Dfl 45 and Dfl 155, respectively, for the HPV test and Dfl 90 and Dfl 200 for the combined test (Pap smear + HPV test). In our basic calculations, a follow-up session for HPV-positive women was restricted to an HPV test and a Pap smear. We repeated the calculations with twice the costs per follow-up session (Dfl 280 instead of Dfl 140). This would be approximately the costs incurred when a colposcopy is added. The results are summarized in Table 7.5.

Options that were more cost-effective than 3-yearly Pap smear screening remained more cost-effective and those that were less cost-effective also remained less cost-effective. The conclusions were, therefore, not affected by considerable changes in the assumptions about the costs of HPV screening.

## DISCUSSION

We produced two model versions that both explained the high observed risk ratios for high-risk HPV types in women with cervical neoplasia compared with women with normal cytology. In addition,

they were both compatible with the 'clearance' rates in repeated HPV tests observed in women with normal cytology. In model A, this clearance resulted from a short duration of harmless HPV infections. In model B, the low sensitivity of the HPV test explained why woman that were HPV positive at a first screening will often be HPV negative at the next one. The effects of HPV screening predicted by the two model versions widely differed. Hence, the high-risk ratios alone were inconclusive for the outcomes expected from HPV screening.

The first non-cross-sectional evidence for the crucial role of high-risk HPV infections for the development of cervical cancer has been found in observational follow-up studies. These studies show only progression to high-grade neoplasias in the presence of (persistent) HPV infections. This concerns women with normal (*16*) and abnormal (*17,18*) cytology. Although these studies are very important for showing that HPV infection precedes the (progression of) neoplasia, they are too small (*16*) or have an inadequate design (*17,18*) for assessing the duration between HPV infection and the development of CIN, and the sensitivity of the HPV test. Nevertheless, they suggest that the sensitivity for progressive HPV infections is high and, in that respect, they support our favourable model version A more than the unfavourable model B. This support emphasizes how worthwhile it is to carry out the required large prospective studies on the association between HPV and cervical neoplasia that hopefully will confirm the 'preliminary' findings.

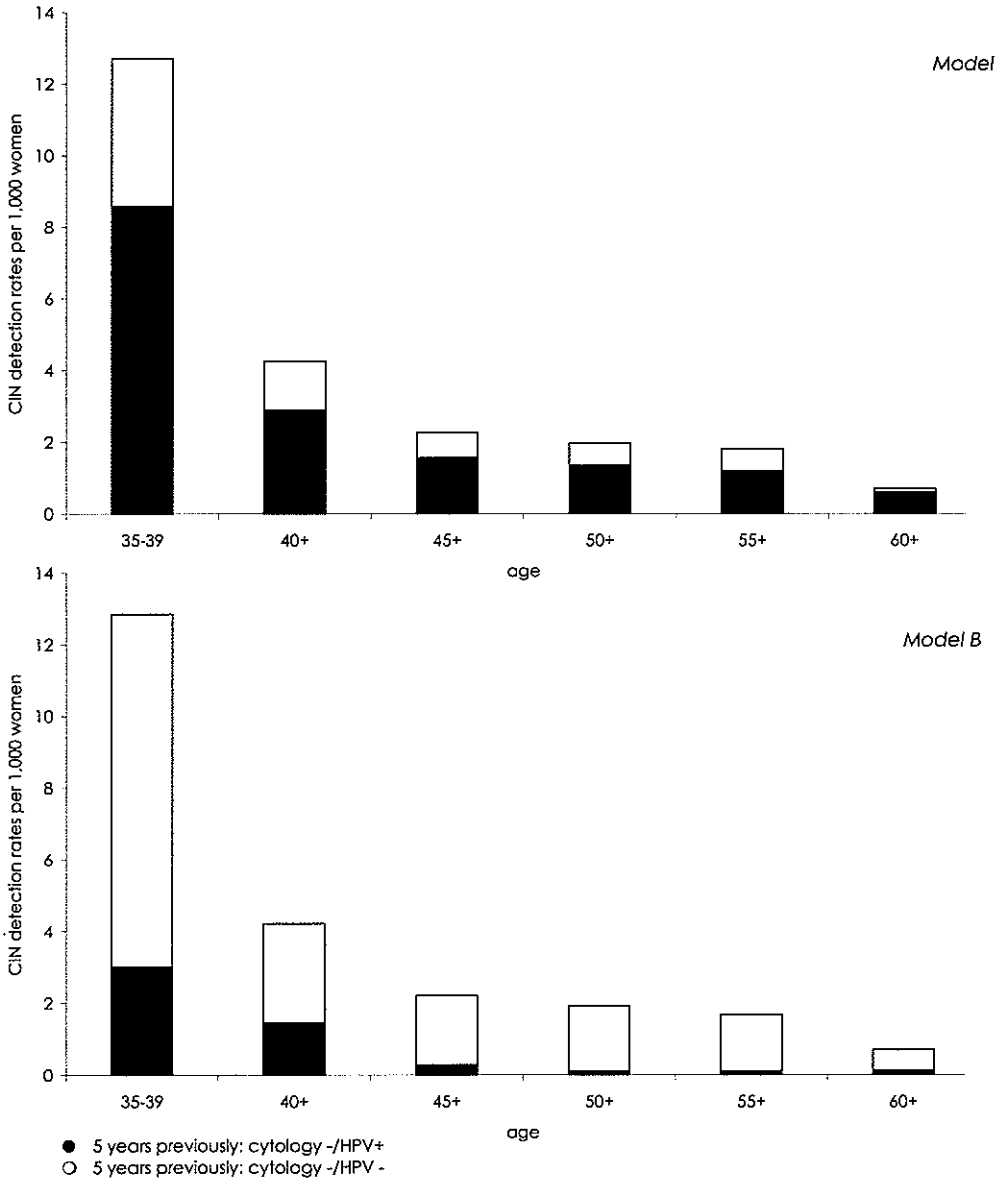
The presented disease model has a number of simplifications. It does, for example, not discern low-grade on high-grade pre-invasive lesions, while HPV-negative CIN cannot become HPV positive. These simplifications, however, are not important for the results, and model refinements will be of little help as long as adequate longitudinal data on HPV detection are not available.

The results of the cost-effectiveness calculations concerning the policies that combine HPV testing and Pap smear screening are complex and their outcomes could not have been predicted easily. For the calculations concerning policies using only the HPV test, it is not surprising that when it takes 10 years for HPV infections to produce CIN, HPV screening can improve Pap smear screening.

This is clearly not the case when HPV infection precedes CIN changes only by 1 year. But it is important to realize that these widely different assumptions are both compatible with the observed very strong association between HPV infection and cervical cancer, even if it is accepted that the HPV infection preceded the neoplastic changes that led to the invasive carcinomas. The work of Jenkins et al (*33*) , who also assessed the effectiveness of HPV testing as a primary screening tool by using a stochastic model, illustrates this issue. The authors did not vary the parameters that are crucial for the outcomes. They used assumptions on the sensitivity of the HPV test that were very similar to those in our model version A. In the sensitivity analysis, the simulated screening situation was further improved (by assuming that 100% of the cancers develop in the presence of high-grade HPV), but lower sensitivity was not tested. As far as duration is concerned, Jenkins' assumptions are intermediate to ours. Although the authors agreed that selection of the progression parameters (which determine the duration of stages) was not unique, they did not vary the progression rate of HPV infection and therefore did not describe the complete range of possible (cost-)effectiveness of HPV screening.

***Figure 7.2***

*Simulated results of a hypothetically observational study using model A and B: age-specific histologically confirmed CIN detection rates at Pap smear screening in women who 5 years previously had had a negative Pap smear, by HPV status 5 years previously and age group at present screening*




To explore the impact of longitudinal data, we simulated an observational cohort study with the two model versions A and B. In the simulation, women who entered the study with negative cytology have a Pap smear 5 years later.

Predicted CIN detection rates in women who at entry were HPV negative and those who were HPV positive were discerned (see Figure 7.2). As the description of cervical neoplasia (CIN and invasive

cervical cancer) of the model was the same in both model versions, the detection rate for CIN at Pap smear screening 5 years after negative cytology was the same. In version A, however, almost 70% of the women with histologically confirmed CIN (low and high grade) came from previously HPV-positive women, whereas in model version B this was only 20%. This reflects a higher predictive value for future CIN of a positive HPV test in version A. The fact that longitudinal outcomes clearly differ in both models means that different longitudinal outcomes can be consistent with present cross-sectional data, and that, once such longitudinal data are available, at least one (and probably both) of models A and B can be rejected. The range of combinations of parameter values on duration of HPV infections and sensitivity of the HPV test that are compatible with observed data will strongly decrease, and better predictions can be made of results expected from HPV screening.

Although the cross-sectional data show a strong association between HPV and cervical neoplasia, the results are insufficient to arrive at recommendations on screening. The discussion, therefore, on the representativeness of the test-positive rates that we aimed at in our simulation (4% in cytologically negative women, 67% in women with CIN and from 83% to 95% in women with invasive cervical cancers) is premature. Nonetheless, it is interesting to assess the influence of lower or higher observed HPV test-positive rates. In women with invasive cancer, the higher HPV positiveness, the better this will be for the effectiveness of HPV screening. Higher test-positive rates in women with normal cytology and in women with CIN, however, can only mean that more women who do not develop cervical cancer will be HPV positive (all women that will develop HPV-positive cervical cancer are already assumed to be HPV positive before the development of the cancer). These women will unnecessarily be detected and followed up, and the negative side-effects and cost of follow-up will increase. In other words, given that HPV infection precedes, for example, 95% of the progressive neoplasias, lower HPV prevalence in the cytologically negative women and in women with CIN implies less harmless and less costly HPV screening.

A modelling approach, as presented in this paper, is useful for a joint analysis of cross-sectional, longitudinal and other relevant epidemiological data. We will adjust our model as soon as new evidence becomes available.

Data from large PCR-based cohort studies will accumulate in the forthcoming years. The fact that many of them are solely focussed on young women should be of major concern. The Copenhagen study (*14*) is restricted to women under 30 years of age, and the median age of the women in the Portland study is 34 years (*20*). Screening for HPV in very young women would cause many women to be followed-up (because of the high prevalence in this age group of HPV infections that will clear) and is therefore not advisable. Moreover, the fact that prevalence is so much higher in younger age groups is also an expression of a different natural history of the HPV infections (at least a higher clearance rate) in this age group. Follow-up results from these women are obviously not transferable to the older age groups. Hence, further cohort studies should aim at women aged 30–60 years.

# REFERENCES


(1) van Ballegooijen M, Koopmanschap M, Habbema J. The management of cervical intraepithelial neoplasia: extensiveness and costs. Eur J Cancer 1995;28A:1703-8.

(2) van Ballegooijen M, Koopmanschap MA, Tjokrowardojo AJ, van Oortmarssen GJ. Care and Costs For Advanced Cervical Cancer. Eur J Cancer 1992;28A:1703-8.

(3) IARC Working Group. IARC monographs on the evaluation of carcinogenic risks to humans. Human papilloma viruses. Lyon: IARC; 1995.

(4) Zur Hausen H. Molecular pathogenesis of cancer of the cervix and its causation by specific human papillomavirus types. In Zur Hausen H, editor. Human Pathogenic Papillomaviruses. Berlin: Springer-verlag; 1994.

(5) Koopmanschap MA, Lubbe KT, van Oortmarssen GJ, van Agt HM, van Ballegooijen M, Habbema JDF. Economic aspects of cervical cancer screening. Soc Sci Med 1990;30:1081-7.

(6) Morrison EA, Ho GY, Vermund SH, Goldberg GL, Kadish AS, Kelley KF, et al. Human papillomavirus infection and other risk factors for cervical neoplasia: a case-control study. Int J Cancer 1991;49:6-13.

(7) Munoz N, Bosch FX, de Sanjose S, Tafur L, Izarzugaza I, Gili M, et al. The causal link between human papillomavirus and invasive cervical cancer: a population-based case-control study in Colombia and Spain. Int J Cancer 1992;52:743-9.

(8) Eluf-Neto J, Booth M, Munoz N, Bosch FX, Meijer CJ, Walboomers JM. Human papillomavirus and invasive cervical cancer in Brazil. Br J Cancer 1994;69:114-9.

(9) de Sanjose S, Munoz N, Bosch FX, Reimann K, Pedersen NS, Orfila J, et al. Sexually transmitted agents and cervical neoplasia in Colombia and Spain. Int J Cancer 1994;56:358-63.

(10) Van den Brule AJ, Walboomers JM, du Maine M, Kenemans P, Meijer CJ. Difference in prevalence of human papillomavirus genotypes in cytomorphologically normal cervical smears is associated with a history of cervical intraepithelial neoplasia. Int J Cancer 1991;48:404-8.

(11) Bergeron C, Barrasso R, Beaudenon S, Flamant P, Croissant O, Orth G. Human papillomaviruses associated with cervical intraepithelial neoplasia. Great diversity and distinct distribution in low- and high-grade lesions. Am J Surg Pathol 1992;16:641-9.

(12) Lorincz AT, Reid R, Jenson AB, Greenberg MD, Lancaster ED, Kurman RJ. Human papillomavirus infection of the cervix: relative risk associations of fifteen common anogenital types. Obstet Gynecol 1992;79:328-37.

(13) Gaarenstroom KN, Melkert P, Walboomers JM, Van Den Brule AJ, Van Bommel PF, Meyer CJ, et al. Human papillomavirus DNA and genotypes: prognostic factors for progression of cervical intraepithelial neoplasia. Int J Gynecol Cancer 1994;4:73-8.

(14) Kjaer SK, van den Brule AJ, Bock JE, Poll PA, Engholm G, Sherman ME, et al. Human papillomavirus--the most significant risk determinant of cervical intraepithelial neoplasia. Int J Cancer 1996;65:601-6.

(15) Bosch FX, Manos MM, Munoz N, Sherman M, Jansen AM, Peto J, et al. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. International biological study on cervical cancer (IBSCC) Study Group. J Natl Cancer Inst 1995;87:796-802.

(16) Rozendaal L, Walboomers JM, van der Linden JC, Voorhorst FJ, Kenemans P, Helmerhorst TJ, et al. PCR-based high-risk HPV test in cervical cancer screening gives objective risk assessment of women with cytomorphologically normal cervical smears. Int J Cancer 1996;68:766-9.

(17) Ho GY, Burk RD, Klein S, Kadish AS, Chang CJ, Palan P, et al. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J Natl Cancer Inst 1995;87:1365-71.

(18) Remmink AJ, Walboomers JM, Helmerhorst TJ, Voorhorst FJ, Rozendaal L, Risse EK, et al. The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease: natural history up to 36 months. Int J Cancer 1995;61:306-11.

(19) Walboomers JM, Husman AM, Snijders PJ, Stel HV, Risse EK, Helmerhorst TJ, et al. Human papillomavirus in false negative archival cervical smears: implications for screening for cervical cancer. J Clin Pathol 1995;48:728-32.

(20) Bauer HM, Hildesheim A, Schiffman MH, Glass AG, Rush BB, Scott DR, et al. Determinants of genital human papillomavirus infection in low-risk women in Portland, Oregon. Sex Transm Dis 1993;20:274-8.

(21) Cuzick J, Szarewski A, Terry G, Ho L, Hanby A, Maddox P, et al. Human papillomavirus testing in primary cervical screening. Lancet 1995;345:1533-6.

(22) Cuzick J, Terry G, Ho L, Hollingworth T, Anderson M. Type-specific human papillomavirus DNA in abnormal smears as a predictor of high-grade cervical intraepithelial neoplasia. Br J Cancer 1994;69:167-71.

(23) Melkert PW, Hopman E, van den Brule AJ, Risse EK, van Diest PJ, Bleker OP, et al. Prevalence of HPV in cytomorphologically normal cervical smears, as determined by the polymerase chain reaction, is age-dependent. Int J Cancer 1993;53:919-23.

(24) Koopmanschap MA, van Oortmarssen GJ, van Agt HMA, van Ballegooijen M, Habbema JDF, Lubbe JTN. Cervical-cancer screening: attendance and cost-effectiveness. Int J Cancer 1990;45:410-5.

(25) van Ballegooijen M, Habbema JD, van Oortmarssen GJ, Koopmanschap MA, Lubbe JT, van Agt HM. Preventive Pap-smears: balancing costs, risks and benefits. Br J Cancer 1992;65:930-3.

(26) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.
(27) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J 1986;293:659-64.
(28) van Oortmarssen GJ, Habbema JD, van Ballegooijen M. Predicting mortality from cervical cancer after negative smear test results. Bmj 1992;305:449-51.
(29) PALGA (Dutch network and national database for pathology). Results of retrieval action on cervical cytology from 1987-1990. Utrecht: SIG/PALGA; 1992.
(30) Jensen OM, Esteve J, Moller H, Renard H. Cancer in the European Community and its member states. Eur J Cancer 1990;26:1167-256.
(31) Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of eighteen major cancers in 1985. Int J Cancer 1993;54:594-606.
(32) PALGA (Dutch network and national database for pathology). Results of retrieval action on cervical cytology from 1977-1994. Utrecht: SIG/PALGA; 1995.
(33) Jenkins D, Sherlaw-Johnson C, Gallivan S. Can papilloma virus testing be used to improve cervical cancer screening? Int J Cancer 1996;65:768-73.

# 8

# EXTENDED DURATION OF THE DETECTABLE STAGE BY ADDING HPV TEST IN CERVICAL CANCER SCREENING

## SUMMARY

The HPV test could improve the (cost-) effectiveness of cervical screening by selecting women with a very low risk for cervical cancer during a long period. An analysis of a longitudinal study suggests that women with a negative Pap smear and a negative HPV test have a strongly reduced risk of developing cervical abnormalities in the years following the test, and that HPV testing lengthens the detectable stage by 2 to 5 years, compared to Pap smear detection alone.

## INTRODUCTION

One of the possible uses for the human papillomavirus test (HPV) is in primary cervical cancer screening in addition to or instead of the current Pap smear (*1-3*).

Introduction of HPV screening should be based on established (cost-) effectiveness. The (cost) effectiveness of HPV testing is primarily determined by the duration of the detectable preclinical stage (the period from the HPV infection to clinical disease), and the sensitivity and costs of HPV testing. To estimate preclinical duration and sensitivity, longitudinal studies on the association between HPV infection and the development of neoplasias are necessary. Several large longitudinal screening studies have started, but no long term results have been reported yet, although smaller longitudinal studies have been published (*4-7*). These studies differ with respect to HPV test used, age range of women, study design, and cytological or histological endpoint, which complicates the comparison and interpretation of these results. It is therefore too early for definite answers on the value of HPV testing in primary screening (*8-10*). But available data can be explored to derive preliminary estimates for parameters that determine the cost-effectiveness of HPV testing. This study investigates the duration of the detectable preclinical stage using the results of Rozendaal et al. (*7*). For these estimates, the 5-year cumulative incidence of cervical intraepithelial neoplasia (CIN) III after a negative Pap smear, the current screening interval in the Netherlands, is compared with the cumulative incidence within a doubled screening interval of 10 years (*3*) in case of a negative Pap smear and a negative HPV test.

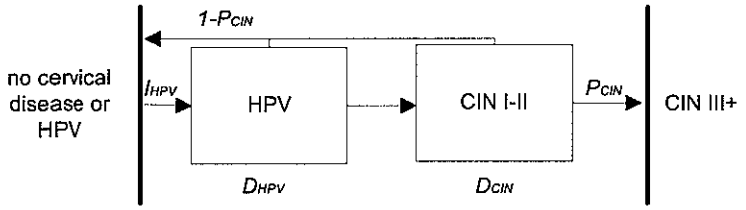
## MATERIAL AND METHODS

The study population, and screening and follow-up results are described by Rozendaal et al. (*6,7*). Briefly, the smears obtained during routine screening from 1988 to 1991 from a cohort of 2250 women aged 34-54 years, that were either normal or that showed borderline nuclear changes were tested for high risk HPV. The women were followed during a mean period of 6.4 years, using screen-detected (histologically confirmed) CIN III as endpoint. Among the 2129 (95%) women with a negative HPV test at baseline, one case of CIN III was diagnosed at a following screening round. Of the 121 women with a positive HPV test result at baseline (5%), 12 women with CIN III were detected later. This resulted in a relative risk of 210, with a 95% confidence interval from 27 to 1600.

The disease model used in this study is schematically presented in Figure 8.1. Women without cervical disease or HPV may become infected with HPV. This infection may clear, or it may progress to low grade CIN. From 'low grade CIN', the disease may regress spontaneously or progress to high grade CIN (corresponding with CIN III), the endpoint of the model. We assume that CIN III cannot develop without HPV infection.

*Figure 8.1*

*Schematic representation of the natural history model for HPV and CIN*




$I_{HPV}$ incidence rate of HPV infections in women without cervical disease or HPV
$P_{CIN}$ probability that HPV will progress to CIN III
$D_{HPV}$ duration of HPV infection preceding CIN
$D_{CIN}$ duration of CIN I/II preceding CIN III

In the model we assumed a constant duration of the HPV infection, and an exponentially distributed duration of low grade CIN with a mean of 6 years. The incidence rate of HPV infections in the age group considered (34-54 years) was set at 5 per 1,000 woman years (11).

Using the results of the Rozendaal study, the duration of the HPV infection and the probability that the HPV infection will progress to CIN III were estimated. On the basis of these estimates it was possible to calculate the cumulative incidence of CIN III within 5 years after the smear was taken, the current screening interval in the Netherlands, per 1,000 cytologically negative women and the cumulative incidence of CIN III within 10 years after the smear was taken per 1,000 cytologically negative/HPV negative women.

Initially, it was assumed that there were no diagnostic errors, i.e. the results of the HPV test and the Pap smear as found by Rozendaal et al. (7) reflected the true disease stage of the women. We used this as our reference model. In alternative models, we studied the consequences of assuming diagnostic errors. We also varied the assumptions on the incidence rate of HPV infections and the duration of low grade CIN. The mathematical description of the model is given in the Appendix.

## RESULTS

In Table 8.1, the results of the reference and alternative models are shown. Using the reference values for the model parameters, the duration of HPV infection before progressing to CIN was estimated at 3.8 years, resulting in a lower cumulative incidence of CIN III in 10 years for women with double negative screening results, than in 5 years after a negative Pap smear and an unknown HPV result.

Next, we dropped the perfect-test assumptions (100% sensitivity) of the HPV test and Pap smear separately, by assuming 50% sensitivity for detecting an HPV infection and 50% sensitivity for detecting CIN I+ respectively. Furthermore, we halved and doubled our assumptions on the incidence rate of HPV infections and the duration of low grade CIN. The estimated range for the duration of HPV before progressing to CIN widened, from 2 to 5 years. Only where the HPV test was assumed to have a sensitivity of 50% for HPV infections that will progress to CIN does the cumulative

*Table 8.1*

*Estimated values for the duration of HPV, the probability that the HPV infection will progress to CIN III, the 5-year cumulative incidence in women with a negative smear and the 10-year cumulative incidence in women with a negative smear and a negative HPV test.*

| | Duration HPV (years) | Probability HPV progresses to CIN III | 5-year cum. incidence CIN III after cyt - (per 1,000 women) | 10-year cum. Incidence CIN III after cyt-/HPV- (per 1,000 women) |
|---|---|---|---|---|
| Reference model[a] | 3.8 | 0.19 | 4.1 | 2.2 |
| Sensitivity HPV test for HPV 50%[b] | 2.2 | 0.08 | 4.6 | 5.2 |
| Sensitivity cytology for CIN I+ 50% | 3.5 | 0.16 | 10.0[c] | 4.0‡ |
| Incidence HPV infections 0.0025 | 2.3 | 0.17 | 4.5 | 1.4 |
| Incidence HPV infections 0.010 | 4.7 | 0.20 | 3.8 | 3.7 |
| Mean duration CIN I/II 2 years | 4.3 | 0.11 | 4.5 | 2.2 |
| Mean duration CIN I/II 10 years | 3.6 | 0.27 | 4.0 | 2.2 |

[a] Reference model: mean duration CIN I/II 6 years, sensitivity HPV test for HPV infection 100%, sensitivity cytology for CIN I+ 100% and incidence HPV infections 0.005 per woman year

[b] Assuming that the one woman that developed CIN III in (7) after HPV negative test at baseline did not have a false-negative HPV test result.

[c] Due to a limited sensitivity of cytology for CIN III only part of these lesions will be detected within the follow-up period considered.

incidence 10 years after a double negative result become slightly higher than the incidence within 5 years after a negative Pap smear.

## DISCUSSION

Our analysis shows duration of the HPV infection before it will progress into CIN of 2 to 5 years. For Pap smear screening, the preclinical duration was the combined duration of CIN and micro invasive cervical cancer, a period estimated at 15 years on average (12-16). Consequently, adding the HPV test to primary screening leads to duration of the detectable preclinical stage of almost 20 years in women aged 34-54 years. Furthermore, the 10-year cumulative incidence in women with a negative Pap smear and HPV test was lower than the 5-year cumulative incidence in women with a negative Pap smear and an unknown HPV result. This high negative predictive value of CIN III in double negative women is the result of a longer preclinical duration and a better selection of women, as women with double negative test results are at lower risk of cervical cancer than women with only negative Pap smear results of whom part will have HPV infections.

These results suggest that an HPV test in combination with the Pap smear, can considerably lengthen the screening interval in double negative women (3).

The confidence interval around the relative risk found by Rozendaal et al. (6) was large (from 27 to 1600). The upperbound of the confidence interval results in an HPV infection duration of 5.5 years. Assuming a relative risk corresponding to the lowerbound, results in a negligible duration. Even then, the 10-year cumulative incidence in double negative women is lower than the 5-year cumulative incidence in women with a negative cytological result, assuming no diagnostic errors. This results

from the very low risk of double negative women of becoming infected with HPV infected and subsequently developing cervical abnormalities in the years following the test. More firm estimates will be obtained on the basis of the results of the ongoing longitudinal studies.

The relative risks found in other studies (4,5), respectively 10.0 and 12.7, are lower than the range studied here. One of the reasons may be that the other studies concern women aged around 20 years. In young women, the occurrence of HPV infections is high (17) and a much higher proportion of these infections are transient (18) compared to older women. Therefore, adding the HPV test in primary screening is not useful for young women (19).

With the current model it is technically not possible to lower the sensitivity of the HPV test and Pap smear simultaneously. Doing this will probably result in an estimate for the duration of the HPV infection of around 2 years. Also, the assumption of a constant duration of the HPV infection can be dropped using a more sophisticated model. However, these refinements pay off only when adequate longitudinal data on HPV detection are available for quantification of the additional parameters.

Women may develop CIN III without first passing the stages CIN I (and even CIN II) (personal communication G.D. Zielinski, C.J.L.M. Meijer). This situation has been represented by assuming a relatively short average duration of low grade CIN of 2 years. Together with the assumption that this stage is exponentially distributed among women, this leads to a situation in which part of the women will develop CIN III shortly after having no neoplasia. Under these assumptions, the duration of the HPV infection before progressing to CIN is estimated to be relatively long, and the selection of low risk women by adding HPV to cytology will be even better (incidence of CIN III 2.2 versus 4.5, Table 8.1).

The endpoint of the model was CIN III as imposed by the data. Invasive cancer is the endpoint to be preferred as prevention of invasive cancer, and therefore death, is aimed at by cervical cancer screening. This endpoint, however, does not yield sufficient power due to the low risk for invasive cancer in Pap smear screened women, unless extremely large and long-term trials are performed.

The current estimate on the duration of HPV before developing CIN is a combined estimate for the duration of HPV for women who will have a regressive CIN III lesion and those that will progress to cervical cancer. To solve the uncertainty on the confounding of regressive CIN III lesions, this prospective analysis with CIN III as endpoint should be accompanied by archival studies, in which retrospectively the HPV status of smears preceding a diagnosis of cervical cancer, is assessed. Zielinski et al. (16) concluded in a retrospective study of 57 women with invasive cervical cancer that the detectable preclinical stage could be prolonged by at least 2 years by adding HPV testing, which corroborates our results. This type of study, however, is susceptible to confounding biases such as selection and length time bias, which may result in an underestimation of the extension of the detectable preclinical stage as cervical cancers found after participation in a screening programme may be selective towards fast growing cancers.

Doubling the screening interval for double negative women will result in cost savings, as half of the screening rounds can be omitted. If, for example, the effectiveness of 10-yearly combined

screening is the same as 5-yearly screening using the Pap smear, the costs of adding the HPV test to the Pap smear must be lower than these savings to be at least a cost-equal alternative. For a full cost-analysis, other costs and savings should also be taken into account, such as the costs of follow up in HPV positive/cytologically negative women and possible savings due to a decrease in detection of regressive cervical lesions because of a longer screening interval.

In conclusion, adding the HPV test to cytology in primary screening for cervical cancer results in an additional duration of the detectable preclinical stage of 2 to 5 years. Consequently, the screening interval for women with cytological and HPV negative test results may be considerably lengthened. These results remain to be confirmed by the large longitudinal studies that are currently underway.

## REFERENCES


(1) Cuzick J, Beverley E, Ho L, Terry G, Sapper H, Mielzynska I, et al. HPV testing in primary screening of older women. Br J Cancer 1999;81:554-8.

(2) Cuzick J. Human papillomavirus testing for primary cervical cancer screening. Jama 2000;283:108-9.

(3) Meijer CJLM, Walboomers JMM. Cervical cytology after 2000: where to go? J Clin Pathol 2000;53:41-3.

(4) Ho GY, Palan PR, Basu J, Romney SL, Kadish AS, Mikhail M, et al. Viral characteristics of human papillomavirus infection and antioxidant levels as risk factors for cervical dysplasia. Int J Cancer 1998;78:594-9.

(5) Liaw KL, Glass AG, Manos MM, Greer CE, Scott DR, Sherman M, et al. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. J Natl Cancer Inst 1999;91:954-60.

(6) Rozendaal L, Walboomers JM, van der Linden JC, Voorhorst FJ, Kenemans P, Helmerhorst TJ, et al. PCR-based high-risk HPV test in cervical cancer screening gives objective risk assessment of women with cytomorphologically normal cervical smears. Int J Cancer 1996;68:766-9.

(7) Rozendaal L, Westerga J, van der Linden JC, Walboomers JM, Voorhorst FJ, Risse EK, et al. PCR based high risk HPV testing is superior to neural network based screening for predicting incident CIN III in women with normal cytology and borderline changes. J Clin Pathol 2000;53:606-11.

(8) Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. Am J Epidemiol 2000;151:1158-71.

(9) Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Jama 2002;287:2382-90.

(10) Mandelblatt JS, Lawrence WF, Womack SM, Jacobson D, Yi B, Hwang YT, et al. Benefits and costs of using HPV testing to screen for cervical cancer. Jama 2002;287:2372-81.

(11) Meijer CJLM. Human papillomavirus and cervical cancer screening: Epidemiological aspects of HPV infection of normal cervical epithelium in relation to the development of abnormal cytology [in Dutch]. Amsterdam: Department of Pathology, University Hospital Vrije Universiteit; 1997.

(12) Gustafsson L, Adami HO. Natural History Of Cervical Neoplasia: Consistent Results Obtained By an Identification Technique [See Comments]. Br J Cancer 1989;60:132-41.

(13) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.

(14) van Oortmarssen GJ, Habbema JD, van Ballegooijen M. Predicting mortality from cervical cancer after negative smear test results. Bmj 1992;305:449-51.

(15) van Oortmarssen GJ, Habbema JD. Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative pap smears. Int J Epidemiol 1995;24:300-7.

(16) Zielinski GD, Snijders PJ, Rozendaal L, Voorhorst FJ, van der Linden HC, Runsink AP, et al. HPV presence precedes abnormal cytology in women developing cervical cancer and signals false negative smears. Br J Cancer 2001;85:398-404.

(17) Melkert PW, Hopman E, van den Brule AJ, Risse EK, van Diest PJ, Bleker OP, et al. Prevalence of HPV in cytomorphologically normal cervical smears, as determined by the polymerase chain reaction, is age-dependent. Int J Cancer 1993;53:919-23.

(18) Evander M, Edlund K, Gustafsson A, Jonsson M, Karlsson R, Rylander E, et al. Human papillomavirus infection is transient in young women: a population-based cohort study. J Infect Dis 1995;171:1026-30.

(19) Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, et al. A systematic review of the role of human papillomavirus testing within a cervical screening programme. Health Technol Assess 1999;3:i-iv, 1-196.

# APPENDIX

## Mathematical formulae of the reference model

If we assume that

- incidence rate of HPV is $I_{HPV}$
- probability that HPV will progress to CIN III is $P_{CIN}$
- duration of HPV preceding CIN is $D_{HPV}$
- duration CIN I /II preceding CIN III is exponentially distributed with mean $D_{CIN}$

Then:

Cumulative incidence of CIN III after x years for women without cervical lesions and without HPV infection at baseline

$$C_{HPV-}(x) = I_{HPV} * P_{CIN} * \int_0^{x-D_{HPV}} (1 - e^{-y*1/D_{CINI-II}})dy$$

Cumulative incidence of CIN III after x years for women without cervical lesions but with HPV infection at baseline

$$C_{HPV+}(x) = I_{HPV} * P_{CIN} * \int_0^{x-D_{HPV}} (1 - e^{-y*1/D_{CINI-II}})dy + P_{CIN} * \frac{1}{D_{HPV}} * \int_{x-D_{HPV}}^{X} (1 - e^{-y*1/D_{CINI-II}})dy$$

Relative risk on CIN III in women not having a cervical lesions but being HPV infected compared to women without cervical lesion and HPV infection after x years

$$rr(x) = \frac{C_{HPV+}(x)}{C_{HPV-}(x)} = \frac{I_{HPV} * \int_0^{x-D_{HPV}} (1 - e^{-y*1/D_{CINI-II}})dy + \frac{1}{D_{HPV}} * \int_{x-D_{HPV}}^{X} (1 - e^{-y*1/D_{CINI-II}})dy}{I_{HPV} * \int_0^{x-D_{HPV}} (1 - e^{-y*1/D_{CINI-II}})dy}$$

Cumulative incidence of CIN III x years after not having a cervical lesion

$$I(x) = p_{HPV-} * C_{HPV-}(x) + p_{HPV+} * C_{HPV+}(x)$$

with $p_{HPV-}$ and $p_{HPV+}$ the percentage of women without and with HPV infection at baseline respectively

# 9

# DISCUSSION

# INTRODUCTION

In this chapter, first the research questions put forward in chapter 1 are answered. The present state of modelling in breast and cervical cancer is then described and priorities for further developments are given. Finally, several general conclusions and recommendations resulting from the work presented in this thesis are given.

# ANSWERING THE RESEARCH QUESTIONS

## I. What is the applicability of the MISCAN breast cancer screening model in new situations?

*The MISCAN simulation programme is well equipped for use in new situations. However, the breast cancer screening model must be carefully checked following any modifications to adapt this to new situations.*

The MISCAN simulation programme was developed for use in different situations. Models can be adapted to specific demography, epidemiology, screening performance and costs. The Dutch breast cancer screening model was adapted in this way for Germany (*1*), Catalonia (Spain) (2), Florence (Italy) (3) and the United Kingdom (*4*).
In chapter 2, the breast cancer screening programme in Navarra was evaluated by implementing the Navarrian characteristics. After these adaptations, however, the screening results in Navarra failed to be satisfactorily reproduced. Various alternative model assumptions on preclinical duration, sensitivity and the existence of a high-risk group could not resolve this problem. Further analysis led to the conclusion that the difficulties in reproducing the Navarrian screening results might also be due to biases in the data, such as overrepresentation of cancers with borderline malignancy at first screening.
The difficulties in Navarra indicate that after adapting the model to new situations, the model cannot be trusted blindly. It must be checked for its ability to reproduce the observed incidence, mortality and screening data of the situation under study. In our experience, both problems in the model and in the data may underlie difficulties in reproducing observed screening results. For example, the natural history structure imposed by the MISCAN model (and other models) may cause limitations in finding explanations for the observed data. The natural history is represented by a (semi) Markov process with discrete state space in which the states are defined according to the generally adopted tumour size categories. A distinction is made between screen-detectable preclinical stages and clinical stages, the latter reflecting the situation following diagnosis on the basis of symptoms (see figure 1.4). The sensitivity depends on the preclinical tumour stage and survival depends on the state in which the woman is diagnosed, either by screen detection or clinically. The natural history modelled this way is the result of both biological, behavioural and

technical factors. A biological factor is tumour growth. The patient- and doctor-delay in responding to breast cancer symptoms are behavioural factors. A technical factor is the smallest size at which a tumour can be detected. In transferring a model to a new situation, the biological factors should remain constant, while behavioural and technical factors may differ between countries/regions and over time. A model in which a more explicit distinction is made between biological, behavioural and technical factors is currently under study at our department. For the Navarrian situation, such a model might estimate which part the observed high detection rate in the first and low detection rate in the subsequent round results from of women and doctors responding relatively late to symptoms and which part from a relatively low technical quality of screening.

## II. How much mortality reduction can be expected in the first years after introduction of screening for breast cancer?

*The mortality reduction in the first years after the start of breast cancer screening will be small and depends, apart from performance of screening and treatment, on several factors such as the fastness in nationwide implementation of the programme. The effect of the screening programme on breast cancer mortality in the Netherlands was only expected to be distinguishable from random fluctuation and autonomous trends after more than 10 years.*

The effect of screening on breast cancer mortality has recently been questioned (*5,6*). Nevertheless, there is convincing evidence that screening does lead to breast cancer mortality reduction (*7-9*). This reduction will be small in the first years of a screening programme due to the time lag between early diagnosis and its effect on mortality. Therefore, these changes can only be distinguished from random fluctuation and trends in the observed mortality rates after a longer period.

For the Netherlands, taking into account the yearly number of breast cancer deaths and the gradual introduction of the screening programme between 1989 and 1997, we estimated that the probability of an effect of the nationwide breast screening programme being distinguishable from random fluctuation would be high in the age group 55-74 8 years after the introduction of the screening programme, that is, from 1997 onwards (Chapter 3). The observed mortality figures in the years 1997, 1998 and 1999 for the age group 55-74 are indeed significantly lower than the mortality in the period before screening, 1986-1988 (*10*). This may be a first indication of the effectiveness of the Dutch nationwide screening programme.

However, other factors may also have contributed to the observed decline in breast cancer mortality, such as changes in background incidence of breast cancer, the trend towards earlier diagnosis e.g. because of women's general higher awareness as regards symptoms, or improvements in treatment. This is illustrated by the 12% reduction in breast cancer mortality observed among the age group of 55-69 within seven years after the introduction of the NHS breast cancer screening programme in England and Wales (*11*). Taking into account the demographic

and epidemiological characteristics of England and Wales, we predicted that the mortality reduction induced by the screening programme at that time would be 8% (Chapter 3). The additional observed mortality reduction might therefore be due to improvements in treatment, such as the widespread adoption of tamoxifen during this period, but random fluctuation may also play a part. Combining individual data on breast cancer mortality, participation and treatment may enable an assessment of the separate effect of screening to be made (9).

## III. Which Pap smear based cervical screening policies are optimal with respect to cost-effectiveness?

*The optimal age range of cervical cancer screening policies increases from age 40-52 years for policies with two examinations during a woman's lifetime to age 20-80 years for policies with more than 20 examinations; the screening interval decreases from 12 to 1.5 years. The incremental costs per life-year gained increase tremendously with increasing number of scheduled examinations. In chapter 5, we estimated the incremental cost-effectiveness ratio to be US$ 6,700 per life-year gained for 2 scheduled examinations during a woman's lifetime and US$ 173,700 for 40 examinations.*

The decision on the number of scheduled examinations could be based on the level of incremental cost-effectiveness that is considered acceptable. For a policy maker, if the decision regarding a policy depends only on a maximal allowed value or threshold value for the incremental costs per life-year gained, then for reference values of $15,000, $30,000 and $60,000, screening policies with five, 10 and 20 scheduled examinations, respectively, and screening intervals of 9, 5 and 3 years, respectively, are optimal.

Inclusion of quality of life aspects will make screening programmes with a large number of scheduled examinations more unfavourable due to the higher number of false-positive test results and overdiagnosis of preclinical cervical lesions that otherwise would have regressed spontaneously.

## IV. What is the cost-effectiveness of the cervical screening practice in different European countries?

*The cost-effectiveness of cervical cancer screening in Europe varies widely, because of considerable variation in the recommended number of scheduled examinations ranging from 7 in the Netherlands and Finland to more than 50 in Germany. Even if the implemented or recommended screening policy has a favourable cost-effectiveness ratio, the cost-effectiveness of actual screening practice is usually much less favourable due to systematic non-attenders and to smears taken outside the target age range or taken after too short an interval.*

The relative risk after negative screening is calculated by dividing the observed incidence after negative screening by the incidence that would have occurred in a situation without screening. This background incidence, however, cannot be observed. Using the incidence in the screening situation will lead to underestimation of the background incidence, as screening will reduce cervical cancer incidence. Our study, and some of the countries participating in the IARC study, used the incidence from the period before screening became widespread. This, however, may also not be appropriate, due to secular changes in incidence over time.
In the Netherlands, cervical cancer mortality started to decrease before screening started. This also holds for the incidence and/or mortality in other European countries. Therefore, the relative risk may have been underestimated in both our study and the IARC study, resulting in too favourable assumptions used to assess the effectiveness of screening so far. If so, to keep the cost-effectiveness on level, fewer smears per women would be indicated.

## VI. What is the evidence for the (cost-)effectiveness of HPV testing in primary cervical cancer screening?

*Results of longitudinal studies published so far indicate that adding the HPV test to primary screening may lead to a considerably longer screening interval for women with both a negative Pap smear and a negative HPV test. These results remain to be confirmed by the ongoing large longitudinal studies. A full cost-effectiveness analysis, including the costs and effects of the follow-up of HPV positive and cytologically negative women, has to be performed before recommendations can be given on the introduction of the HPV test in primary screening.*

The HPV test has the potential to improve primary screening. The results in chapter 7 show that cross-sectional data on HPV allow no conclusions on the (cost-)effectiveness of the use of the HPV test in primary screening. It was concluded that only results of large longitudinal screening studies on the association between HPV infection and the development of neoplasias would enable firm conclusions on the (cost-)effectiveness of HPV screening. Since then, large longitudinal studies have started, for which no long-term results are yet available. In the meantime, however, several small longitudinal studies have been published. Results of these studies indicate that if the HPV test is added in primary screening, the screening interval for women with a negative Pap smear and a negative HPV test may be lengthened considerably to achieve equal effectiveness (Chapter 8). When the results are confirmed by the longitudinal studies, cost-effectiveness estimates can be used to assess the optimal length of the screening interval.

# THE USE OF MODELLING IN EVALUATION OF BREAST AND CERVICAL CANCER SCREENING

## Modelling in breast cancer screening: present state and future

Several sophisticated breast cancer models (22-34) have been developed for the evaluation of the cost-effectiveness of different screening policies for breast cancer. In ongoing breast cancer screening programmes, modelling is used in the monitoring of screening programmes by comparing the expected and observed short- and long term results. In the yearly reports of the National Evaluation Team for Breast Cancer screening in the Netherlands, short-term screening results such as detection rates and interval cancer rates were compared to the results predicted by MISCAN (*10,35-44*). It appeared that the detection rates of subsequent screenings and interval cancer rates were lower than predicted by the MISCAN model. Also, the stage distribution of cancers detected at screening was less favourable than expected. The new breast cancer model under development at our department, in which an explicit distinction is made between the biological, behavioural and technical factors, can be used in exploring explanations for these differences; false reassurance is an example of this (*45*). MISCAN predictions suggested that screening would have a significant influence on breast cancer mortality from 1997 onwards (*35*). Now that death statistics have become available for 1998, evaluation of the long term screening results, including breast cancer mortality reduction, can be initiated. Preparations for a detailed analysis of breast cancer mortality have been started (9).

Alongside the monitoring of the results of ongoing screening programmes, breast cancer models are used for investigation of the cost-effectiveness of new techniques for screening, diagnostics and treatment. A recent example is the investigation of the cost-effectiveness of using stereotactilly guided core-needle biopsies as a diagnostic procedure for nonpalpable breast lesions, instead of surgical excision biopsy following wire localisation (*46*). This change was predicted to improve quality of life and reduce the total costs of diagnosis and treatment of breast carcinoma at the expense, however, of a slight increase in breast cancer mortality because of a loss in sensitivity.

In the future, the cost-effectiveness of the use of digital mammography instead of the conventional analogue screen film mammography may become a modelling research topic. At this moment, we are merely at the beginning of the digital mammography era. Already, however, several promising features of digital mammography are becoming evident, such as better image quality, ease of image manipulation and archiving, and image availability.

Modelling is also of use to study the cost-effectiveness of lowering the starting age of breast cancer screening and of optimising screening in high risk groups for breast cancer. There is an ongoing debate on the effectiveness of breast cancer screening for women in their forties. Most of the randomised trials conducted with mammographic screening were not designed to provide insight into the effectiveness of screening under the age of 50. For these trials, subgroup analyses only were possible. These comprised only a small number of women and therefore yielded non-

significant and opposite outcomes. A trial is currently running in the United Kingdom to study the effect on breast cancer mortality of mammographic screening in women before age 50. However, the results of this trial will not become available until after 2010. Even if the effectiveness of screening women before age 50 is established, its cost-effectiveness must still be assessed. The cost-effectiveness in women under 50 is expected to be less favourable than for women above 50, due to a lower incidence and possibly lower sensitivity in the younger women.

Women with a familial history of breast cancer are at a high risk for breast cancer. The (cost-) effectiveness of intensive screening (with or without chemoprevention) of these women should be closely monitored taking into account new developments, and compared to prophylactic surgery, an option that should preferably be avoided. The use of magnetic resonance imaging (MRI) instead of mammography may be an alternative for these women (*47,48*). The (cost-) effectiveness of these developments is currently being monitored at our department with the use of MISCAN (*49*).

## Modelling in cervical cancer screening: present state and future

As described in Chapter 1, most of the current efforts in modelling cervical cancer screening concern the natural history of the HPV infection and possibilities for the use of the HPV test in cervical cancer screening (*50-54*). An effective and affordable prophylactic vaccine for HPV would change the discussion on cervical cancer prevention completely. The first results of an HPV type 16 vaccine in healthy women have recently been published (*55*).

The effectiveness of a vaccine will depend on the high risk genotypes included in the vaccine. Increasing the number of types is also likely to increase the cost of vaccination. Moreover, a vaccine against a limited number of genotypes could increase the absolute importance of other types in causing cervical cancer through a number of mechanisms (*56*). Further, a small part of the invasive cancers (<5%) develop without prior HPV infection or with an unknown high risk genotype. Therefore, a vaccination and screening programme may be combined; the vaccine will reduce the incidence of the disease and screening will detect the cases that are not prevented. The resultant decrease in incidence may, however, be detrimental to the cost-effectiveness of a parallel screening programme. Determining the optimum combination of vaccination and screening is a future challenge in cervical cancer modelling.

The largest part of adverse effects of cervical cancer screening occur in women with false-positive test results and in women detected at screening with preclinical cervical lesions that would have regressed spontaneously if not detected. False positive test results lead to additional smears, referrals for colposcopy and other diagnostic procedures. As no distinction can be made between progressive and regressive preclinical lesions, both will be treated, leading to overtreatment. In the optimisation of screening policies, these unfavourable side-effects should ideally be taken into account. Therefore, the cost-effectiveness should also take health related quality of life into account. However, no measurements of quality of life in the health states produced and prevented by cervical cancer screening have been reported to date. Tentative calculations using quality of life measurements related to corresponding breast cancer screening states show a

Chapter 5 shows that the incremental cost-effectiveness ratio of cervical cancer screening policies in Europe varies by more than a factor of 20, assuming a constant coverage among countries and no opportunistic screening.

In practice, the cost-effectiveness of cervical screening in different European countries is also influenced by the coverage and the number of excess smears, i.e. smears taken outside the target age range or taken after too short an interval without almost any additional effect (Chapter 6). Differences in coverage from 50% to 82% were observed. Moreover, non- participating women were found to be at a high risk for cervical cancer (*12-14*).

Differences in excess smear use from 3 to 130% were found. These differences partly compensated the differences in number of scheduled examinations, as differences in excess smears combined with the number of scheduled examinations, resulted in an up to fourfold difference between the various European countries with regard to the number of smears taken over the course of a woman's lifetime.

To improve the cost-effectiveness of screening programmes, measures should be taken to increase the coverage and decrease the number of excess smears. Also, when evaluating new screening techniques, whether or not these will influence these parameters should be taken into account. For example, self sampling for HPV may be a method to reach women that do not participate in cervical cancer screening programmes (*15*).

## V. What is the incidence after negative screening in the Netherlands and does this correspond to the results from a multicountry analysis on which most models for cervical cancer screening have based their assumptions?

*Dutch screening data show a low relative risk for cervical cancer incidence throughout a large number of years after negative screening using the Pap smear. This corresponds to the result from a multicountry analysis for the first years after negative screening. After 4 years, the relative risk observed in the multicountry study was higher than in the Netherlands. However, in both analyses the effectiveness of cervical cancer screening may have been estimated too favourably due to overestimation of the background incidence.*

The effectiveness of the Dutch screening practice in selecting screen-negative women, who are at a low risk for cervical cancer and require rescreens only after a certain period, is shown in chapter 4. Screen-negative women were found to have a reduced risk for more than 10 years.

This decrease in relative risk depends on the duration of the preclinical stage and the sensitivity of the screening test. Therefore, estimating the relative risk after negative screening gives useful information for parameterisation and validation of a model. The duration and sensitivity estimates of most models in cervical cancer screening are validated against the IARC estimates of the relative risk (*16-19*), which are based on the results of eight screening programmes (*20,21*).

considerable influence on the effectiveness of cervical screening. Increasing the smear frequency may lead to negative incremental health effects (*57*). Careful assessment of the quality of life aspects of cervical cancer screening is therefore indicated. At our department, a research proposal has been developed to this end. Quality of life aspects should also be taken into account in the discussion on the introduction of the HPV test in cervical cancer screening. The use of the HPV test will generate a new group of HPV-positive women, which should receive follow-up. As cervical cancer is a rare complication of an HPV infection, a considerable number of unnecessary follow-up procedures will result, accompanied by the extra concern and anxiety for the women in question.

Pap smear screening quality depends on a series of successive actions, starting with smear taking, followed by processing the cellular material, evaluation of the smear and communicating the smear result. Improvements in each of these procedures will lead to improvement in the overall performance of Pap smear screening. Current developments focus on the improvement of the preparation of the cellular material (liquid based screening) and evaluation of the smear (computer aided screening). With the use of modelling, the trade off between sensitivity and specificity for these or other new methods can be expressed in costs per (quality adjusted) life year gained to support the decision making process on the introduction of these techniques (*58*).

## Uncertainty analysis

Uncertainty about the true numeric values of the model parameters (e.g. demographic, epidemiological and screening characteristics) is known as parameter uncertainty. To deal with parameter uncertainty, a sensitivity analysis or an uncertainty analysis can be performed.

In a univariate sensitivity analysis the values of the parameters are successively changed to assess the impact of uncertainty on the model outcomes. In a multivariate sensitivity analysis several parameters are involved simultaneously. In this way the combined influence on the outcome measures of changing these parameters can be investigated. The computation time increases, however, as a function of the number of uncertain parameters and the values considered for every parameter. When considering large numbers of parameters and values, uncertainty analysis may be more appropriate than sensitivity analysis. In uncertainty analysis, the uncertainty is specified by a multivariate probability distribution of the parameter values. Parameter values are repeatedly drawn from the multivariate probability distribution, and the model outcomes are calculated for that draw of parameter values.

*Table 9.1*

*Uncertainty analysis for cervical cancer screening in the Netherlands: frequencies (%) for which a policy is preferred to a subset of policies. A policy is preferred when: (i) its iCER does not exceed a given threshold value, and (ii) the iCERs of the other policies are smaller than or equal to the iCER of the preferred policy. For each threshold value the highest frequency is underlined. The frequencies in each column add up to 100%.*

| Screening policy: Number of invitations: first age (interval) last age | Threshold value of the iCER: | | |
|---|---|---|---|
| | 15,000 | 30,000 | 60,000 |
| ≤5° | <u>56.5</u> | 6.5 | |
| 6: 32 (7) 67 | 21.0 | 3.0 | |
| 7: 32 (6) 68 | 13.0 | 5.0 | |
| 8: 27 (7) 76 | 4.5 | 18.0 | 3.0 |
| 9: 27 (6) 75 | 2.0 | 15.5 | 3.0 |
| 10: 27 (5) 72 | 3.0 | <u>39.5</u> | 10.0 |
| 15: 22 (4) 78 | | 12.0 | <u>47.5</u> |
| 20: 20 (3) 77 | | 0.5 | 29.5 |
| 25: 20 (2.5) 80 | | | 5.5 |
| 30: 22 (2) 80 | | | 1.5 |
| 40: 20 (1.5) 78.5 | | | |

° Only a screening policy with five invitations was included 5:30(9)68.

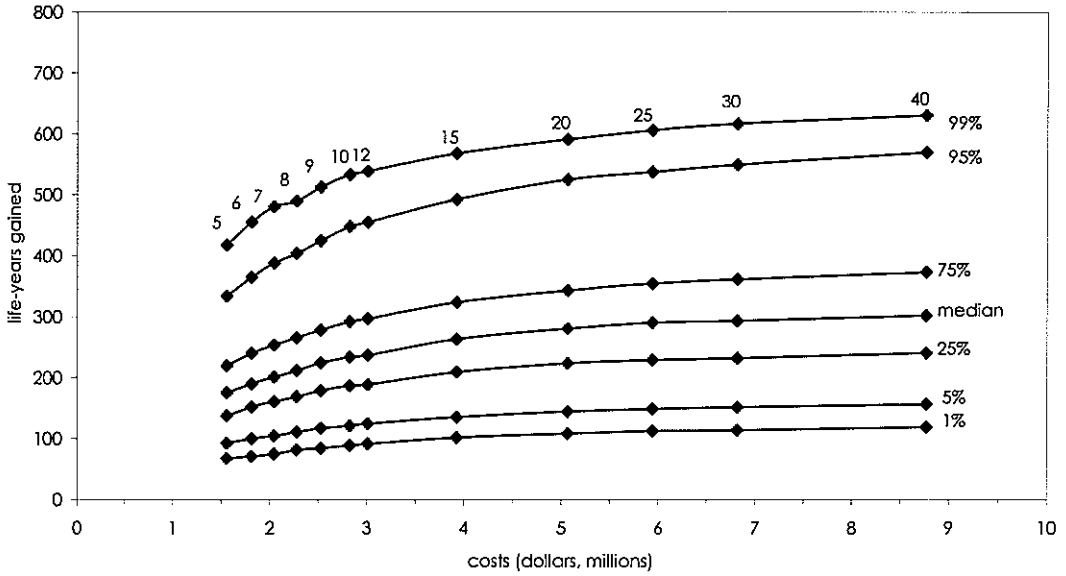
An uncertainty analysis was performed on the results of the cost-effectiveness analysis of cervical cancer screening described in chapter 5. Some of the model parameters could be specified quite precisely; for example, this was the case for the demographic parameters and the parameters describing the prognostic effect of treatment of non-invasive lesions. Other parameters were uncertain. For these parameters, the uncertainty around the point estimates that were used in Chapter 5 needed to be quantified. The parameters included in the uncertainty analysis are the lifetime risk for cervical cancer for the different birth cohorts, the proportion of progressive lesions occurring at higher ages, the mean duration of the disease stages, survival after clinical detection, attendance at screening, the relative risk for cervical disease in never-attenders, the sensitivity and the specificity of the screening procedure, the improvement in prognosis after screen detection of invasive preclinical cancers and the costs of palliative treatment in non-curable disease. If data were available, statistical analyses were performed to assess the uncertainty around the point estimate (for example the mean duration and sensitivity of the preclinical disease stages), and assumptions on the uncertainty were made by experts for the remaining parameters.

The costs and effects of the screening policies that were found to be optimal in Chapter 5, were predicted again, taking into account the parameter uncertainty. A large number of life histories were simulated to obtain these predictions, in order to reduce the influence of random fluctuation.

The results showed that the uncertainty about the parameters induces a large uncertainty in the predicted number of life-years gained, the 5% percentile being about 50% lower and the 95% percentile being 80% higher than the median number of life years gained as illustrated in figure 9.1. The impact of parameter uncertainty on the costs is much smaller, with a standard deviation of 9%

***Figure 9.1***

*Uncertainty analysis for cervical cancer screening in the Netherlands: percentiles of the empirical distribution of the life years gained, and the mean costs for the screening policies included. The number of scheduled examinations is indicated for each policy. Costs (US$ in millions) and effects (life-years gained) are per 1,000,000 women in the simulated general population per year of the screening program.*




for 5 invitations decreasing to 6% for 20 or more invitations, compared to a standard deviation of 35% for the number of life-years gained.

The uncertainty in choice of screening policies is shown in table 9.1 for various threshold values of the incremental cost per life year gained (iCER). The large uncertainty about the preferred policy for given threshold values is a direct consequence of the almost fourfold difference between the 5% and 95% percentiles in the number of life years gained. It is reassuring to note that the policies that are preferred with the highest frequency for the threshold values of $15 000 and $30 000 correspond to the optimal screening policies, without taking uncertainty into account, to wit the policies with 5 and 10 scheduled examinations (Chapter 5). Interestingly, at the $60 000 threshold value, a less intensive screening policy with 15 scheduled examinations was preferred most often when taking uncertainty into account, compared to 20 scheduled examinations when not taking uncertainty into account (Chapter 5).

Besides quantification of the uncertainty in the estimated cost and effects, the parameters that are the main cause of the uncertainty can be identified in an uncertainty analysis. In our study, most of the uncertainty was due to a small number of parameters. The most important parameter is the risk of developing cervical cancer if screening would not exist, especially for cohorts of women born after 1948. The second parameter is the incidence of progressive preclinical stages after the age of 55. This parameter especially influences the effectiveness of screening in older women. This kind of information can be used to select the parameters regarding which additional efforts to reduce this

uncertainty could be undertaken. If it is not possible to reduce the uncertainty for these parameters, the uncertainty must be accepted as inherent to the decision making process in that situation.

Besides parameter uncertainty, there is also uncertainty about the structure of the model, although these two types of uncertainty are sometimes closely associated. An example of model uncertainty concerns the asymptomatic, preclinical stage of cervical cancer. Several different hypothesised forms ('structures') are plausible given the available data. The only appropriate way to deal with uncertainty about the structure of the model is computation of the outcomes under each alternative structural assumption. Therefore, in designing a new model, the focus should not be on a single structure. A more general model, in which several structures of the process under study can be embedded, is better equipped to deal with model uncertainty.

## Validity of models

Screening recommendations should be based on validated models. For an example of the consequences of using insufficiently validated models for making recommendations, see Appendix 1. Recommendations should be made with care if uncertainty exists about the underlying process. For example, it has been suggested that women older than 50 years of age might be withdrawn from the cervical cancer screening programme in the United Kingdom, if they had a current negative smear together with a recent history of negative smears or together with a negative HPV test (*51*). The model used assumes that most invasive cancers occurring over age 50 will have already been detected by screening before age 50. Hence this model is a priori bound to predict only small increases in incidence when women are withdrawn from screening before the recommended age of 64. The authors acknowledge that the clinical course of disease in older women and the natural course of HPV infection are not well understood, which makes their recommendations speculative (see Appendix 2).

In general, validation of a model, consisting of demonstration of its face validity, and its internal and external validity, should take place before using the model for prediction and evaluation. Furthermore, after adaptation of a model to a new situation the model outcomes have to be checked against observed results.

## Communicating and understanding complex models

In scientific research, not only the results but also the methods used to obtain the results should be available in the public domain. By providing information about the model used, interested parties can perform an independent check of the model by reproducing the results using the method provided. Therefore, a full description of the model structure and input characteristics should be given. This may also diminish feelings of having to deal with a 'black box'.

In the past, scientists were dependent on the willingness of scientific journals to publish descriptions of the model used. Most medical journals only allowed a short description instead of a detailed

explanation of the model in question. Nowadays, a host of new possibilities to provide information are available, of which the website of the journal is by far the most preferred.

Two general descriptions of the MISCAN simulation programme have been published in scientific journals (*59,60*). Furthermore, detailed reports on the breast cancer and cervical cancer model used in the decision making process on the national screening programmes for breast and cervical cancer have been published in Dutch (*61-64*). More efforts should be made to communicate the MISCAN programme and disease models internationally. This may be done by putting the full information on a website.

Furthermore, discussions on the model and its outcomes should be encouraged with persons interested in modelling outcomes, such as clinicians, to solve miscommunications arising from the different professional backgrounds.

A very useful initiative in both respects is the 'model profiler' that has been designed and is currently being developed within the framework of the Cancer Intervention and Surveillance modelling NETwork (CISNET) initiated by the National Cancer Institute (USA) (http://cisnet.cancer.gov). The model profiler provides a unified way of disseminating specific model information, as well as a means for comparisons between models. This initiative enables detailed description of the model structure and the input parameters to be provided.

# CONCLUSIONS AND RECOMMENDATIONS

## Conclusions

- Validated (breast) cancer models should not be transferred blindly to other situations.
- The costs per life-year gained of cervical cancer screening programmes with a large number of scheduled examinations are considerably higher than in case of a small number of scheduled examinations.
- Evaluating the effectiveness of cervical cancer screening programmes is hampered by uncertainty on the incidence that would have occurred in a situation without screening.
- If an HPV test is used in addition to cytology, the screening interval may be considerably lengthened for women with negative test results for both cytology and HPV testing.
- Mathematical models are an important tool in the evaluation and monitoring of screening for breast and for cervical cancer.

## Recommendations

- A breast cancer model with a clear distinction between biological, behavioural and technical factors should be developed and tested in its ability to explain observed breast cancer screening results from different situations.

- Linked individual data on breast cancer mortality, participation in breast cancer screening and treatment should be available to separate the effect of breast cancer screening on the breast cancer mortality from that of using different treatment modalities.
- As soon as long-term follow-up data of large HPV screening trials are available a cost-effectiveness evaluation should be performed to determine the optimal role of the HPV test in cervical screening.
- Models should be validated before they are used for evaluation and prediction.
- Detailed information on models should be publicly available.

## REFERENCES


(1) Beemsterboer PM, de Koning HJ, Warmerdam PG, Boer R, Swart E, Dierks ML, et al. Prediction of the effects and costs of breast-cancer screening in Germany. Int J Cancer 1994;58:623-8.

(2) Beemsterboer PMM, Warmerdam P, Boer R, Borras JM, Moreno V, Viladiu P, et al. Screening for breast cancer in Catalonia, which policy is to be preferred? Eur J Public Health 1998;8:214-46.

(3) Paci E, Boer R, Zappa M, de Koning HJ, van Oortmarssen GJ, Crocetti E, et al. A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy. Eur J Cancer 1995;3:348-53.

(4) Boer R, de Koning HJ, Threlfall A, Warmerdam P, Street A, Friedman E, et al. Cost effectiveness of shortening screening interval or extending age range of NHS breast screening programme: computer simulation study. BMJ 1998;317:376-9.

(5) Gotsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355:129-34.

(6) Giard RWM, Bonneux LGA. Borstkankerscreening onvoldoende effectief. Ned Tijdschr Geneeskd 2001;145:2205-08.

(7) Duffy SW. Interpretation of the breast screening trials: a commentary on the recent paper by Gotsche and Olsen. The Breast 2001.

(8) Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002;359:909-19.

(9) de Koning HJ, Fracheboud J, Verbeek AL, Rutgers EJ, van der Maas PJ. De wetenschappelijke basis van het bevolkingsonderzoek naar borstkanker in Nederland. Ned Tijdschr Geneeskd 2002;146:1034-41.

(10) LETB (Landelijke Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): Fracheboud J, Otto SJ, Groenewoud JH, van Ineveld BM, Broeders MJM, Verbeek AL, et al. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (IX). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 2001.

(11) Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. BMJ 1995;311:1391-5.

(12) 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.

(13) Magnus K, Langmark F, Andersen A. Mass screening for cervical cancer in Ostfold county of Norway 1959-77. Int J Cancer 1987;39:311-6.

(14) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.

(15) Nobbenhuis MA, Helmerhorst TJ, van den Brule AJ, Rozendaal L, Jaspars LH, Voorhorst FJ, et al. Primary screening for high risk HPV by home obtained cervicovaginal lavage is an alternative screening tool for unscreened women. J Clin Pathol 2002;55:435-9.

(16) Eddy DM. The frequency of cervical cancer screening. Comparison of a mathematical model with empirical data. Cancer 1987;60:1117-22.

(17) Gustafsson L, Adami HO. Cytologic screening for cancer of the uterine cervix in Sweden evaluated by identification and simulation. Br J Cancer 1990;61:903-8.

(18) Gyrd-Hansen D, Holund B, Andersen P. A cost-effectiveness analysis of cervical cancer screening: health policy implications. Health Policy 1995;34:35-51.

(19) van Oortmarssen GJ, Habbema JD. Duration of preclinical cervical cancer and reduction in incidence of invasive cancer following negative pap smears. Int J Epidemiol 1995;24:300-7.

(20) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for squamous cervical cancer: duration of low risk after negative results of cervical cytology and its implication for screening policies. Br Med J 1986;293:659-64.

(21) IARC Working Group on Evaluation of Cervical Cancer Screening Programmes. Screening for cancer of the uterine cervix. Lyon; 1986.
(22) van Oortmarssen GJ, Habbema JDF, Lubbe JTN, van der Maas PJvd. A model-based analysis of the HIP project for breast cancer screening. Int J Cancer 1990;46:207-13.
(23) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JC, Collette HJ, Hendriks JH, 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.
(24) Chiacchierini RP, Lundin FE. Benefit/risk ratio of mammography. In Logan WW, editor. Breast carcinoma: The radiologist's expanded role. New York: Wiley; 1977. p. 15-28.
(25) Chiacchierini RP, Lundin FE. Risk-benefit analysis for reduced dose mammography. In Logan WW, Muntz EP, editors. Reduced dose mammography. New York: Masson Publishing; 1979.
(26) Chiacchierini RP, Lundin FE, Scheidt PC. A risk/benefit analysis by life table modeling of an annual breast screening program which includes X-ray mammography. In Nieburgs HE, editor. Prevention and detection of cancer. Part II: Detection, Volume 2. New York: Marcel Dekker; 1980. p. 1741-62.
(27) Shwartz M. A mathematical model used to analyze breast cancer screening strategies. Operations Res 1978;26:937-55.
(28) Knox EG. Simulation studies of breast cancer screening programmes. London: Oxford University Press; 1975.
(29) Eddy DM. A mathematical model on the efficacy of breast cancer screening. In Feig SA, McLelland R, editors. Breast carcinoma: Current diagnosis and treatment. New York: Masson Publishing; 1983. p. 339-49.
(30) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. Journal of Epidemiology and Biostatistics 1997;2:25-35.
(31) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the preclinical screen-detectable phase. Journal of Epidemiology and Biostatistics 1997;2:9-23.
(32) Day NE, Duffy SW. Trial design based on surrogate endpoints - application to comparison of different breast screening frequencies. J Royal Stat Soc, Series A 1996;159:49-60.
(33) Jansen JT, Zoetelief J. Assessment of lifetime gained as a result of mammographic breast cancer screening using a computer model. Br J Radiol 1997;70:619-28.
(34) Parmigiani G. Modeling in medical decision making. Chichester: John Wiley& Sons, Ltd; 2002.
(35) van den Akker-van Marle E, de Koning H, Boer R, van der Maas P. Reduction in breast cancer mortality due to the introduction of mass screening in The Netherlands: comparison with the United Kingdom. J Med Screen 1999;6:30-4.
(36) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): de Koning HJ, Boer R, Ineveld BM, van, Bruyn AE, de, Maas PJ, van der, Fracheboud J. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (II). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 1993.
(37) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): de Koning HJ, Boer R, Ineveld BM, van, Bruyn AE, de, Maas PJ, van der, Fracheboud J. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (III). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 1994.
(38) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): de Koning HJ, Beemsterboer PMM, Boer R, van Ineveld BM, Bruyn AE, de, Maas PJ, van der. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (IV). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 1995.
(39) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): Fracheboud J, Beemsterboer PMM, Boer R, van Ineveld BM, Colette HJA, Verbeek ALM. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (V). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 1996.
(40) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): de Koning HJ, Boer R, van Ineveld BM, Beemsterboer PMM, Verbeek ALM, Hendriks JHCL. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (VI). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 1997.
(41) LETB (Landelijke Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): Fracheboud J, Boer R, Verbeek ALM, Hendriks JHCL, Broeders MJM, Beemsterboer PMM, et al. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (VII). Rotterdam: Erasmus Universiteit Rotterdam: Instituut Maatschappelijke Gezondheidszorg; 1999.
(42) LETB (Landelijke Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): Fracheboud J, Groenewoud JH, Boer R, Broeders MJM, Baan CA, Verbeek ALM, et al. Landelijke evaluatie van bevolkingsonderzoek naar borstkanker in Nederland (VIII). Rotterdam: Erasmus Universiteit Rotterdam, Instituut Maatschappelijke Gezondheidszorg; 2000.
(43) LETB (Landelijk Evaluatie Team voor bevolkingsonderzoek naar Borstkanker): Groenewoud JH, Fracheboud J, Boer R, van Ineveld BM, de Bruyn AE, van der Maas PJ, et al. Landelijk

bevolkingsonderzoek naar borstkanker in Nederland. Optimalisatiestudie 1999-2001. Rotterdam/Nijmegen; 2002.
(44) Fracheboud J, de Koning HJd, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL. Nation-wide breast cancer screening in The Netherlands, results of initial and subsequent screenin 1990-1995. National Evaluation Team for Breast Cancer Screening. Int J Cancer 1998;6:132-8.
(45) Boer R, de Koning H, van Oortmarssen G, Warmerdam P, van der Maas P. Stage distribution at first and repeat examinations in breast cancer screening. J Med Screen 1999;6:132-8.
(46) Groenewoud JH, Pijnappel RM, van den Akker-van Marle ME, Birnie E, Buijs-van der Woude T, Mali WPTM, et al. Cost-effectiveness of large-core needle biopsy for nonpalpable breast lesions compared to surgical biopsy. submitted.
(47) Kuhl CK, Schmutzler RK, Leutner CC, Kempe A, Wardelmann E, Hocke A, et al. Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. Radiology 2000;215:267-79.
(48) Stoutjesdijk MJ, Boetes C, Jager GJ, Beex L, Bult P, Hendriks JH, et al. Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. J Natl Cancer Inst 2001;93:1095-102.
(49) Kriege M, Brekelmans CT, Boetes C, Rutgers EJT, Oosterwijk JC, Tollenaar RAEM, et al. MRI screening for breast cancer in women with familial or genetic predisposition: design of the Dutch National Study (MRISC). Familial Cancer 2001;1:163-8.
(50) van Ballegooijen M, van den Akker-van Marle ME, Warmerdam PG, Meijer CJ, Walboomers JM, Habbema JD. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. Br J Cancer 1997;76:651-7.
(51) Sherlaw-Johnson C, Gallivan S, Jenkins D. Withdrawing low risk women from cervical screening programmes: mathematical modelling study. Bmj 1999;318:356-60.
(52) Myers ER, McCrory DC, Nanda K, Bastian L, Matchar DB. Mathematical model for the natural history of human papillomavirus infection and cervical carcinogenesis. Am J Epidemiol 2000;151:1158-71.
(53) Mandelblatt JS, Lawrence WF, Womack SM, Jacobson D, Yi B, Hwang YT, et al. Benefits and costs of using HPV testing to screen for cervical cancer. Jama 2002;287:2372-81.
(54) Kim JJ, Wright TC, Goldie SJ. Cost-effectiveness of alternative triage strategies for atypical squamous cells of undetermined significance. Jama 2002;287:2382-90.
(55) Koutsky LA, Ault KA, Wheeler CM, Brown DR, Barr E, Alvarez RN, et al. A controlled trial of a human papillomavirus type 16 vaccine. N Engl J Med 2002;347:1645-51.
(56) Garnett GP, Waddell HC. Public health paradoxes and the epidemiological impact of an HPV vaccine. J Clin Virol 2000;19:101-11.
(57) van Ballegooijen M. Effects and costs of cervical cancer screening. Department of Public Health. Rotterdam: Erasmus University; 1998.
(58) Meerding WJ, van Ballegooijen M, Habbema JDF. Performance and cost-effectiveness of liquid based cytology. Histopathology 2002;41 (Suppl.2):494-501.
(59) Habbema JD, van Oortmarssen GJ, Lubbe JT, van der Maas PJ. The MISCAN simulation program for the evaluation of screening for disease. Comput Methods Programs Biomed 1985;20:79-93.
(60) Loeve F, Boer R, van Oortmarssen G, van Ballegooijen M, Habbema J. The MISCAN-COLON simulation model for the evaluation of colorectal cancer. Comput Biomed Res 1999;32:13-33.
(61) van der Maas PJ, van Ineveld BM, van Oortmarssen GJ, de Koning HJ, Lubbe JTN, Habbema JDF, et al. De kosten en effecten van bevolkingsonderzoek op borstkanker. Interimrapport. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit; 1987.
(62) Habbema JDF, Lubbe JTN, van Agt HME, van Ballegooijen M, Koopmanschap MA, van Oortmarssen GJ. Kosten en effecten van bevolkingsonderzoek op baarmoederhalskanker. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit; 1988.
(63) de Koning HJ, van Ineveld BM, van Oortmarssen GJ, Boer R, Collette HJA, Verbeek ALM, et al. De kosten en effecten van bevolkingsonderzoek naar borstkanker. Eindrapport. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit; 1990.
(64) van Ballegooijen M, Boer R, van Oortmarssen GJ, Koopmanschap MA, Lubbe JTN, Habbema JDF. Bevolkingsonderzoek op baarmoederhalskanker: leeftijdsgrenzen en intervallen. Een geactualiseerde kosten-effectiviteitsanalyse. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit; 1993.

# APPENDIX 1

## COMMENTS ON DUFFY/CHEN MARKOV CHAIN MODELS FOR PROGRESSION OF BREAST CANCER

## LETTER

Chen and colleagues recently published two articles on Markov chain models for investigation of the natural history of breast cancer in terms of tumour size, lymph node spread and histological grade(*1,2*). We think the models do not agree as well with observed data as has been suggested and that the effectiveness of shortening the screening interval for women in their forties is overestimated.

In the article on tumour attributes and the preclinical screen-detectable phase Chen *et al.* (*1*) check the appropriateness of their models by comparing the observed and expected number of cases for two of their models and by calculating the $\chi^2$ test statistic for these models. They conclude that there is a significant difference between observed and expected numbers only in the 50-59 age group in the model for size alone (M-S-2 model) and that the fit in the model for size and node status together (M-N-S model) is good. We disagree with this conclusion on several grounds. First, we do not understand the very low expected numbers of interval cancers and clinical cancers arising in the control group before screening with size < 2 cm and node involvement in the M-N-S model (Table 12). These expected numbers result from the low estimate for the 1 year transition probabilities from the initial states 'preclinical, N(-), < 2 cm' and 'preclinical, N(+),< 2cm' to the final state 'clinical, N(+), <2cm'. We think that these estimates do not give the maximum likelihood since higher values for these transition probabilities will improve the fit of the model considerably, as expected numbers of 0.00015 and 0.0013, compared with observed numbers of 8 and 23 respectively, clearly indicate that the model is not correct in this aspect. Second, we are even more concerned about the correctness of the use of the goodness-of-fit test, with respect to both the calculation of the $\chi^2$ test statistic and the assessment of the *p*-value. According to the expected and observed number of cases based on the M-N-S model (Table 12) we calculated the $\chi^2$ test statistic to be 62.4 and 833,544.3 instead of 31.1 and 25.4 for the age groups 40-49 and 60-69 respectively. Consequently, the null hypothesis that the observed data coincide with model expectations must be rejected for these age-groups.

We have serious doubts on the number of degrees of freedom used in determining the *p*-values. If the model and its parameters are specified, e.g. fitted on another independent data set, then, if *k* independent data categories are involved in the validation procedure, the test statistic has a $\chi^2$ distribution with *k* degrees of freedom. However, in this case the *r* model parameters are maximum-likelihood estimates based on individual data from the same data set as the *k* data categories used for validation. The distribution of the test statistic will then fall between a $\chi^2$ distribution with *k-r* degrees of freedom and a $\chi^2$ distribution with *k* degrees of freedom (*3,4*). Therefore the numbers of degrees of freedom are certainly not 10 and 20 for the M-S-2 model and the M-N-S model respectively, but fall between 5 and 10 for the M-S-2 model (five model parameters) and between 5 and 20 for the M-N-S model (15 model parameters). As a consequence, the observed and expected numbers are also significantly different for the 40-49 age group in the M-S-2 model, and the fit of the M-N-S model for the 50-59 age group becomes doubtful.

Due to the unsatisfactory validation results of the two presented models, we doubt the validity of the other models, for which the results of the goodness-of-fit test are not shown, and therefore the value of the conclusions concerning the process of dedifferentation.

In the accompanying paper the Markov chain models are used for prediction of outcomes for different screening regimes (2). The models predict remarkably low mortality reductions for a 3 year regime in the 40-49 age group, some models predict even higher mortality than in a situation without screening. The authors do not explain how the model can predict screening to cause breast cancer mortality, instead of the intended prevention of it. This is probably due to the fact that the model assumes a death rate starting at the moment of screen detection. In this way women are at risk of dying from breast cancer during their lead time, even if they would not have been diagnosed in a situation without screening. If this effect appeared in reality, it would provide a strong argument against screening. Without such detrimental effect, we expect the increase of efficacy by shortening the screening interval to be much more moderate than presented. Obviously, shortening the screening interval cannot possibly cause an increase in the percentage of mortality reduction which is more than proportional to the increase in the number of screenings performed.

Although modelling is a useful tool in the evaluation and planning of cancer screening it should be applied carefully and extensive validation of models is required.

## REFERENCES


(1) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the preclinical screen-detectable phase. Journal of Epidemiology and Biostatistics 1997;2:9-23.

(2) Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. Journal of Epidemiology and Biostatistics 1997;2:25-35.

(3) Mood AM, Graybill FA, Boes DC. Introduction to the theory of statistics. Third ed: Mc-Graw-Hill Book Company, Singapore; 1974.

(4) Kendall, Stuart. The Advanced Theory of Statistics, Vol. 2, Distribution Theory. New York: Hafner Publishing Company; 1961.

# APPENDIX 2

## WITHDRAWING LOW RISK WOMEN FROM CERVICAL SCREENING PROGRAMMES; CONCLUSIONS CANNOT YET BE DRAWN

## LETTER

Sherlaw-Johnson et al. evaluated policies for withdrawing women from the cervical cancer screening programmes before the recommended age of 64, using a mathematical model (*1*). Their results were obtained with specific and uncertain model assumptions, which were insufficiently subjected to validation and sensitivity analysis.

From the description of the model in cited earlier papers, most new cases of cervical intraepithelial neoplasia seem to originate at younger ages. The duration is assumed to be independent of age and very long on overage (50 years for cervical intraepithelial neoplasia grade III). This implies that most invasive cancers occurring over age 50 started as cervical intraepithelial neoplasia before age 50, which could thus be detected by screening before age 50. Hence this model is bound to predict only small increases in incidence when women are withdrawn from screening before the recommended age of 64.

The sensitivity analysis considers only small adaptations of this basic assumption. Other models, for which detailed analysis of screening data and data on the incidence of cancer was used, resulted in much lower estimates of the mean duration of cervical intraepithelial neoplasia (*2-4*). These models would predict less favourable effects of withdrawal policies.

Present data on human papillomavirus allow for widely different models, some of which are and some of which are not favourable for use in screening (*5*). Given this uncertainty, it is not yet possible to come to conclusions about the impact of withdrawing women from cervical screening programmes if results of their smear test and a simultaneous test for high risk types of human papillomavirus are negative.

## REFERENCES


(1) Sherlaw-Johnson C, Gallivan S, Jenkins D. Withdrawing low risk women from cervical screening programmes: mathematical modelling study. BMJ 1999;318:356-61.

(2) Brookmeyer R, Day NE. Two-stage models for the analysis of cancer screening data. Biometrics 1987;43:657-69.

(3) Gustafsson L, Adami HO. Natural history of cervical neoplasia: consistent results obtained by an identification technique [see comments]. Br J Cancer 1989;60:132-41.

(4) van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559-65.

(5) van Ballegooijen M, van den Akker-van Marle ME, Warmerdam PG, Meijer CJ, Walboomers JM, Habbema JD. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. Br J Cancer 1997;76:651-7.

# SUMMARY

Cancer is an important public health problem. In the Netherlands, the cancer mortality rate was 271 per 100,000 men and 218 per 100,000 women in the year 1998. Lung, prostate and colorectal cancer are the principal cancers in men, and breast, lung and colorectal cancer in women.
In the Netherlands screening programs for breast and cervical cancer were introduced in order to reduce morbidity and mortality due to these cancers.
A screening programme should only be introduced after its effectiveness has been established, preferably by randomised controlled trials or otherwise by evidentially convincing observational studies. Empirical studies give an answer to the value of a specific screening policy in a particular epidemiological situation. It is not possible to perform empirical studies for all of the alternatives. Modelling is used to extrapolate the results to other policies and situations.
Costs of a screening programme are also important. In a cost-effectiveness analysis, costs and effects of interventions are compared. The resulting cost-effectiveness ratio can be used to decide between different screening policies, and also for the comparison of screening with other health interventions.
Various possibilities for the use of mathematical models in the evaluation of cancer screening exist. Mathematical models can be used for 1) data analysis, 2) evaluation/monitoring and 3) planning/optimisation. In data analysis, models are used to test hypotheses about the natural history of the disease, characteristics of the screening test and the association between early detection and risk of dying from the cancer. In evaluation, the results of a screening programme are compared to the model outcomes. If they differ, the model can be used in exploring possible reasons for these differences. Using modelling for planning and optimisation results in cost-effectiveness estimates and optimal strategies for screening programmes for a specific situation.
In this thesis different mathematical models are used in studying screening for breast and cervical cancer. The microsimulation program MISCAN is frequently used. This model is developed at the Department of Public Health, Erasmus MC, for the evaluation of screening for cancer.

In chapter 2 the results of the breast cancer screening programme in Navarra (Spain) were compared with predictions from MISCAN, based on the demographic, epidemiological, and screening characteristics of Navarra. The detection rate in the first round was higher than predicted, while the rate in the subsequent round was lower than expected. Alternative assumptions could not satisfactorily explain the first and second round results together. Nevertheless, the annual mortality reduction was expected to range between 17 to 23%.
The effects on the breast cancer mortality will be small in the first years after the start of screening (Chapter 3). Accordingly, it was expected by the MISCAN model that the reduction in breast

cancer mortality due to the Dutch nationwide breast screening programme, which started around 1989, would only be visible from 1997 onwards. In England and Wales a 12% mortality reduction was reported within seven years after the introduction of screening. Modelling results showed that 70% of this percentage was expected to be attributable to screening, and the remaining part to improvements in treatment of breast cancer.

Cervical cancer screening uses the Pap smear as screening test. When abnormalities are detected, the woman is advised to have further diagnostic evaluations. When no abnormalities are found in the smear, the probability of developing cervical cancer in the years after the smear is very small. Comparing the number of cervical cancers after negative screening with the number of cervical cancers in a situation without screening in the Netherlands, indeed show a strongly reduced risk of cervical cancer for a period of more than 10 years following a negative smear (Chapter 4).

A precise estimate of this reduction is hampered through lack of information about the background incidence, i.e. the incidence in a situation without screening. This incidence will be different from the observed incidence in a situation with screening, as an effective screening programme for cervical cancer screening will reduce the incidence. It is shown that taking the incidence from the period before screening was introduced can result in an overestimate of the reduction in incidence due to cervical cancer screening, as there is evidence that the incidence has decreased independently of screening. Because this trend has often not been accounted for, several analyses may have overestimated the background risk and consequently the effect of screening.

In many countries in Europe, North America and Australia screening programmes for cervical cancer are running. However, recommended screening policies differ widely. The number of scheduled Pap smears per woman varies from seven in the Netherlands and Finland, compared to more than 50 Pap smears in Germany. The recommended interval between Pap smears is one year in Germany, while it is five years in the Netherlands and Finland. The age range in which Pap smears are taken varies accordingly. The starting age varies between 20 and 30 years, and the last screening age between 59 years and no upper age limit.

Using the MISCAN model, the costs and effects of almost 500 screening policies are determined (Chapter 5). The costs and effects of the different policies were compared to identify the efficient policies, i.e. no alternative policy exists that results in more life-years gained for lower costs. Fifteen efficient screening policies have been identified, considering two to 40 scheduled examinations. The incremental cost-effectiveness ratio increased from $ 6,700 per life-year gained (two scheduled examinations) to $ 173,700 per life-year gained (40 scheduled examinations).

Although for most countries the recommended screening policy was reasonably close to an efficient policy, the high number of smears scheduled per woman implied an unfavourably high cost-effectiveness ratio.

In practice the cost-effectiveness of screening is negatively influenced by deviation from the recommended strategy. Part of the women does not participate in screening. This is unfavourable

for the effectiveness of the screening programme, especially because non-participating women have a higher risk for cervical cancer. Other women have more smears during lifetime than recommended. These smears add little to the effectiveness, while the costs of screening will increase proportionally to the number of smears (Chapter 6).

The human papillomavirus (HPV) is the main causative factor for invasive cervical cancer. The use of HPV testing in cervical cancer screening is currently being investigated. In chapter 7 we show, with the use of the MISCAN model, that cross-sectional data do not allow firm conclusions on the (cost-) effectiveness of the HPV test. It was possible to construct two widely different model quantifications, one favourable for HPV screening and one unfavourable, which were both compatible with the observed HPV prevalence in women with and without cervical neoplasia. In the favourable model the duration of the HPV infection preceding the development of cervical cancer was assumed to be 10 years, while the unfavourable model assumes a duration of one year.
For conclusions on the (cost-)effectiveness of the use of the HPV test in cervical cancer screening, longitudinal data are needed. By now several large longitudinal HPV screening studies are underway, but their results are only available around 2005. In the meantime, we investigated the natural history of the HPV infection in combination with the (precursors of) cervical cancer using the results of a small longitudinal study. On the basis of this study (Chapter 8), the duration of the HPV infection that will progress into cervical neoplasia was estimated between two to five years.

As illustrated above, models are important in the evaluation of cancer screening. In this thesis, models were used for data analysis, evaluation and optimisation. However, there are some remaining topics concerning the use of models that need special attention. These are uncertainty analysis, the validity of models and the communication of complex models (Chapter 9).
When using models, it is important to investigate the consequences for the model outcomes caused by the uncertainty around the parameters used as model input. Therefore, uncertainty analysis can be used, in which the uncertainty around the parameters is specified and the influence on the model outcomes is assessed. Uncertainty analysis can also be used to identify the parameters that are the main cause of uncertainty. This kind of information can be used to select parameters regarding which additional research to reduce this uncertainty could be undertaken. Otherwise, the uncertainty must be accepted as inherent to the decision making process in that situation.
Before its use for prediction and evaluation, the models have to be validated adequately. Therefore, the structure of the model should make sense to people who have a good knowledge of the problem (face validity). Furthermore, the model should reproduce the data used to estimate the parameters for the model (internal validity) and the model predictions should compare well to empirical data that are not used for parameter estimation of the model, if available (external validity).

However, also validated models have to be checked against observed results after adaptation of the model to a new situation.

In scientific research, not only the results but also the methods used to obtain the results should be available in the public domain. In this way, interested parties can perform an independent check of the model. This may also diminish feelings of having to deal with a 'black box'. In the past, it was regularly difficult to have detailed model descriptions published in the scientific journals. Most medical journals only allowed a short description. Nowadays, a host of new possibilities to provide information are available, for example by putting information on a website. The website of the journal is by far the most preferred.

Conclusions

- Validated (breast) cancer models should not be transferred blindly to other situations.
- The costs per life-year gained of cervical cancer screening programmes with a large number of scheduled examinations are considerably higher than in case of a small number of scheduled examinations.
- Evaluating the effectiveness of cervical cancer screening programmes is hampered by uncertainty on the incidence that would have occurred in a situation without screening.
- Mathematical models are an important tool in the evaluation and monitoring of screening for breast and for cervical cancer.

Recommendations

- As soon as long-term follow-up data of large HPV screening trials are available a cost-effectiveness evaluation should be performed to determine the optimal role of the HPV test in cervical screening.
- Models should be validated before they are used for evaluation and prediction.
- Detailed information on models should be publicly available.

# SAMENVATTING

Kanker is een belangrijk gezondheidsprobleem. In 1998 overleden in Nederland 271 op de 100.000 mannen en 218 op de 100.000 vrouwen aan kanker. Long-, prostaat- en dikke darmkanker zijn de drie voornaamste kankers bij mannen en voor vrouwen zijn dit borst-, long- en dikke darmkanker.
In Nederland zijn landelijke bevolkingsonderzoeken naar borst- en baarmoederhalskanker geïntroduceerd met als doel de morbiditeit en mortaliteit ten gevolge van deze kankers te reduceren.
Voordat een dergelijk screeningsprogramma geïntroduceerd kan worden, moet de effectiviteit van een programma aangetoond zijn, bij voorkeur met behulp van gerandomiseerde studies of anders met duidelijk overtuigende observationele studies. Empirische studies laten de effectiviteit van een bevolkingsonderzoek in een bepaalde epidemiologische situatie zien. Het is niet mogelijk om empirische studies voor alle alternatieve strategieën uit te voeren. Wiskundige modellen bieden hier uitkomst. Met behulp van modellen kunnen de resultaten van empirische studies geëxtrapoleerd worden naar andere strategieën en situaties.
De kosten van een screeningsprogramma zijn ook van belang. In kosten-effectiviteitsstudies worden de kosten en effecten van gezondheidsinterventies aan elkaar gerelateerd. De resulterende kosten-effectiviteitsratio kan worden gebruikt voor het vergelijken van verschillende screeningsstrategieën met elkaar, maar ook voor de vergelijking van screening met andere gezondheidsinterventies.
Er zijn diverse mogelijkheden voor het gebruik van wiskundige modellen bij de evaluatie van screening voor kanker. Wiskundige modellen kunnen worden onderverdeeld in modellen gebruikt voor 1) data analyse, 2) evaluatie en 3) optimalisatie. Modellen voor data analyse worden gebruikt om hypothesen te toetsen omtrent het natuurlijk beloop van de ziekte, karakteristieken van de screeningstest en de relatie tussen vroege opsporing van een kanker en de kans om aan die kanker te overlijden. Met behulp van evaluatiemodellen worden voorspellingen gedaan over de resultaten van een screeningsprogramma. Deze kunnen vervolgens vergeleken worden met de gerealiseerde uitkomsten van een screeningsprogramma. Als de resultaten afwijken van de verwachtingen, kunnen met behulp van het model mogelijke redenen hiervoor worden geëxploreerd.
Bij optimalisatie van screening kunnen modellen gebruikt worden om voor verschillende screeningsstrategieën de kosten en effecten te bepalen, en op basis daarvan een optimale screeningsstrategie voor een bepaalde situatie vast te stellen.
In dit proefschrift worden diverse wiskundige modellen gebruikt ten behoeve van de evaluatie van screening op borst- en baarmoederhalskanker. Hierbij is veelvuldig gebruik gemaakt van het

microsimulatieprogramma MISCAN. Dit programma is ontwikkeld door het instituut Maatschappelijke Gezondheidszorg, Erasmus MC, voor evaluatie van screening op kanker.

In hoofdstuk 2 zijn de eerste resultaten van screening op borstkanker in Navarra (Spanje) vergeleken met de verwachte resultaten die MISCAN genereert op basis van de demografische, epidemiologische en screeningskarakteristieken in Navarra. Het bleek echter dat het aantal gedetecteerde kankers in de eerste ronde van screening hoger was dan verwacht, terwijl het aantal gedetecteerde kankers in de tweede ronde van het screeningsprogramma lager was dan verwacht. Diverse alternatieve modelaannamen konden de geobserveerde gegevens niet verklaren. Ondanks dat, wordt een jaarlijkse reductie in de borstkankersterfte van 17 tot 23% voorspeld.
Na de invoering van screening op borstkanker, zal het enige tijd duren voordat er een reductie in de borstkankersterfte waar te nemen is (hoofdstuk 3). Met behulp van MISCAN werd voorspeld dat het na de invoering van het landelijke bevolkingsonderzoek naar borstkanker in 1989, tot 1997 zou duren voordat er een reductie in de borstkankersterfte waarneembaar was. In Engeland en Wales werd al binnen zeven jaar na de introductie van screening een sterfte reductie van 12% waargenomen. Modelberekeningen lieten echter zien dat 70% van dit percentage toe te schrijven is aan screening. De overige reductie wordt toegeschreven aan een verbetering van de behandeling voor borstkanker.

Screening op baarmoederhalskanker vindt plaats door middel van het maken van uitstrijkjes. Wanneer er afwijkingen in het uitstrijkje worden gevonden, wordt de vrouw geadviseerd nader diagnostisch onderzoek te ondergaan. Als er geen afwijkingen worden geconstateerd in het uitstrijkje, wordt de kans op het ontwikkelen van baarmoederhalskanker in de eerste jaren na het negatieve uitstrijkje klein geacht. Uit een vergelijking van het aantal gevallen van baarmoederhalskanker in de jaren na een negatief uitstrijkje met het aantal gevallen van baarmoederhalskanker dat ontstaat in een situatie zonder screening in Nederland, blijkt inderdaad dat het risico op baarmoederhalskanker sterk gereduceerd is gedurende een lange periode (meer dan 10 jaar) na een negatief uitstrijkje (hoofdstuk 4). Een nauwkeurige inschatting van deze reductie wordt echter bemoeilijkt door het ontbreken van gegevens over de hoogte van de incidentie van baarmoederhalskanker wanneer er geen screening zou plaatsvinden, de zogenaamde achtergrondincidentie. Deze incidentie zal niet gelijk zijn aan de incidentie in een situatie met screening, aangezien een effectief screeningsprogramma voor baarmoederhalskanker de incidentie zal verminderen. Wij hebben laten zien dat het nemen van de incidentie uit de periode voor screening, kan leiden tot een overschatting van de reductie in incidentie door baarmoederhalskankerscreening, aangezien er aanwijzingen zijn dat de incidentie sindsdien onafhankelijk van screening gedaald is. Omdat er doorgaans geen rekening is gehouden met deze trend, zal in verscheidene studies de achtergrondincidentie en daarmee het effect van screening zijn overschat.

In verschillende landen in Europa, Noord-Amerika en Australië zijn screeningsprogramma's naar baarmoederhalskanker geïntroduceerd. Deze laten echter een grote variëteit zien. Het aanbevolen aantal uitstrijkjes per vrouw varieert van zeven in Nederland en Finland tot meer dan 50 in Duitsland. Het geadviseerde interval tussen de uitstrijkjes is minimaal één jaar in Duitsland en maximaal vijf jaar in Nederland en Finland. De leeftijdsgroep waarbinnen de uitstrijkjes worden gemaakt varieert overeenkomstig: de beginleeftijd varieert tussen 20 en 30 jaar, en de leeftijd waarop de laatste screening plaatsvindt is minimaal 59 jaar, terwijl er in sommige landen geen maximumleeftijdsgrens is gesteld.
Met behulp van MISCAN zijn voor 500 screeningsstrategieën de kosten en effecten bepaald (hoofdstuk 5). Op basis van de verwachte kosten en effecten zijn de screeningsstrategieën waarvoor geen alternatieven bestaan met hogere opbrengsten voor dezelfde kosten, of lagere kosten bij gelijkblijvende effectiviteit vastgesteld. Er waren 15 zogenaamde efficiënte strategieën, variërend van twee tot 40 uitnodigingen. De incrementele kosten-effectiviteitsratio van deze strategieën nam toe van $ 6.700 per gewonnen levensjaar (twee uitnodigingen) tot $ 173.700 per gewonnen levensjaar (40 uitnodigingen).
Ondanks dat voor de meeste landen de gebruikte screeningsstrategie redelijk dicht bij een efficiënte strategie gesitueerd was, leidde het grote aantal aanbevolen uitstrijkjes per vrouw tot ongunstige kosten-effectiviteitsratio's.
In praktijk zal de kosten-effectiviteit van screeningsprogramma's nadelig beïnvloed worden door vrouwen die zich niet aan het screeningsprogramma houden. Een deel van de vrouwen neemt niet deel aan het screeningsprogramma. Dit is nadelig voor de effectiviteit van het screeningsprogramma, des te meer vanwege het feit dat vrouwen die niet deelnemen aan screening een hoger risico op baarmoederhalskanker blijken te hebben. Andere vrouwen laten meer uitstrijkjes maken dan aanbevolen. Deze uitstrijkjes leiden in de praktijk tot een nauwelijks hogere effectiviteit terwijl de kosten van screening wel proportioneel toenemen met het aantal uitstrijkjes (hoofdstuk 6).

Het humaan papillomavirus (HPV) is de belangrijkste risicofactor voor het krijgen van baarmoederhalskanker. De mogelijkheid om de HPV-test op te nemen in het bevolkingsonderzoek naar baarmoederhalskanker wordt momenteel onderzocht. In hoofdstuk 7 laten wij met behulp van MISCAN zien dat op basis van de beschikbare cross-sectionele gegevens er geen uitspraak is te doen over de toegevoegde waarde van de HPV-test. Het was mogelijk om twee verschillende modellen te construeren, één uitermate gunstig voor HPV-screening en één uitermate ongunstig, waarbij beide modellen in overeenstemming zijn met de waargenomen HPV-prevalentie in vrouwen met en zonder een voorstadium van baarmoederhalskanker. In het gunstige model wordt aangenomen dat de duur van HPV voorafgaand aan de voorstadia van baarmoederhalskanker 10 jaar bedraagt, terwijl dit in het ongunstige model één jaar is.
Om een uitspraak te doen over de (kosten-)effectiviteit van het gebruik van de HPV-test in screening op baarmoederhalskanker, zijn longitudinale gegevens nodig. Er zijn inmiddels

verschillende langdurige longitudinale studies gestart, maar de resultaten van deze studies zullen pas rond 2005 beschikbaar komen. In de tussentijd hebben we het natuurlijk beloop van de HPV-infectie in combinatie met (de voorstadia van) baarmoederhalskanker bestudeerd met behulp van de resultaten van een beperkte longitudinale studie. Op basis van deze studie (hoofdstuk 8), werd de duur van de HPV-infectie voorafgaand aan het ontstaan van de voorstadia van baarmoederhalskanker tussen twee tot vijf jaar geschat.

Uit het voorgaande blijkt dat modellen een waardevolle bijdrage kunnen leveren in de evaluatie van screeningsprogramma's voor borst- en baarmoederhalskanker. Toch zijn er enige onderwerpen op dit gebied die nog verdere aandacht behoeven. Dit betreft onzekerheidsanalyses, de validiteit van de modellen en de communicatie van complexe modellen (hoofdstuk 9).

Bij het gebruik van modellen is het van belang de gevolgen van de onzekerheid rondom de invoerparameters voor de modeluitkomsten te bepalen. Met behulp van onzekerheidsanalyses kan de onzekerheid rondom parameters die gebruikt worden als modelinvoer worden gespecificeerd, waarna de gevolgen van deze onzekerheid voor de modeluitkomsten kunnen worden bepaald. Onzekerheidsanalyses kunnen ook worden gebruikt om die invoerparameters te identificeren die het grootste deel van de onzekerheid rondom de modeluitkomsten bepalen. Voor deze modelparameters kan vervolgens worden geprobeerd om de onzekerheid in te perken. Mocht dit niet mogelijk zijn dan moet de onzekerheid geaccepteerd worden als zijnde inherent aan het beslissingsproces in die situatie.

Voordat modellen gebruikt worden voor aanbevelingen op het gebied van bevolkingsonderzoek, dienen deze adequaat gevalideerd te worden. Hiervoor dient de modelstructuur valide gevonden te worden door personen met kennis van zaken op dit gebied. Daarnaast moet het model de gegevens die gebruikt zijn voor de kwantificering van het model kunnen reproduceren (interne validiteit) en zou het model ook externe datasets, dat wil zeggen de data die niet gebruikt zijn voor de schatting van de parameters, moeten kunnen reproduceren (externe validiteit).

Echter ook voor gevalideerde modellen geldt dat na aanpassing voor nieuwe situaties, de modeluitkomsten eerst zorgvuldig dienen te worden vergeleken met geobserveerde gegevens, voordat het model gebruikt wordt voor voorspellingen.

In wetenschappelijk onderzoek dienen niet alleen de resultaten, maar ook de methode die gebruikt is om deze resultaten te genereren openbaar te zijn. Op deze manier kunnen geïnteresseerden een onafhankelijke controle van het model uitvoeren, en kan het begrip van het model en de resultaten vergroot worden, wat eventuele 'black box' gevoelens kan verminderen. In het verleden is het regelmatig moeilijk gebleken om gedetailleerde modelbeschrijvingen te publiceren in de wetenschappelijke tijdschriften. Tijdschriften accepteerden voornamelijk korte beschrijvingen.

Tegenwoordig zijn er echter nieuwe mogelijkheden om deze informatie te verschaffen, bijvoorbeeld door informatie op een website te plaatsen. Hierbij heeft de website van het tijdschrift de voorkeur.

Conclusies

- Gevalideerde (borstkanker)modellen kunnen niet zonder meer worden gebruikt in andere situaties.
- De kosten per gewonnen levensjaar van screeningsprogramma's op baarmoederhalskanker met een groot aantal uitstrijkjes zijn beduidend hoger dan bij een klein aantal uitstrijkjes.
- De evaluatie van de effectiviteit van het bevolkingsonderzoek naar baarmoederhalskanker wordt bemoeilijkt door de onzekerheid over de incidentie van baarmoederhalskanker die zou zijn opgetreden in een situatie zonder screening.
- Wiskundige modellen zijn een belangrijk hulpmiddel in de evaluatie van screening op (borst- en baarmoederhals-) kanker.

Aanbevelingen

- Zodra longitudinale gegevens van grote, langdurige gerandomiseerde HPV-onderzoeken beschikbaar zijn, moet een kosten-effectiviteitsanalyse worden uitgevoerd om de optimale inzet van de HPV test in het bevolkingsonderzoek naar baarmoederhalskanker te bepalen.
- Modellen moeten gevalideerd zijn voordat zij gebruikt worden voor evaluatie en predictie.
- Gedetailleerde informatie over modellen zou publiekelijk beschikbaar moeten zijn.

# PUBLICATIONS

## THIS THESIS

### This thesis is based on the following papers and manuscripts

M.E. van den Akker-van Marle, C.M.M. Reep –van den Bergh, R. Boer, A. Del Moral, N. Ascunce, H.J. de Koning. Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low at the second round. International Journal of Cancer (1997), 73, 464-469 *(Chapter 2)*.[1]

M.E. van den Akker-van Marle, H.J. de Koning, R. Boer, P.J. van der Maas. Reduction in breast cancer mortality due to the introduction of mass screening in the Netherlands; comparison with the UK. Journal of Medical Screening (1999), 6, 30-34 *(Chapter 3)*.

M.E. van den Akker-van Marle, M. van Ballegooijen, J.D.F. Habbema. Incidence of cervical cancer after negative screening in the Netherlands. British Journal of Cancer (2003), 88, 1054-1057 *(Chapter 4)*.[2]

M.E. van den Akker-van Marle, M. van Ballegooijen, G.J. van Oortmarssen, R.Boer, J.D.F. Habbema. Cost-effectiveness of cervical cancer screening: comparison of screening policies. Journal of the National Cancer Institute (2002), 94, 193-204 *(Chapter 5)*.[3]

M. van Ballegooijen, E. van den Akker-van Marle, J. Patnick, E. Lynge, M. Arbyn, A. Antilla, G. Ronco, J.D.F. Habbema. Overview of important cervical cancer screening process values in European Union countries, and tentative predictions of the corresponding effectiveness and cost-effectiveness. European Journal of Cancer (2000), 36, 2177-2188 *(Chapter 6)*.[4]

M. van Ballegooijen, M.E. van den Akker, P.G. Warmerdam, C.J.L.M. Meijer, J.M.M. Walboomers, J.D.F. Habbema. Present evidence on the value of HPV testing for cervical cancer screening: a model-based exploration of the (cost-)effectiveness. British Journal of Cancer (1997), 76, 651-657 *(Chapter 7)*.[2]

M.E. van den Akker-van Marle, M. van Ballegooijen, L. Rozendaal, C.J.L.M. Meijer, J.D.F. Habbema. Extended duration of the detectable stage by adding HPV test in cervical cancer screening. *Submitted (Chapter 8)*.

M.E. van den Akker-van Marle, R. Boer, G.J. van Oortmarssen. Comments on Duffy/Chen Markov chain models for progression of breast cancer. Letter to the Editor. Journal of Epidemiology and Biostatistics (1998), 3, 423-427 *(Appendix 1)*.

M.E. van den Akker-van Marle, M. van Ballegooijen, R. Boer, G.J. van Oortmarssen, J.D.F. Habbema. Withdrawing low risk women from cervical screening programmes. Conclusions cannot yet be drawn. British Medical Journal (1999), 319, 58 *(Appendix 2)*.[5]

# OTHER PUBLICATIONS

M. van Ballegooijen, A.B. Bos, M.E. van den Akker, G.J. van Oortmarssen, R. Boer, W.J. Meerding, J.D.F. Habbema. Een eerste beschrijving van de praktijk van het bevolkingsonderzoek naar baarmoederhalskanker in Nederland in 1994 op grond van gegevens uit het centrale PALGA. MGZ rapport 97.1. Rotterdam, 1997.

M.E. van den Akker-van Marle, C.M.M. Reep-van den Bergh, J.J. Polder, R. Boer, P.M.M. Beemsterboer, P.G. Warmerdam, H.J. de Koning. Cost-effectiveness analysis of breast cancer screening in European pilot projects. Evaluation of the early detection programme for breast cancer in Navarra, Spain. MGZ rapport 97.29. Rotterdam, 1997.

A.B. Bos, M. van Ballegooijen, G.J. van Oortmarssen, M.E. van Marle, J.D.F. Habbema, E. Lynge. Non-progression of cervical intraepithelial neoplasia estimated from population-screening data. British Journal of Cancer (1997), 75, 124-130.

J. Cuzick, P. Sasieni, P. Davies, J. Adams, C. Normand, A. Frater, M. van Ballegooijen, E. van den Akker. A systematic review of the role of human papillomavirus testing within a cervical screening programme. Health Technology Assessment (1999), 3, 1-196.

J. Cuzick, P. Sasieni, P. Davies, J. Adams, C. Normand, A. Frater, M. van Ballegooijen, E. van den Akker-van Marle. A systematic review of the role of human papilloma virus (HPV) testing within a cervical screening programme: summary and conclusions. British Journal of Cancer (2000), 83, 561-565.

M. van Ballegooijen, A.J. Bouwman-Notenboom, M.E. van den Akker-van Marle, A.B. Bos, G.J. van Oortmarssen, R. Boer, W.J. Meerding, W.A. Stolk, J.D.F. Habbema. De praktijk van het bevolkingsonderzoek naar baarmoederhalskanker in Nederland in 1996 en 1997: Landelijke cijfers en regionale vergelijkingen. Rotterdam, 2000.

A.B. Bos, M. van Ballegooijen, M.E. van den Akker-van Marle, A.G.J.M. Hanselaar, G.J. van Oortmarssen, J.D.F. Habbema. Endocervical status is not predictive of the incidence of cervical cancer in the years after negative smears. American Journal of Clinical Pathology (2001), 115, 851-855.

A.B. Bos, M. van Ballegooijen, M.E. van den Akker-van Marle, J.D.F. Habbema. Minder Pap-2 uitslagen ('lichte afwijkingen') in het bevolkingsonderzoek naar baarmoederhalskanker sinds de invoering van nieuwe richtlijnen in 1996. Nederlands Tijdschrift voor Geneeskunde (2002), 146, 1586-1590.

A.B. Bos, M. van Ballegooijen, M.E. van den Akker-van Marle, A.G.J.M. Hanselaar, G.J. van Oortmarssen, J.D.F. Habbema. Incidentie van baarmoederhalskanker even groot na een negatief uitstrijkje met of zonder endocervicale cellen. Nederlands Tijdschrift voor Geneeskunde (2002), 146, 1591-1594.

W.J. Meerding, M. van Ballegooijen, M.P.M. Burger, M.E. van den Akker-van Marle, W.G.V. Quint, J.D.F. Habbema. Human papillomavirus testing for triage of women referred because of abnormal smears: a decision analysis considering outcomes and costs. Journal of Clinical Epidemiology (2002), 55, 1025-1032

# DANKWOORD

Mijn eerste woord van dank gaat uit naar degenen die me ingewijd hebben in de wereld van de borst- en baarmoederhalskankerscreening: Petra Beemsterboer, Peter Warmerdam, Gerrit van Oortmarssen (zal het ons ooit lukken om samen een artikel af te ronden?), Rob Boer, Harry de Koning, Marjolein van Ballegooijen en Dik Habbema. De laatste twee personen hebben ook bij de afronding van dit proefschrift ondanks hun drukke werkzaamheden tijd gevonden om elke keer opnieuw mijn stukken van commentaar te voorzien.
Ook aan degenen met wie ik de afgelopen jaren heb samengewerkt, zowel nationaal als internationaal, ben ik dank verschuldigd. Het was goed om ook buiten de eigen (model)muren te kijken, ik heb hier veel van geleerd. Met name wil ik noemen de samenwerking met de afdeling Pathologie van het VU Medisch Centrum, professor Chris Meijer en zijn collega's. Onze discussies waren heftig, maar brachten ons ook verder.

Het instituut Maatschappelijke Gezondheidszorg was een ideale plek om al deze jaren te vertoeven, van de brede discussies op het inhoudelijke vlak, tot de ontspanning buiten het werk tijdens Beachvolleybaltoernooien en survivalweekenden in de Ardennen.
Een speciaal woord voor mijn kamergenoten, bedankt voor alle gezelligheid in de afgelopen jaren. Met mijn kamergenoten van het eerste uur, Anita Bos en Carin Reep-van den Bergh, heb ik een onvergetelijke tijd gehad, waarin veel lief en leed gedeeld werd. En ook de kamergenoten die mij de afgelopen periode hebben moeten verdragen, bedankt hiervoor. Merel van Dijk, jouw bananenbrood, worteltjestaart en bleke betten hebben zeker een bijdrage aan dit proefschrift geleverd, en Gerrit Draisma, bedankt voor je altijd relativerende opmerkingen.

Buiten het werk, bedank ik iedereen die regelmatig of incidenteel de zorg voor onze kinderen overneemt. Alleen als je met een gerust hart je kinderen achterlaat, is het mogelijk om werk te combineren met kinderen.
Familie en vrienden, bedankt dat jullie onvermoeibaar bleven vragen hoe het met het proefschrift ging, en daarnaast ook de broodnodige afleiding gaven. Dit geldt zeker voor mijn paranimfen Carina van Vliet en Caroline de Vocht-van Rossum. Carina, zo'n promotie kan toch niet moeilijker zijn dan de Afsluitdijk met tegenwind of kamperen op Schiermonnikoog zonder geld en geldautomaat? Bedankt dat je me ook hierbij terzijde wil staan. Caroline, alhoewel het bij volleybal zo was dat ik de ballen opzette en jij ze afmaakte, zal ik het nu beiden zelf moeten doen. Maar dat je ook hier naast me staat, zal me zeker helpen. Bedankt hiervoor.

Lieve pappa en mamma, alweer een proefschrift af! Zou het laatste nu ook nog komen? Voor mij was dit alles niet mogelijk geweest zonder jullie onophoudelijke steun, variërend van de uitgebreide telefoongesprekken tot het uitleggen van de eerste universitaire calculussommen.
Ruben en Meike, voor jullie breken de gouden tijden nu aan, eindelijk zal de computer wat vaker beschikbaar zijn om Robbie Konijn te spelen.
Lieve, lieve Jaap, bedankt voor alles.

# Mathematical Modelling in Evaluation of Screening for Breast and Cervical Cancer

# CURRICULUM VITAE

Elske van den Akker-van Marle werd geboren op 28 mei 1971 te Alphen aan den Rijn. In 1989 haalde zijn haar Atheneum β diploma aan het Christelijk Lyceum te Alphen aan den Rijn. Tijdens haar middelbare schooltijd won zij met twee medescholieren de door staatssecretaris Ginjaar-Maas ingestelde prijs voor meisjesteams bij de Wiskunde Olympiade.

In 1989 startte zij met de studie Econometrie aan de Vrije Universiteit te Amsterdam. Vanaf 1990 studeerde zij daarnaast Theologie. In 1994 rondde zij haar studie Econometrie af met als specialisatie 'Operations Research', na een stage in het Waterland Ziekenhuis te Purmerend. De afstudeerstage betrof onderzoek naar mogelijkheden voor capaciteitsplanning in het Waterland Ziekenhuis.

Van oktober 1994 tot en met april 2003 was zij in dienst bij het Instituut Maatschappelijke Gezondheidszorg aan de Erasmus Universiteit Rotterdam. Hier hield zij zich bezig met de evaluatie van bevolkingsonderzoek naar borst- en baarmoederhalskanker met behulp van wiskundige modellen. In 1997 bracht zij in het kader van het Young Scientist Summer Programme 3 maanden door aan het International Institute for Applied System Analysis (IIASA) in Laxenburg (Oostenrijk).

In november 2002 is zij in dienst getreden bij TNO Preventie en Gezondheid waar zij zich bezighoudt met kosten-effectiviteitsstudies op het gebied van de Jeugdgezondheidszorg.