Evaluation of screening programmes

Studies on breast cancer and prostate cancer

Publication of this thesis was financially supported by the Department of Public Health, Erasmus University, Rotterdam, and the 'Stichting Urologisch Wetenschappelijk Onderzoek'.

The research on breast cancer was financed by: Deutsche Mammographie Studium, Catalan Agency of Technology Assessment, and the Health Insurance Executive Board of the Netherlands.
The research on prostate cancer was carried out in collaboration with the Department of Urology, Erasmus University and Academic Hospital, Rotterdam and financed by the Dutch Cancer Society and the Dutch Prevention Fund.

Cover illustration: 'Fakkeloptocht', painting by Carolien Beemsterboer
Cover design: Carolien Beemsterboer
Layout: Bon Mot, Rotterdam
Printed by: Print Partners Ipskamp, Enschede

ISBN 90-9012612-0
NUGI 757

Evaluation of screening programmes

Studies on breast cancer and prostate cancer

Evaluatie van screening programma's

Onderzoek naar borstkanker en prostaatkanker

Proefschrift

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus
Prof.dr P.W.C. Akkermans, M.A.
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 19 mei 1999 om 13:45 uur

door

Petronella Margretha Maria Beemsterboer

geboren te Heemskerk

Promotiecommissie

Promotor	:	Prof.dr P.J. van der Maas
Overige leden	:	Prof.dr F.H. Schröder
		Prof.dr A.L.M. Verbeek
		Prof.dr F.F.H. Rutten
		Dr H.J. de Koning (tevens co-promotor)

• Contents •

1

INTRODUCTION

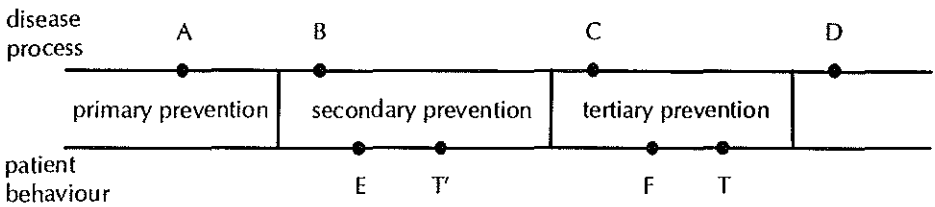
INTRODUCTION

In the past century treatment and primary prevention of disease has focussed on decreasing mortality rates (Wolleswinkel-van den Bosch, 1998). The current challenge is directed towards secondary prevention. Screening for disease is becoming increasingly part of medical practice in the Western world. Screening for cervical cancer with PAP smears and for lung cancer with chest X-rays were the first examples of cancer screening that were expected to reduce mortality (Boucot, 1948; Papanicoulou and Traut, 1941; Victor, 1955). Although many in the medical profession had great expectations there were also opponents, at least for lung cancer screening (Boucot and Sokoloff, 1955). The natural course and the screening tools (chest X-rays) available for the detection of lung cancer seemed not to be able to detect the cancer in a phase where treatment resulted in substantially improved prognosis. The results of clinical trials, published approximately 30 years later, showed no overall decrease in lung cancer mortality and screening for lung cancer never became established as a public health service (Early Lung Cancer Cooperative Study, 1984). For cervical cancer however, the characteristics of screen test in combination with the natural history of the disease (i.e. the long duration of the screen detectable period), is estimated to reduce mortality by 75% (100% attendance of screening) (van Ballegooijen, 1998). Evidence about the effectiveness of cervical cancer screening is provided by historical studies, case-control studies, analysis of data from large screening programmes and analysis of the natural history of cervical cancer using mathematical models.

Mammography as a screening tool with acceptable test characteristics was developed in the sixties (Gershon-Cohen *et al.*, 1961). At that time, testing new techniques in Randomised Clinical Trials had become common practice. This explains why the estimates of the effectiveness of breast cancer screening were much more evidence based than those of cervical and lung cancer screening.

In figure 1.1 the conceptual framework of screening is depicted. For lung cancer, because the pre-clinical detectable stage (B-C) is so short and lacks a suitable test to detect the disease early, no improvement in prognosis can be achieved. For breast and cervical cancer, screening by mammography and cervical smears can detect and treat the disease, while it is still in the pre-clinical phase. This results, on average, in improved prognosis if treated and thus prevents deaths from the disease.

For breast cancer, national screening programmes have been introduced in the Netherlands and in many European countries (Fracheboud *et al.*, 1998; Shapiro *et al.*, 1998).



A-B:	period of increased risk
B:	first detectable patho-fysiologic change
B-C:	pre-clinical detectable stage
C:	first symptoms and/or signs
D:	stable outcome (partial or complete recovery or death)
E:	moment of early detection
F:	moment when patients will seek medical advice because of signs or symptoms
T:	time of onset of therapy
T':	onset of therapy after early detection
E-F:	lead time
C-F:	"patient delay"
E-T' and F-T:	period of diagnostic procedures and doctors' delay

Figure 1.1
Conceptual frame work of screening, the disease process and patient behaviour

Table 1.1
Favourable and unfavourable effects of introducing cancer screening

EVALUATION OF SCREENING	
Favourable effects	Unfavourable effects
• Cancer mortality reduction	• Lead time (years)
• Prevention of advanced disease	• Attending screening (pain, anxiety, false positive results)
• Less intensive treatment[1]	• Overdiagnosis
• Cost savings	• Costs
	• Secondary effects of screening (e.g. radiation risk[1], opportunistic screening)

[1] in the case of breast cancer

Swedish randomised trials have shown that a breast cancer mortality reduction of approximately 30% can be achieved by mammography for women aged 50-69, of whom 70% attend the screening (Nystrom *et al.*, 1993). The favourable outcome of a trial with regard to mortality is however not a guarantee of a successful nation-wide screening programme. Before the introduction of a screening programme in the total population, all positive and negative effects of screening for population health status should be quantified. In table 1.1, these favourable and unfavourable effects are summarised. The current state of health care relating to the disease, i.e. no available screening programme is the baseline for a comparison with a situation where screening is available.

The advantage of evaluating screening programmes and other health care facilities using a uniform outcome measurement (cost per Quality adjusted life year saved) ensures that the outcomes of health services can be compared and provides relevant information for assisting health policy priorities. The consequences for population health status can therefore be better assessed. This is currently, especially relevant as technical developments are resulting in new diagnostic tools and thus new possibilities for early detection. Because of a limited health care budget a uniform outcome is expected to facilitate considerations on the introduction of both screening programmes and new curative treatments.

New challenges for cancer screening are in the area of colorectal cancer, melanoma and prostate cancer. Screening is, in regard to these diseases still however in a process of development. There are many questions that have to be resolved, before policy makers can decide about the introduction of these programmes. In this thesis the differences in the phase of evaluation between two cancer-screening programmes (breast cancer and prostate cancer) are used to illustrate aspects of the evaluation of cancer screening programmes.

EVALUATION OF BREAST CANCER SCREENING

In the Netherlands a thorough evaluation of all effects and costs in the total female population was carried out before the decision to introduce screening as a public health service was made, (de Koning, 1993). This cost-effectiveness analysis was also used to determine the most appropriate age group and screening interval (de Koning *et al.*, 1991). The MISCAN (MIcrosimulation SCreening ANalysis) programme takes into account all favourable and unfavourable effects of the introduction of screening using the situation where no screening programme exists, as a reference (Habbema *et al.*, 1985; van Oortmarssen *et al.*, 1990).

Using the computer simulation package MISCAN, individual life histories are generated, representing the demography, mortality of all causes and incidence and mortality from breast cancer. In the disease part of the programme the relevant stages of breast cancer are discerned and the natural history is simulated as a progression through these stages. Key parameters in the model of the performance of screening are mean duration of screen-detectable preclinical disease, sensitivity and improvement of prognosis for screen detected cancers. Changing one of the input parameters, e.g. sensitivity of the screening, age group or interval enables the evaluation of the impact of different scenarios.

Some examples of evaluation issues that were (in part) studied in this thesis, in the phase after the introduction of screening programme as a national health care service in the Netherlands, are described below. An important question in regard to screening evaluation is whether the results of cost-effectiveness analyses in one country can be extrapolated to other countries with a different health care setting, epidemiology of the disease and perhaps other test characteristics of mammography. In a study by van Ineveld et al. it was shown that the cost-effectiveness of breast cancer screening might differ by a factor of 3-5 between different countries of the European Union (van Ineveld *et al.*, 1993). In this thesis factors are explored that determine whether and how results of cost-effectiveness analyses and thus evaluation of screening programmes can be extrapolated. This is illustrated using a cost-effectiveness analysis from Germany, in a decentralised screening setting and in Spain where there is a centralised screening system, but with lower levels of incidence and mortality of breast cancer.

After screening was introduced in the Netherlands for all women aged 50-69 with an interval of 2 years, the National Evaluation Team on Breast cancer screening (NETB) analysed the early outcomes of the screening (de Koning *et al.*, 1995; Fracheboud *et al.*, 1998). At the end of 1997 all 750.000 women in the target age group had been invited at least once to participate in the national screening programme. During the first few years the evaluation of the national screening programme focused on the results of the first screening examinations. These results were very encouraging. The subsequent screens are however also of the most importance for the effectiveness of the screening programme (reduction of breast cancer mortality). After implementation of the programme, approximately 90% of all screening examinations consisted of subsequent screens. The cancers should be detected so early in their development that the expected breast cancer mortality reduction should result. This depends however on the sensitivity of the screening test and the biological growth of the tumours. The results of subsequent screens so far have not been as encouraging as those of the first screens (Fracheboud *et al.*, 1998).

After the introduction of the Dutch screening programme, the results of a combined analysis of the Swedish screening trials were published (Nystrom *et al.*, 1993). These new results were implemented in the MISCAN model in order to make new estimates of the expected mortality reduction resulting from the Dutch national screening programme (de Koning *et al.*, 1995). Other issues that require continuous attention are new screening techniques such as computed assisted diagnosis (Karssemeijer and Hendriks, 1997) and the discussion about screening women in their forties and the expansion of screening to higher ages (Anonymous, 1997).

Before the introduction of breast cancer screening in the Netherlands, all effects mentioned in table 1 were quantified and used in the cost effectiveness analysis (de Koning, 1993). The quantification of secondary effects had however been based on assumptions that had not yet been studied or were judged to be of minor importance at the time. Two of these topics, namely opportunistic screening (screening outside a programme) and radiation risk of mammography were further studied in this thesis.

The effect of introducing breast cancer screening in certain age groups (50-69) might change the use of mammography in other age groups and in the target population. In the cost-effectiveness analysis a reduction of mammography requests was assumed in the target population. In this thesis the use of mammography as requested by general practitioners was analysed before and after the introduction of the national screening programme.

The radiation dose used in modern mammography was assumed to be negligible at the time the decision was made to introduce breast cancer screening (Health Council of the Netherlands, 1987). As a result of new techniques (the use of grids to prevent scatter radiation) and continuous improvement of image quality, the radiation dose had probably increased from 0.5 mGy to about 2 mGy per examination. These developments had also accelerated, as a result of the introduction of screening programmes. In this thesis, the radiation risk of mammography and the implications for screening programmes with different age groups and intervals is further studied.

EVALUATION OF PROSTATE CANCER SCREENING

Screening for prostate cancer is much less developed than for breast cancer. Randomised Controlled Trials are now being conducted (Auvinen *et al.*, 1996; Gohagan *et al.*, 1994) but to date no country has implemented a national screening programme. In Europe, many urologists and policy makers seem to share the opinion that screening should only be introduced after it has been proven to save lives. Evidence about the

effectiveness of screening in reducing prostate cancer mortality will only be available approximately 10 years after the start of the randomised screening trials, i.e. from about 2005. But for prostate cancer screening, saving lives is not the only issue in a policy decision to introduce screening. All favourable and unfavourable effects of the screening (see table 1) should be identified and quantified.

For the evaluation of prostate cancer screening a MISCAN (MIcro Simulation SCreening ANalysis) disease module for prostate cancer was developed. In this module the relevant stages of the disease are integrated and validated with information about incidence, mortality, stage distribution and survival of prostate cancer in the situation where no screening programme exists. Information about detection rates and stage distribution of the trial is also a continuous source of clinical input for the model.

In addition a quality of life study is also conducted alongside the trial. Quality of life before and after participating in the screening (Essink-Bot *et al.*, 1998) and before and after primary treatment has been studied (Madalinska *et al.*). Furthermore, quality of life of metastasised prostate cancer will also be studied.

Another issue concerns the quantification of changes in assessment and therapy as a result of screening. The health care costs of diagnosis and treatment of prostate cancer have also to be estimated. All the information will be combined resulting in cost-effectiveness ratios for prostate cancer screening. It is expected that, as for breast cancer, this analysis will also help to decide about a lower and upper age-limit of the target population and the screening interval. Using this information will support a decision about the introduction of a nation-wide screening programme for prostate cancer.

This thesis contains studies on some of the elements in table 1.1. Examples are the analysis of the most optimal combination of screening tests and the observation of PSA (Prostate Specific Antigen) among men before and after participating in the screening trial and in the total population as a baseline value for the situation without screening. Determination of PSA in serum is a very simple screening test compared to mammography and a cervical smear. A widespread use of the screen test in the control arm of the trial (opportunistic screening) might affect the mortality reduction achieved in the trial. If large numbers of opportunistic screening occur this might be taken into account separately in the cost-effectiveness analysis. With the introduction of a screening programme this 'opportunistic' screening is (in part) replaced by an organised screening. Thus part of costs of screening is compensated for by abolishing opportunistic screening (which is generally less efficient).

From the evaluation of breast cancer screening it was obvious that many of the extra costs (approximately 40%) induced by screening could be compensated for by prevention of advanced disease and its costs (de Koning *et al.*, 1992). Similarly for pros-

tate cancer, quantification of advanced disease and its costs is also one of the aspects for consideration in the screening evaluation. Therefore in this thesis the course, care and costs of advanced prostate cancer were studied.

RESEARCH PURPOSES

The goal of this thesis is to elucidate some aspects of evaluation of screening programmes. The evaluations of breast cancer and of prostate cancer screening are in very different phases. The differences are used to illustrate the challenges for the evaluation of both screening programmes. Especially the experience of the evaluation of breast cancer screening can be applied to prostate cancer.

The following research purposes will be addressed in this thesis:

- To identify and quantify factors that influence the cost-effectiveness of (breast) cancer screening programmes in different countries and thus health care settings (Part I)
- To quantify secondary effects (radiation risk of mammography and opportunistic screening) of the introduction of a national breast cancer screening programme (Part II)
- To quantify factors influencing the performance of a prostate cancer screening trial or a future programme (Part III)

STRUCTURE OF THIS THESIS

Chapter 2 presents the costs and effects of breast-cancer screening in Germany where health care is decentralised. The effect of sensitivity, specificity and attendance rate on the effects and costs is quantified. In chapter 3, breast cancer screening programmes are evaluated to compare extension of screening of the age group 50-64 to higher or lower ages in a country with a lower level of incidence and mortality (Spain; Catalonia).

In chapter 4, the possible effects of radiation exposure in a breast cancer-screening programme are explored for different scenarios. Radiation risk is age-dependent and the effectiveness of screening is also age-dependent. Furthermore, radiation risk is proportional to the total dose and thus dependent on the age-borders and interval. The effects of both factors combined, on the balance between breast cancer deaths induced and prevented is investigated. In chapter 5 we evaluate breast cancer screening outside an

organised screening programme by the general practitioner in the target age group of the national screening programme and in adjacent age groups.

In part III the screening tests as used in the Rotterdam part of the European Randomised Study on Screening for Prostate Cancer (ERSPC) are evaluated (chapter 6). We look at different scenarios by weighing screening procedures, biopsies and cancers detected. In chapter 7 the use of 'screening tools', especially PSA (Prostate Specific Antigen) of the trial population before randomisation and during the study is quantified. Additionally the base line (situation without screening) with regard to the assessment of prostate disease is established. In chapter 8 the course of advanced prostate cancer is described, including the care given and the costs involved.

In chapter 9 the most important aspects of both screening programmes are discussed to establish their relevance for the evaluation. It is obvious that the evaluation of the screening programmes is in totally different phases. For prostate cancer the evaluation will be used to decide about whether a prostate cancer-screening programme should be introduced and if so, how. For breast cancer the issue is how to monitor and optimise an already established national screening programme.

References

Anonymous, National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40-49, January 21-23, 1997. National Institutes of Health Consensus Development Panel. *J Natl Cancer Inst*, 89, 1015-26 (1997).

Auvinen, A., Rietbergen, J.B., Denis, L.J., Schroder, F.H. and Prorok, P.C., Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. *J Med Screen*, 3, 97-104 (1996).

Boucot, K.R., Mass surveys in case-finding techniques for pulmonary neoplasms. *J Am Med Wom Assoc*, 3, 485-492 (1948).

Boucot, K.R. and Sokoloff, M.J., Is survey cancer of the lung curable? *Dis Chest*, 27, 369-388 (1955).

de Koning, H.J., The effects and costs of breast cancer screening, *Dept. of Public Health*, Erasmus University Rotterdam, Rotterdam (1993).

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M.M. and van der Maas, P.J., Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst*, 87, 1217-23 (1995).

de Koning, H.J., Fracheboud, J., Boer, R., Verbeek, A.L., Collette, H.J., Hendriks, J.H., van Ineveld, B.M., de Bruyn, A.E. and van der Maas, P.J., 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*, 60, 777-80 (1995).

de Koning, H.J., van Ineveld, B.M., de Haes, J.C., van Oortmarssen, G.J., Klijn, J.G. and van der Maas, P.J., Advanced breast cancer and its prevention by screening. *Br J Cancer*, 65, 950-5 (1992).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

Early Lung Cancer Cooperative Study, Early lung cancer detection: summary and conclusions. *Am Rev Respir Dis*, 130, 565-570 (1984).

Essink-Bot, M.L., de Koning, H.J., Nijs, H.G., Kirkels, W.J., van der Maas, P.J. and Schroder, F.H., Short-term effects of population-based screening for prostate cancer on health-related quality of life. *J Natl Cancer Inst*, 90, 925-31 (1998).

Fracheboud, J., de Koning, H.J., Beemsterboer, P.M.M., Boer, R., Hendriks, J.H.C.L., Verbeek, A.L.M., de Bruyn, A.E. and P.J., v.d.M., Nation-wide breast cancer screening in the Netherlands: results of initial and subsequent screening 1990-1995. *Int J Cancer*, 75, 694-698 (1998).

Gershon-Cohen, J., Hermel, H.B. and Berger, S.M., Detection of breast cancer by periodic X-ray examination: A five-year survey. *JAMA*, 176, 1114-1116 (1961).

Gohagan, J.K., Prorok, P.C., Kramer, B.S. and Cornett, J.E., Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute [see comments]. *J Urol*, 152, 1905-9 (1994).

Habbema, J.D., van Oortmarssen, G.J., Lubbe, J.T. and van der Maas, P.J., The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed*, 20, 79-93 (1985).

Health Council of the Netherlands, The early detection of breast cancer (in Dutch), Health Council of the Netherlands, The Hague (1987).

Karssemeijer, N. and Hendriks, J.H., Computer-assisted reading of mammograms. *Eur Radiol*, 7, 743-8 (1997).

Madalinska, J.B., Essink-Bot, M.L., de Koning, H.J., Schroder, F.H. and van der Maas, P.J., Health related quality of life in patients with either screen-detected or clinically diagnosed prostate cancer before undergoing primary treatment. *(in preparation)*.

Nystrom, L., Rutqvist, L.E., Wall, S., Lindgren, A., Lindqvist, M., Ryden, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J. and et, a., Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. *Lancet*, 341, 973-8 (1993).

Papanicoulou, G.N. and Traut, H.F., The diagnostic value of vaginal smears in carcinoma of the uterus. *Am J Obstet Gynecol*, 42, 193 (1941).

Shapiro, S., Coleman, E.A., Broeders, M., Codd, M., de Koning, H.J., Fracheboud, J., Moss, S., Paci, E., Stachenko, S. and Ballard-Barbash, R., Breast cancer screening programs in 22 countries: current policies, administration, and guidelines. *Int J Epidemiology*, in press (1998).

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

van Ineveld, B.M., van Oortmarssen, G.J., de Koning, H.J., Boer, R. and van der Maas, P.J., How cost-effective is breast cancer screening in different EC countries? *Eur J Cancer*, 12, 1663-8 (1993).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

Victor, A.B., The early diagnosis of primary lung cancer. *Dis Chest*, 27, 389-409 (1955).

Wolleswinkel-van den Bosch, J.H., The epidemiological transition in The Netherlands [Thesis], *Department of Public Health*, Erasmus University Rotterdam, Rotterdam (1998).

• Part I •

Identification and quantifying factors that influence the cost effectiveness of (breast) cancer-screening programmes in different health care settings

2

PREDICTION OF THE EFFECTS AND COSTS OF BREAST CANCER SCREENING IN GERMANY

Summary

Although breast cancer screening programmes are now being introduced it is still debated whether this is an appropriate policy for all European countries. Taking into account empirical data from 2 regional pilot screening projects, this study has evaluated the effects and costs of a nation-wide breast cancer screening programme in Germany. Special attention was paid to the decentralised German health-care system and to the influence of attendance, interval and age group. The recent results of the analysis of the Swedish randomised screening trials were used to estimate the improvement in prognosis after early detection of breast cancer.

Our analysis shows that a programme providing for the screening of women aged 50-69 at 2-year intervals might be expected to result in a decrease in mortality from breast cancer estimated at 11% for the total German population, representing 2,100 deaths from breast cancer prevented each year. The cost per life-year gained was assessed at between DM 18,800 and DM 25,300 for this scenario; 2 to 3 times less favourable than in the UK and The Netherlands. The sensitivity of mammography was estimated to be 12% lower than in The Netherlands and the attendance rate was calculated at 47% on average. A greater effort to ensure the quality of the screening programme and to improve the invitation system might finally lead to much better results. The mortality reduction might be as much as 18% if the attendance and the sensitivity of the screening could be improved to the Dutch level.

Acknowledgements

We thank Mr. B.M. van Ineveld for his valuable suggestions on cost aspects and Mrs. J.C.J.M. de Haes for her research on the quality of life.

INTRODUCTION

Mammographic screening has been shown to be effective in reducing breast-cancer mortality for women aged 50 to 69 in several West European countries (Nystrom *et al.*, 1993; UK Trial of Early Detection of Breast Cancer Group, 1993). Still, it is not evident that a decision for screening would be appropriate for all European countries since they might show large differences regarding epidemiology, screening and health care in general (van Ineveld *et al.*, 1993). In The Netherlands a detailed cost-effectiveness analysis was carried out to predict the favourable and unfavourable effects of breast-cancer screening (de Koning, 1993; de Koning *et al.*, 1991). These predictions led to the decision to introduce a national screening programme, the results of which appeared to prove that the predictions had been accurate (NETB (National Evaluation Team for Breast cancer screening), 1993).

We used the same approach to make predictions about the main consequences of breast cancer screening in one of the largest West European countries, Germany. Since 1971 the opportunity has been afforded for regular cancer-related health check-ups each year, with mammography for high-risk groups only and if symptoms or complaints were present (Robra, 1993). In recent years experimental mammography screening has been carried out in 2 regions (DMS (Deutsche Mammographie-Studie), 1991, 1992, 1993), which served as the basis for this analysis.

The specific influence of a more decentralised screening organisation, including private practices, was taken into account. Another important issue was the influence of attendance rates, which are likely to be lower in screening programmes without strict invitation systems. Results are presented on the main favourable and unfavourable effects of screening women aged 50 to 69 at 2-year intervals and on quality of life and cost-effectiveness. Because it is still debated whether younger women should also be screened, we considered alternative scenarios including women under 50 and/or different screening intervals. We also examined the country-specific characteristics of screening for breast cancer in Germany and the requirements for obtaining the largest possible benefit from breast-cancer screening.

MATERIAL AND METHODS

Cost effectiveness analysis and baseline assumptions

The design of an analysis of the effects and costs of breast cancer screening has been described by de Koning et al. (de Koning *et al.*, 1991). In short, the effects and costs of different policies are compared with a situation in which mass screening is not applied. The description of the latter situation is based on national data on assessment and treatment. Cost and effect estimates for screening are derived form the results of screening trials. The estimate of improvement in prognosis after early detection is based on results from the recent analysis of the Swedish randomised trials (Nystrom *et al.*, 1993).

For the baseline assumptions on screening and the natural history of breast cancer we used the cost effectiveness approach and MISCAN model, which were also used to evaluate screening in The Netherlands (de Koning, 1993; van Oortmarssen *et al.*, 1990). All factors were adjusted for the German situation where available data allowed. This method has been shown to be useful for predicting the effects and cost of screening in Australia, done with the same base-line assumptions adjusted as required (Carter *et al.*, 1993). The disease model is based on a 3-stage division of the development of invasive breast cancer (stages reflect the size of the tumour). A proportion of the invasive breast cancers was assumed to be preceded by a screen-detectable ductal carcinoma in situ (dCIS). No changes were made regarding the impact on the quality of life in the different phases (de Haes *et al.*, 1991). Only the extent to which these states would be prevented or induced by screening in the German setting was taken into account.

For the analysis it was assumed that screening would start in 1994, with a build-up period of 5 years, based on the experience in The Netherlands and the UK. After this build-up period it was assumed that the programme would be carried out in the whole of Germany for a total period of 27 years.

Demography and epidemiology of breast cancer in Germany

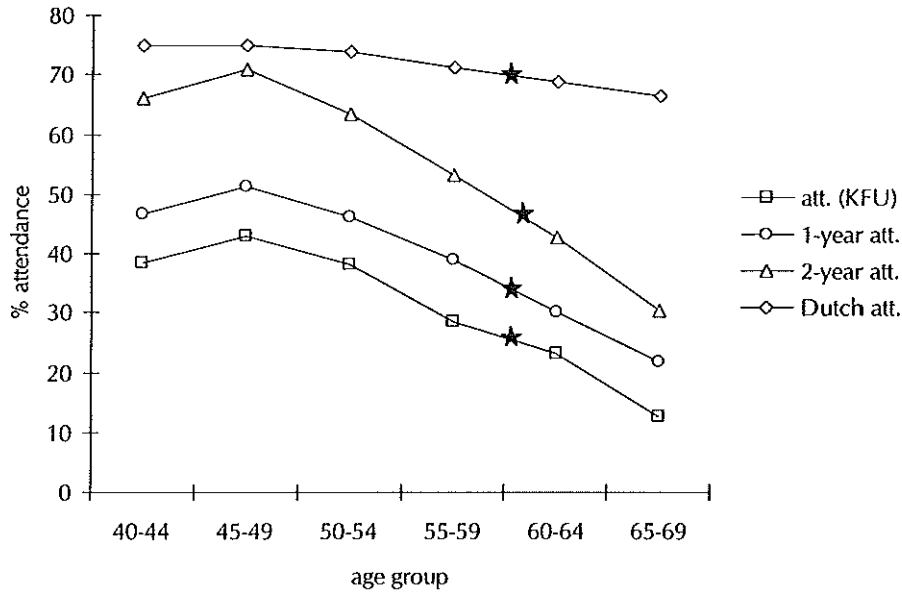
The 1989 (5-year) age-specific distribution of the total German female population was used and the death rates from other causes than breast cancer were based on the 1990 (5-year) age-specific rates from the former Federal Republic of Germany. Detailed data were lacking the former German Democratic Republic. Breast-cancer mortality rates for the total German population were calculated on the basis of 1990 data from the former FRG and 25% lower rates were used for the former GDR (Statistisches Bundesamt,

1992; Statistisches Bundesamt, 1991). The "onset" of disease (or preclinical incidence) could be assessed by using the clinical age-specific incidence figures from the Saarland cancer registry and the expected duration of preclinical breast cancer as estimated in detail in earlier analysis with empirical data on screening from the Dutch trials. The mean duration of preclinical disease was estimated in detail in earlier analyses with empirical data on screening from the Dutch trials. The mean duration of preclinical disease was estimated to range from 2.7 (age 50) to 6.2 (age 70) years (Statistisches Amt des Saarlandes, 1987, 1988, 1989, 1990; van Oortmarssen *et al.*, 1990). Stage distribution of breast cancer was assumed to be the same as in The Netherlands, although it was difficult to make an accurate comparison on account of the diversity of sources (Netherlands Cancer registry, 1992; Leonhardt, 1988; Paterok et al., 1992).

Since data on the incidence of breast cancer in the whole of Germany were lacking, the Saarland registry was used to determine breast-cancer incidence at the national level. Regional differences in incidence were assumed to be negligible on the basis of the regional mortality data for breast cancer (Smans *et al.*, 1992). First, the Saarland incidence data were used to predict the mortality from breast cancer in Saarland (Statistisches Amt des Saarlandes, 1987, 1988, 1989, 1990). Since the predicted mortality was too high, the relative survival rates of breast cancer patients had to be adjusted slightly downwards for all ages and specifically for the younger age group to make the simulated breast-cancer mortality fit the data observed in Saarland. With the mortality incidence distribution from Saarland, it was possible to estimate national clinical incidence form the national mortality data. The mortality rates predicted by MISCAN and those of the German national data were comparable.

Screening in Germany

In 2 pilot regions, Aurich and Braunschweigh, the Deutsche Mammographie-Studie (DMS) has been carried out since 1990 (DMS (Deutsche Mammographie-Studie), 1991, 1992, 1993). The DMS study is part of the Krebs-Früherkennungsuntersuchungen (KFU) programme, which affords women the opportunity to see a doctor for regular yearly examinations. The expected age-specific attendance rates were based on attendance rates observed in the KFU programme in 1985-1986 and on rates obtained from a recent telephone survey (Berghof and Robra, 1988; Robra, personal communication). Because it was likely that these 2 sources of data would respectively underrate and overrate attendance, we decided to use the average. Moreover, as attendance rates obtained form the telephone survey only concerned women aged 55 to 74, those for women age 40 to 54 were extrapolated from the data of the KFU. On the basis of the data thus obtained, the attendance rates for women attending breast cancer screening



att. (KFU) = attendance of KFU-programme, 1985-1986 (1-year interval)
1-year att. = 1-year average attendance based on KFU-programme and telephone survey
2-year att. = 2-year average attendance based on KFU-programme and telephone survey
Dutch att. = attendance of Dutch screening trials (2-year interval)

Figure 2.1
Different attendance patterns for screening programmes with average values for women aged 50–69 (depicted by ★)

programmes at least every other year (2-year attendance) or every year (1-year attendance) were calculated. The 1-year attendance rate of the KFU, considered at least a reasonable one, was used as a lower boundary. Attendance showed a steep decrease with increasing age, so relatively more younger women attend the programme (figure 2.1).

In the DMS study women were asked by their general practitioner or specialist to participate and undergo a 2-view mammography. Women with suspicious lesions were either directly for biopsy or received additional examinations first. The sensitivity and the positive predictive values (PPV) of screening were derived from the experience obtained in the first years of the pilot project, including 31,000 first screening examinations and 16,000 subsequent ones until the first quarter of 1993 (DMS (Deutsche Mammographie-Studie), 1991, 1992, 1993; Swart, 1992; Swart, personal communication). On the basis of these data the sensitivity for all stages was lowered by 12% compared to the Dutch sensitivity. The sensitivity of the screening test was 0.73 on average

depending on the stage in which the breast cancer was detected, ranging from 0.35 for dCIS to 0.62 for tumours < 1 cm and 0.84 for tumour ≥ 1 cm. In the MISCAN model the sensitivity of a screening examination is defined as the probability of detecting a preclinical breast cancer. No data were available on the assessment or biopsy of breast cancer in normal medical practice in Germany, so data applicable to The Netherlands were used. As a consequence the PPV of a biopsy in a situation without screening was higher for tumours < 1 cm than in the screening programme and almost the same for tumours larger than 1 cm.

Cost of screening, assessment and treatment

In Germany screening will be attached to a decentralised health-care system. While in The Netherlands special screening units have been built, screening in Germany is carried out in the private practices of gynaecologists and radiologists. The existing mammographic centres often have too high a capacity for an efficient programme to be carried out, and this is the most important reason for the higher cost of screening in such a system. The cost estimates of screening were different according to 2 calculations. In the first case, the amount of DM 121 per screening was reached by taking the tariff of DM 110 quoted in the DMS study and raising it by DM 11 for the cost of external quality assurance, double reading and training (DMS (Deutsche Mammographie-Studie), 1991, 1992, 1993). As a variant the Dutch cost structure of screening was used. It was not changed, but applied to a decentralised system by a correction of 1.6 based on data from a French decentralised screening system (Lancry and Fagnani, 1989; van Ineveld *et al.*, 1993).

The cost and use of all procedures concerning the assessment and treatment of breast cancer have been studied extensively for The Netherlands (de Koning, 1993). The evaluation of the use of different therapeutic procedures in Germany was based on published data, since national data and registries were lacking. The available information was comparable with that from Dutch medical practice (Granetzny *et al.*, 1991; Leonhardt, 1988; Paterok *et al.*, 1992). The tariffs of different additional examinations and treatments covered by sickness funds in The Netherlands were related to the tariffs in Germany and minor adjustments were made (Mundenbruch, 1990). Because the structure of care was based on the Dutch situation, all costs were corrected for the purchasing power parities (PPP) of the cost of health care (OECD (Organization for Economic Co-operation and Development), 1991). This PPP was 1.19, which means that health care in Germany is 19% more expensive. [Costs are presented in "Deutsch Mark"(DM) in this study.]

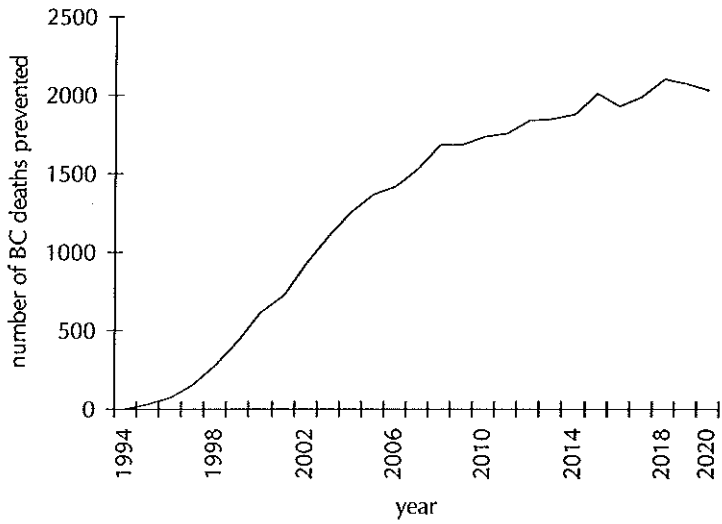


Figure 2.2
Yearly number of breast cancer deaths prevented by screening women aged 50–69 with a 2-year interval, 1994–2020

Table 2.1
Stage distribution in a situation with and without screening and of the screen detected cancers in Germany in the year 2000

	without screening		with screening (incl. screen-detected)		screen-detected	
	N	(%)	N	(%)	N	(%)
dCIS	1,840	(4.6)	2,630	(6.2)	930	(14.0)
Invasive < 1 cm	3,430	(8.5)	4,690	(11.1)	1,390	(21.0)
Invasive 1-2 cm	9,970	(24.7)	11,680	(27.7)	2,760	(41.7)
Invasive ≥ 2 cm	25,070	(72.2)	23,100	(54.9)	1,540	(23.3)
Total	40,310		42,100		6,620	

RESULTS

Effects on incidence and mortality

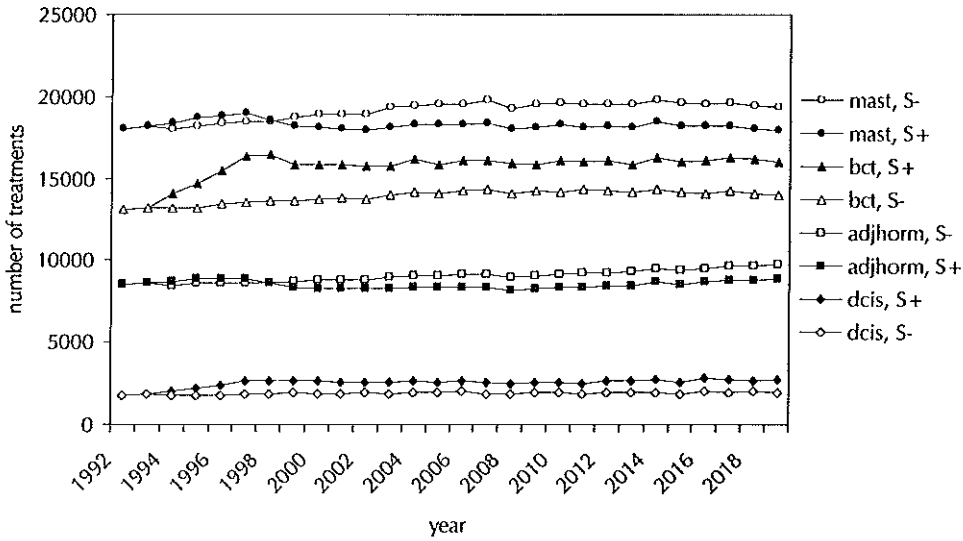
In the year 2000 breast cancer will be diagnosed in about 40,310 women if no screening is available. When a screening programme is implemented in Germany the number of newly diagnosed breast-cancer cases will eventually increase by about 2-3%. In a situation where women aged 50-69 are screened every 2 years, 42,100 cases will be diagnosed each year, 6,620 of them being detected by screening, that is 16% of all breast cancers diagnosed each year. To assist with these diagnoses, about 5.2 million women have received an invitation and about 2.4 million women have had a mammography (0.6 million first screens and 1.8 million subsequent screens). Ultimately a total of 2,100 deaths from breast cancer will thus be prevented each year, and the reduction of mortality from breast cancer is expected to be 11% in the total German population by the year 2020 (figure 2.2).

A comparison between the stage distribution for 1 year in a situation with and without screening and of the screen-detected cases (50-69) is summarised in table 2.I. The stage distribution is much more favourable for screen-detected cancers. Since only 16% of all breast cancers are detected by screening, the effect on the overall stage distribution is relatively small.

Effects on health care

With the introduction of a screening programme the use of health-care services for diagnosis and treatment will increase. In a programme form women aged 50-69 at 2-year intervals, 2.6% of all women screened will have additional examinations. After the first screen this proportion will be 4.5%, because of a prevalence screen. A biopsy will be carried out in 0.6% of all women screened. The number of diagnostic procedures outside screening will decrease, because some women will be diagnosed with breast cancer in the screening programme. The proportion of diagnostic procedures for non-palpable breast cancers (< 1 cm) will increase, however, because smaller cancers will be detected by screening.

Although the absolute number of treatments for breast cancer will rise since more breast cancers will be detected, the proportions between the various therapies used will change (figure 2.3), because the implementation of a screening programme will result in a more favourable stage distribution. In comparison with a situation without screening, the use of breast conserving therapy will increase by about 15% after the build-up period while 8% fewer mastectomies will be carried out. The treatment of



S- = in a situation without screening
S+ = in a situation with screening
mast = mastectomy
BCT = Breast Conserving Therapy
adjhorm = adjuvant hormonal therapy
dCIS = treatment for dCIS

Figure 2.3
Yearly number of different treatments with and without screening of women aged 50–69 with a 2-year interval, 1994–2020

dCIS will increase by 36% and the treatment of advanced disease will decrease by 11%.

Cost effectiveness and quality of life

In table 2.2 a situation with screening and one without are compared for all costs over a period of 27 years. The largest extra cost is caused by the screening procedure itself (DM 4,010 million). The costs of assessment or biopsy, primary therapy and follow-up are all higher in a situation with screening. The increase in costs of assessment and biopsy is almost as great as the increase in costs due to treatment (DM 290 million and DM 320 million respectively). On account of the relatively large number of false positives, the additional cost of assessment and biopsy is high in a situation with screening. Moreover, a greater number of procedures for non-palpable breast cancers are carried

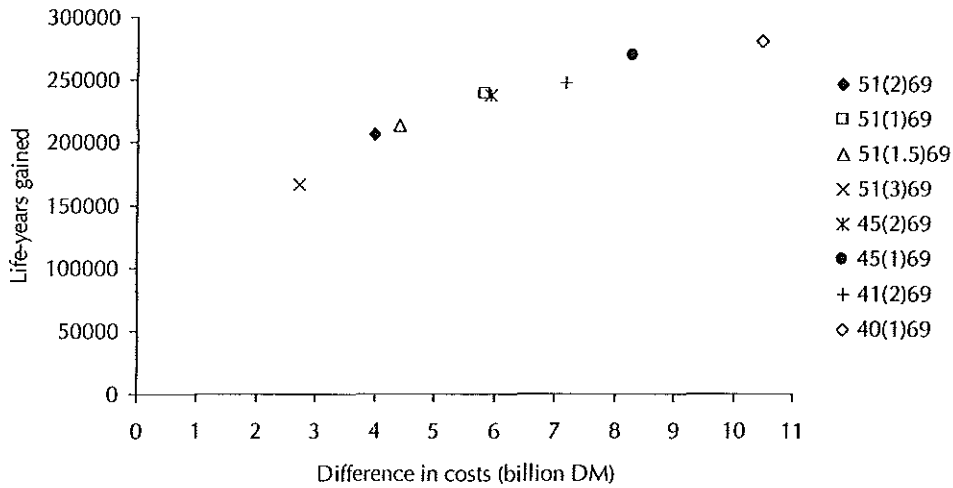
out and this also involves a higher cost. In a situation with screening the overall costs for breast-cancer diagnoses and treatment is 11% higher than in a situation without screening. The cost of advanced disease is DM 890 million lower in a situation with screening. This reduction is proportional to the decline in mortality since all women prevented from dying of breast cancer would have had palliative treatment. In total, the savings are so small that, when a screening programme is implemented in Germany, the overall extra costs will be only 2.5% lower than the costs incurred by the programme itself (DM 3,890 million and DM 4,010 million respectively). The mean number of life-years gained by the earlier detection of breast cancer was estimated at 15.8 years per breast-cancer death prevented (no discounting). The costs per life-year gained for screening women aged 50-69 at a 2-year interval were DM 18,800 (5% discounted).

When the costs of screening were assumed to be higher than in the basic scenario, as in all likelihood they will be when a national decentralised screening programme is implemented, the balance between costs and effects was even worse (DM 25,300 per life-year gained).

Table 2.2
Effects on mortality, costs and cost-effectiveness of breast cancer screening of women 50-69 with a 2-year interval in Germany during 27 years. Cost amounts are expected differences between a situation with and without screening (in Million DEM), 5% discounting, cost per screen DEM 121.

	screening versus no screening	no screening
Breast cancer deaths prevented[1]	54321	
Life-years gained[1]	860273	
Cost of screening	+4,010	0
Cost of assessment/biopsy	+290	7,350
Cost of primary therapy	+320	9,770
Cost of follow-up	+170	1,850 [2]
Cost of advanced disease	−890	15,860
Difference in cost	+3,890	34,830 (total cost)
Breast cancer deaths prevented	19,600	
Life-years gained	206,500	
Cost per life-year gained (CE-ratio, DEM)[3]	18,800	
Quality adjusted life-years gained	197,000	
Cost per quality adjusted life-year gained (DEM)	19,800	

[1]no discounting
[2]only cost of follow-up of women after 1992 are included
[3]DEM 25,300 if cost of screening based on Dutch cost-structure and decentralized health care system



51(3)69 = a scenario with screening for women aged 51–69 with a 3-year interval

Figure 2.4
Difference in extra costs and life-years gained for different interval and age group scenarios

To assess the quality of life a comparison was made between a situation with screening and one without. The largest gain in quality-adjusted life-years is caused by the less frequent use of palliative therapy in a situation with screening. On the other hand, many quality-adjusted life-years are lost, since a screening situation results in a greater number of life-years after disease. Overall, 500 quality-adjusted life-years are lost with screening, so a small correction of 4.8% was carried out on the total number of life-years gained (table 2.2). The costs per quality-adjusted life-year gained amount to DM 19,800 for the basic scenario and DM 26,500 for the variant with the higher costs of screening (in both cases 5% discounted).

Influence of screening interval

If the screening interval is shortened and screening is carried out more frequently, a greater number of life-years are gained. The curved line in figure 2.4 shows that additional costs will be incurred while proportionally less life-years will be gained. This can be seen in the scenario involving age group 50-69, with intervals of 3, 2, 1.5 and 1 years. If the screening interval is reduced from 3 to 2 years, the extra costs per extra life-year are DM 29,400. If the interval is reduced from 2 to 1.5 years the extra costs are even higher (DM 69,900). The figure also shows that the costs of implementing screen-

ing policies for women aged 45+ or 40+ are relatively high compared to the benefits in terms of mortality reduction. For the variants with a lower starting age the extra costs are very high, with a maximum of DM 96,200 per extra life-year gained when the interval is shortened from 2 years to 1 in the age group 40-69. Is has to be taken into account that according to the average values of the KFU programme and the telephone survey (figure 2.1), the attendance rates vary proportionally to the changing interval.

Influence of attendance rates on breast-cancer-induced mortality

Attendance rates in the Dutch screening trials were 70% on average, while in the German basic scenario they were 47% and in the KFU programme approximately 26% (figure 2.1). On the basic of these rates, the mortality reduction is 17%, 11% and 7% respectively and the number of breast-cancer deaths prevented about 30,200, 19,600 and 11,600 respectively for the basic scenario. In Germany a steep decrease in attendance is observed starting from age 55, while in The Netherlands proportionately more older women attend the screening. There are however only slight differences in the CE (cost-effectiveness) ratios.

DISCUSSION

The predicted balance between the effects and costs of breast-cancer screening in Germany for women aged 50-69 at 2-year intervals is less favourable than in other European countries. In Germany the CE ratio is between DM 18,800 and DM 25,300 per life-year gained for this basic scenario if the costs of a decentralised health-care system are taken into account. This figure is DFL 7,650 (Dutch florins) for the same programme in The Netherlands (exchange rate 1994, 1 DFL = 0.89 DM) (de Koning *et al.*, 1991). The predicted reduction in mortality from breast cancer was 11% in the total female population for this programme and less than the 16% predicted for The Netherlands.

The effects and cost effectiveness of screening for breast cancer are determined by different factor. Firstly, the CE ratio depends on the epidemiology of the disease, incidence and mortality. If the incidence is high, more cases can be detected and the effects of screening for disease might be greater, thus preventing more women from dying of breast cancer. In Germany, since national data on the incidence of breast cancer were lacking we predicted the national incidence using the survival rates for Saarland and the national mortality rates. We may have underestimated or overestimated the effects if the survival rates for Saarland are not representative of all of Germany. However, because of the absence of data, this problem cannot be addressed. We nonethe-

less anticipate minor differences in survival deriving also from minor regional differences in breast-cancer mortality (Smans *et al.*, 1992). The extrapolation is not likely to have influenced the results of the CE analysis very much.

Another factor that influences the effects of screening is the quality of the screening test, its sensitivity and specificity. A third one is the behaviour of the target population with respect to attendance. All these aspects are negatively related to the CE ratio. In Germany the incidence of breast cancer and the mortality incurred by this disease are about 20 tot 25% lower than in The Netherlands, so the detection rates will be lower and relatively fewer cancers will be found at the same cost. Sensitivity and attendance rates also have lower values than in The Netherlands, with the difference that these aspects can be influenced.

On the basis of the German pilot regions, the sensitivity of screening for breast cancer is estimated to be 12% lower than in The Netherlands. Improving sensitivity would have a positive effect by increasing the number of early-detected breast cancers and thus reducing mortality. In Germany only 16% of all breast cancers are screen-detected while in The Netherlands, where sensitivity is higher, the proportion is 24%. Two factors to be taken into account are the higher proportion of older women living in Germany and their relatively low attendance. If more older women attended the programme the effect of an improvement of sensitivity on reducing mortality would be more than proportional.

Specificity affects the induced costs of the screening programme. In Germany a considerable number of additional examinations (2.6% of all screens) is carried out, although the percentage of biopsies (0.6% of all screens) is at the same level as in The Netherlands. Incidence is about 20 to 25% lower, so more women have to undergo a biopsy without having breast cancer. In The Netherlands, a very efficient referral system after screening (0.8% of all screened women are referred) reduces the costs of diagnosing breast cancer in a situation with screening. In Germany the costs of diagnosing breast cancer are much higher when screening is carried out than when it is not.

The attendance rate has a large impact on reducing mortality from breast cancer. The attendance assumed for this analysis in Germany was low compared to attendance in other European countries. The largest difference regarding the invitation system of the KFU programme is that women are not actively invited to participate. If a screening programme for breast cancer is implemented in Germany a personal invitation is expected to lead to much higher attendance rates. However, it is impossible to predict attendance in this situation since the health-care system is different. One way to increase the overall attendance might be to use population registers or registers of the sickness funds (Krankenkassen) or of family practices and to promote health education. In the national programmes of both The Netherlands and the UK, this approach results

in an attendance of 70% or more (Chamberlain *et al.*, 1993; de Koning, 1993). Our results show that with an average attendance rate of 70% the mortality reduction would increase to 17%. Finally, the cost of the screening procedure is high. If it were possible to save money on the screening examination the cost-effectiveness of this programme could be influenced.

Our results also show that in Germany it is advisable to have an extensive system to ensure the quality of screening, as in other European countries. The predictive values of different procedures must be improved, thereby reducing the absolute number of additional examinations through increased sensitivity. This can be achieved by the extensive training of all people involved in screening and by using equipment of high quality. The decentralised health-care system in Germany presents a difficulty in this respect. One solution might be to introduce referral centres to which all women with an abnormality on mammography would be sent. Specialised doctors would have to be involved to advise for additional examinations and biopsy.

The KFU programme, in which a general check-up is offered to all inhabitants to screen for different kinds of cancer, has been carried out in Germany for about 20 years. However, no favourable effect on breast cancer mortality has yet been demonstrated (Robra, 1993). A better option might be to introduce a specially organised national screening programme for breast cancer, which would replace the breast-cancer "screening" in the KFU programme. It would be strongly advisable to carry out this screening programme in Germany only when the quality of screening is improved and attendance is heightened, especially among older women. Performing a cost-effectiveness analysis would provide much information relevant to policy decisions before a national screening programme is implemented. This study has demonstrated that the results of a cost-effectiveness analysis are dependent on the country and the health service system for which the analysis is carried out. Without the use of a model it is extremely difficult to predict the overall effects of breast-cancer screening.

If sensitivity, specificity and attendance have the same value as in The Netherlands the mortality reduction in Germany will be even higher than in The Netherlands (18%) on account of the age structure of the German population. The CE ratio will improve as well (DM 15,700 per life gained). If all aspects of the Dutch screening programme except the demography and epidemiology of breast cancer were applied to Germany (costs included), the costs per life-year gained would be DFL 8,900. This finding shows that because of differences in demography and epidemiology the CE ratio would probably not be as favourable as in The Netherlands.

References

Berghof, B., Robra, B-P., Krankheitsfrüherkennung Krebs Frauen und Männer. Aufbereitung und interpretation der Untersuchungsergebnisse aus den gesetzlichen Früherkennungsmassnahmen 1985-1986, Deutscher Ärzte-Verlag, Cologne 42-49 (1988).

Carter, R., Glasziou, P., van Oortmarssen, G., de Koning, H., Stevenson, C., Salkeld, G. and Boer, R., Cost-effectiveness of mammographic screening in Australia. *Aust J Public Health*, 17, 42-50 (1993).

Chamberlain, J., Moss, S.M., Kirkpatrick, A.E., Michell, M. and Johns, L., National Health Service breast screening programme results for 1991-2 [published erratum appears in BMJ 1993 Aug 28;307(6903):543] [see comments]. *Bmj*, 307, 353-6 (1993).

de Haes, J.C., de Koning, H.J., van Oortmarssen, G.J., van Agt, H.M., de Bruyn, A.E. and van Der Maas, P.J., The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer*, 49, 538-44 (1991).

de Koning, H.J., de, The effects and costs of breast cancer screening, *Dept. of Public Health*, Erasmus University Rotterdam, Rotterdam (1993).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

DMS (Deutsche Mammographie-Studie), Erprobung von Qualitätssicherungsmassnahmen zur Einführung der Mammographie in das Krebsfrüherkennungsprogramm. Sachbericht, Hanover (1991, 1992, 1993).

Granetzny, A., Peek, U. and Komin, R., [Surgical and adjuvant therapy of breast cancer] Chirurgische und adjuvante Therapie des Mammakarzinoms. *Zentralbl Chir*, 116, 1111-23 (1991).

Lancry, J.P. and Fagnani, F., Evaluation économique du dépistage systématique des cancers du sein et du col de l'utérus, COMAC-HSR, Paris (1989).

Leonhardt, A., [Clinical detection of 2,511 breast cancers 1977-1986. Efficiency of legal early cancer detection studies] Die klinische Erkennung von 2511 Mammakarzinomen der Jahre 1977 bis 1986. Die Effizienz der gesetzlichen Krebs-Fruherkennungsuntersuchungen. *Geburtshilfe Frauenheilkd*, 48, 92-8 (1988).

Mundenbruch, R., Gegenüberstellung mit Abrechnungshinweisen, BMÄ E-GO EBM, Zauner, Dachau (1990).

NETB (National Evaluation Team for Breast cancer screening), National evaluation of screening for breast cancer in the Netherlands, Dept. of Public Health, Erasmus University, Rotterdam (1993).

Nystrom, L., Rutqvist, L.E., Wall, S., Lindgren, A., Lindqvist, M., Ryden, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J. and et, a., Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. *Lancet*, 341, 973-8 (1993).

OECD (Organization for Economic Co-operation and Development), Health data file. CREDES, OECD, Paris (1991).

Paterok, E.M., Rosenthal, H., Richter, S. and Sabel, M., [Breast cancer: trends from 1964 to 1990. Results of a long-term study]

Mammakarzinom: Trends von 1964 bis 1990. Ergebnisse einer Langzeitstudie. *Rontgenpraxis*, 45, 325-9 (1992).

Robra, B.-P., *Evaluation des Deutschen Krebsfrüherkennungsprogramms; Struktur, Prozess, Resultate und Entwicklungsperspektiven*, Wirtschaftverlag NW, Verlag für neue Wissenschaft, Bremerhaven (1993).

Smans, M., Muir, C.S. and Boyle, P., *Atlas of cancer mortality in the European Economic Community*, Oxford University Press, Oxford (1992).

Statistisches Amt des Saarlandes, Morbidität und Mortalität an Bösartigen Neubildungen im Saarland 1986, 1987, 1988, 1989. Jahresbericht des Saarlandischen Krebsregisters. Sonderhefte, Saarbrücken (1987, 1988, 1989, 1990).

Statistisches Bundesamt, *Statistisches Jahrbuch 1991 für das vereinte Deutschland*, Metzler Poeschel, Stuttgart (1992).

Statistisches Bundesamt, *Todesursachen 1990*, Metzler Poeschel, Stuttgart (1991).

Swart, E., Ansätze zur Qualitätssicherung ärtzlichen Handelns in der ambulanten Versorgung am Beispiel der Früherkennungsmammographie - Methodische Vorraussetzungen und praktische Erfahrungen, Medizinische Hochschule Hanover, Hanover (1992).

UK Trial of Early Detection of Breast Cancer Group, Breast-cancer mortality after ten years in the UK Trial of the early detection of breast cancer. *Breast*, 2, 13-20 (1993).

van Ineveld, B.M., van Oortmarssen, G.J., de Koning, H.J., Boer, R. and van der Maas, P.J., How cost-effective is breast cancer screening in different EC countries? *Eur J Cancer*, 12, 1663-8 (1993).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

3

SCREENING FOR BREAST CANCER IN CATALONIA: WHICH POLICY IS TO BE PREFERRED?

Summary

Background: the effects and costs of different policies for breast cancer screening in Catalonia (Spain) were analysed, to give a basis for setting priorities and deciding on the introduction of a screening programme.
Methods: the MISCAN (MIcrosimulation SCreening ANalysis) model of the natural history of breast cancer was used. The epidemiology of breast cancer in Catalonia and the demography of the Catalan population was taken into account as well as the results on mortality reduction from a Swedish overview of breast cancer screening trials.
Results: the reduction in breast cancer mortality in the total female population due to a screening programme for the age group 50-64 years would be 16, 12 and 9%, with screening intervals of 1, 2 and 3 years respectively. The cost-effectiveness ratios (CE ratios) for these scenarios were 924,000, 730,000 and 719,000 pesetas (Pts) per life-year gained respectively (5% discounting). The most cost-effective screening scenario is the one in which women aged 50-69 years are screened with an interval of 3 years with a mortality reduction of approximately 12% in the total female population (CE ratio = 694,000 Pts). Screening until the age of 69 years (2-year interval) was almost as cost-effective as screening the age group 50-64 years with a 2-year interval, with a reduction in breast cancer mortality of 15%. Extension to under the age of 50 years resulted in diverging results depending on the assumptions for improvement in prognosis for younger women (40-49 years).
Conclusion: if the extension of a 2 yearly screening programme for women aged 50-64 years is considered (mortality reduction of 12%), extension to older women would be more advisable, based on proven benefits and costs, than extension to younger age groups.

Acknowledgements

We would like to thank J. Galceran, A. Izquierdo, M. Marzo, G. Perez and J. Espinàs for providing data on cancer incidence, mortality and breast cancer screening pilot projects. We also thank J. J. Polder for working on the costs. This project was partially financed by the Catalan Agency for Technology Assessment.

INTRODUCTION

The cost effectiveness of breast cancer screening can vary substantially between countries, depending on, for example health care system, costs of health care, the screening programme and epidemiology of the disease. In Europe and the USA different screening policies exist and in the UK and The Netherlands these policies were in part based on cost-effectiveness analyses (de Koning *et al.*, 1991; Forrest, 1987).

In this study a detailed cost-effectiveness analysis was carried out for one region of Spain, Catalonia, where all relevant aspects were taken into account in as much detail as possible. In Spain, breast cancer screening activities are organised according to the regional health care organisation. Pilot projects on breast cancer screening have been started in different parts of Spain (Ascunce *et al.*, 1994; Servei de programmes especials, 1993). In Catalonia, a pilot project on breast cancer screening was started for women aged 50-64 years in Molins de Rei in the metropolitan area of Barcelona in 1992 (Marzo, 1995). The upper age limit in all pilot projects in Spain is 64 years, while the starting age is 45 or 50 years. The screening interval in these pilot projects is 2 years for the age group 50-64 years and 1 or 2 years for the age group 45-49 years.

In this study we show the reduction in breast cancer mortality and the cost-effectiveness ratio (CE ratio) of different screening intervals and age groups to be screened in Spain, taking into account the age-specific reductions in breast cancer mortality from a Swedish overview of screening trials (Nystrom *et al.*, 1993). Additional scenarios are examined in which screening women under the age of 50 years is as effective as for women aged 50-69 years.

METHODS

The MISCAN model approach to assessing the benefits of screening

To predict the number of breast cancer deaths prevented for all separate screening policies, the computer simulation MISCAN (MIcrosimulation SCreening ANalysis) model of the natural history of breast cancer was used. A detailed description of the earlier MISCAN model is given by van Oortmarssen et al (van Oortmarssen *et al.*, 1990). In the present model, breast cancer has four invasive, screen-detectable, pre-clinical states ($\leq$0.5 cm, 0.5-1 cm, 1-2 cm and $>$2 cm) and one non-invasive state, Ductal Carcinoma

In Situ (DCIS). The MISCAN model has been validated with data from the Dutch pilot screening projects from Utrecht en Nijmegen, which started in 1974 and 1975 (Collette *et al.*, 1984; Verbeek *et al.*, 1984). These data allowed an estimation of the sensitivity of mammography in the different states and the mean duration of the screen-detectable, pre-clinical period by age (van Oortmarssen *et al.*, 1990).

The level of mortality, underlying incidence and survival of breast cancer are important for the proportion of favourable and unfavourable effects of breast cancer screening. With the MISCAN model it is possible to take into account the demography and epidemiology of breast cancer. By generating individual life histories a dynamic population is simulated. The characteristics of the screening programme (attendance, interval and age groups) and the screening test (sensitivity) and assessment procedures (e.g. biopsy) are considered (see table 3.1). The value of the improvement in prognosis by stage after detection of breast cancer by screening was determined in a separate analysis (de Koning *et al.*, 1995). Based on the breast cancer mortality reduction reported in a Swedish overview of randomised breast cancer screening trials, the improvement

Table 3.1
Demography, incidence, mortality, survival and clinical stage distribution of breast cancer in Catalonia as used in the MISCAN model

Population 1991 (female)	
Total	3,096,552
Age 40-44 years	205,716
Age 45-49 years	184,291
Age 50-64 years	522,638
Age 65-69 years	160,973
Breast cancer incidence (per 100,000 person-years)	
Crude rate	75.4
Age standardized (European)	66.4
Mortality from breast cancer (per 100,000 person-years)	
Crude rate	32.8
Age standardized (European)	27.3
Five year relative survival rate	
Age <65 years	0.721
Age ≥ 65 years	0.693
Clinical stage distribution (%)	
DCIS	3.7
≤ 0.5 cm (T1a)	1.4
0.5-1 cm (T1b)	6.2
1-2 cm (T1c)	32.4
>2 cm (T2+)	56.3

Table 3.2
Characteristics of the screening for first and subsequent screenings or different age groups

	First screening (%)	Subsequent screening (%)
Attendance rate (average)		
Age 40-49 years	75	75
Age 50-64 years	70	70
Age 50-69 years	69	69
Referral rate		
Age 50-64 years (2 year interval)	6.2	3.6
Age 50-69 years (2 year interval)	6.3	3.6
Positive predictive value (PPV) of advice for biopsy		
Age 40-49 years	21	33
Age 50-64 years	35	55
Age 50-69 years	37	58
Detection rate (per 1000 examinations at 2 year interval)[4]		
Age 40-44 years	0.7	0.7
Age 45-49 years	1.7	1.4
Age 50-54 years	2.4	1.6
Age 55-59 years	2.9	1.8
Age 60-64 years	4.5	2.7
Age 65-69 years	6.5	3.6
Age 50-64 years[5]	3.3	2.2
Age 50-69 years[5]	4.1	2.6
	Age 45-49 years[2]	Age 50-69 years
Sensitivity of screen test[1]		
DCIS	0.32	0.40
≤ 0.5 cm (T1a)	0.52	0.65
0.5-1 cm (T1b)	0.64	0.80
1-2 cm (T1c)	0.72	0.90
> 2 cm (T2+)	0.76	0.95
Improvement in prognosis due to screen detection[3]		
DCIS	1.000	1.000
≤ 0.5 cm (T1a)	0.310	0.892
0.5-1 cm (T1b)	0.230	0.814
1-2 cm (T1c)	0.070	0.567
> 2 cm (T2+)	0.050	0.395

[1] Sensitivity of screen test is the probability of a positive screen result when screening a woman with screen-detectable pre-clinical cancer.
[2] For the age group 40-44 years the values for sensitivity were 60% of those for the age group 50-69 years.
[3] Improvement in prognosis is the reduction in risk of dying from breast cancer compared to a situation without screening, detected in that stage.
[4] Note: the possibility of an increase in underlying incidence over time, as seen in some parts of Spain, is not taken into account, due to lack of data.
[5] During the first 5 years of the programme.

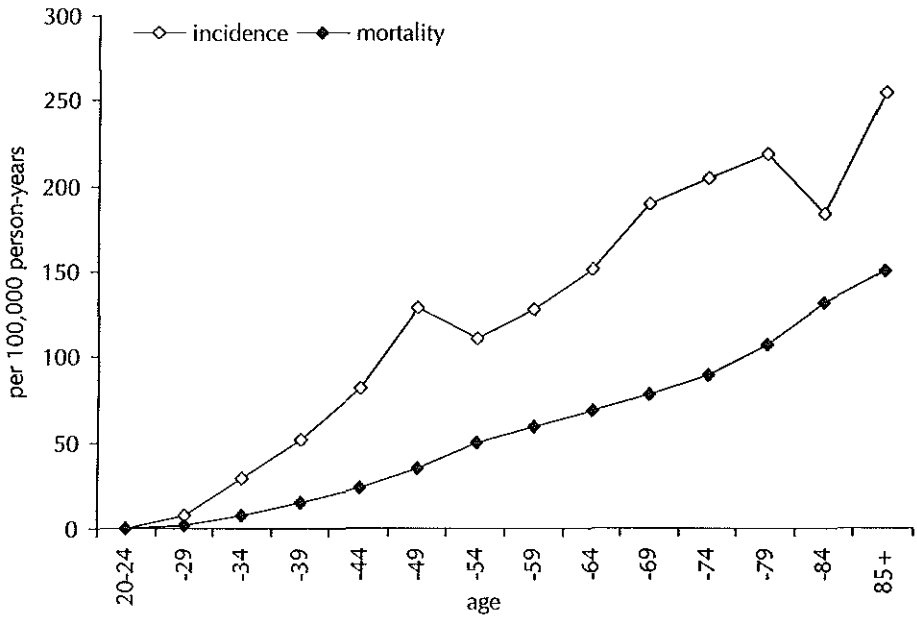


Figure 3.1
Breast cancer incidence and mortality in Catalonia (the MISCAN model)

in prognosis was estimated much higher in age groups over 50 years than for women aged 40-49 years (see table 3.2) (de Koning *et al.*, 1995; Nystrom *et al.*, 1993).

With the use of this model we were able to predict many changes resulting from the introduction of a screening programme in a population in detail. The estimates we present for the screening programmes with different age groups and intervals are for a programme starting in 1995 and carried out over a period of 27 years. This long period was chosen because of the gradual introduction of screening and the time lag between the introduction of screening and its effect on breast cancer mortality. The effects resulting from the screening programme that extend beyond the 27 years screening programme were also accounted for.

Demography and epidemiology of breast cancer in Catalonia

We used the life table and female population structure to simulate the demography of the female population in Catalonia (Servei d'Informacio Sanitaria, 1993). The incidence of breast cancer was used from the cancer registries in Girona and Tarragona for the years 1985-1989 (figure 3.1) (Borras and Galceran, 1993; Viladiu *et al.*, 1994). Because of small differences between rural and urban areas, the overall incidence was used. The

mortality from breast cancer was based on the vital health statistics of Catalonia (figure 3.1) (Servei d'Informacio Sanitaria, 1993).

The clinical stage distribution was used to estimate the mean duration of the screen-detectable, pre-clinical phase of the different states in the model. Data on the tumour size, lymph nodes and distant metastasis (TNM stage) were available for all women hospitalised for breast cancer in Girona. These data were compared to less detailed data (local, regional and distant) of all incident cases in the Cancer Registry of Girona (Viladiu *et al.*, 1994). We used the TNM stage information of all patients that were hospitalised during 1989, because the differences were small. The age- and stage-specific survival from breast cancer as used in the model was in accordance with the 5 year relative survival rate from Girona for the years 1985-1989 (Viladiu *et al.*, 1994). With the incidence and survival of breast cancer, the mortality of breast cancer, as found in the vital statistics, was predicted by the MISCAN model. The parameters used in the MISCAN model are shown in table 3.1.

The screening process

With regard to the screening process, we used the data of the small pilot project in Molins de Rei and the published results of a breast cancer screening programme in Navarra (Ascunce *et al.*, 1994). We assumed two-view mammography for the first screening and one-view screening for subsequent examinations. The number of women referred to the hospital for additional examinations after the first screen was 6% in Navarra, which was similar to the percentage of referrals found in Molins de Rei after a two-view examination (Marzo, 1995). We used a referral rate of 6.2% for the first screening and 3.6% for subsequent screening for women aged 50-64 years. The values for referral after subsequent screens and for the other age groups (50-69 and 40/45-64 years) were extrapolated using proportions of the Dutch situation, because no data for Catalonia and Navarra were available. All referred women received a clinical mammography and 12% of them also needed a biopsy. The attendance rate in Molins de Rei was 68% (Marzo, 1995) and in Navarra 86% (Ascunce *et al.*, 1994). In the scenario of screening women aged 50-69 years with a 2-year interval we used an attendance rate of 70% on average. The attendance showed a slight decrease with increasing age, from 75% at age 50 years to 65% at age 70 years. The attendance rate for women aged 40-49 years was 75% for all ages. We took into account that not all women will attend the screening every round. The percentage of attenders that will attend the next screening is estimated at 88% and the percentage of women that did not attend but will attend next time is estimated at 24% for each screening round, based on the pilot project and the national screening programme in The Netherlands (de Koning *et al.*, 1995).

The positive predictive values of an advice for biopsy were estimated to be 0.35 for the first screening and approximately 0.55 for subsequent screening examinations. The sensitivity was assumed to be as high as in The Netherlands and was dependent on the size of the tumour (T stage). The characteristics of the screening programme in Catalonia as used in the MISCAN model are summarized in the upper part of *table 2*. The age-specific detection rates were predicted by the MISCAN model given the input parameters clinical stage distribution, incidence of breast cancer and sensitivity of screening (lower part of table 3.2).

(Sensitivity) analyses

The basic scenario was screening with an interval of 2 years for women aged 50-64 years. In the sensitivity analyses, the age groups and intervals were changed to determine the impact on mortality reduction of breast cancer in the total female population and on the CE ratio. The intervals were varied from 1 to 3 years for the age groups 50-64 and 50-69 years. For the age groups 40-64 and 45-64 years only scenarios with a 2-year interval were analysed; because of the shorter pre-clinical phase for the age group 40-49 years a 3-year interval was assumed to be too long. We also studied the impact of a 12% lower sensitivity and an average attendance rate of 50% for women aged 50-69 years.

Recently, an updated analysis of Swedish results showed a reduction in breast cancer mortality of 23% (95% CI 0.59-1.01) in women of 40-49 years of age at entry invited for screening compared to women not invited, which was much higher than the relative risk (RR) of 0.90 (95% CI 0.65-1.24) in 1993 (Nystrom, 1996; Nystrom *et al.*, 1993). As an optimistic variant we used an improvement in prognosis for women aged 40-49 years as high as for women aged 50 years and over (de Koning *et al.*, 1995).

Costs

For this analysis, the Dutch cost structure was used, because extensive research on this aspect has been carried out in The Netherlands (de Koning *et al.*, 1992; de Koning *et al.*, 1991). All costs due to the screening, assessment and treatment of breast cancer were considered (de Koning *et al.*, 1991). We adapted this to the Spanish situation by using the gross domestic product-purchasing power parities (GDP-PPP) of 1991(OECD (Organisation for Economic Co-operation and Development), 1992) and the most recently published correction factor for health care PPP of 1990 (OECD (Organisation for Economic Co-operation and Development), 1993). All costs are presented in pesetas (Pts, exchange rate in 1991 100 Pts ~ $ = 1) (OECD (Organisation for Economic Co-

Table 3.3
Costs due to breast cancer for the total programme (1995-2021) of screening women aged 50-64 years at an interval of 2 years (5% discounting)

		Costs (*10^6 Pts)				
	Average cost (Pts)	Screening	(%)	No screening	(%)	Difference (total cost)
Screening[1]	4,000	12,237	(11)	–	–	12,237
Assessment[2]						
– Screening	454,530	3,236	(3)	–	–	3,236
– Outside screening	324,620	21,211	(19)	22,861	(23)	–1,650
Primary treatment plus follow-up[3]	599,300	31,737	(29)	30,511	(31)	1,226
Palliative treatment	1,808,960	41,355	(38)	43,953	(45)	–2,598
Total		109,776	(100)	97,324	(100)	12,451

[1] Included invitations, two-view mammography on first screening, one-view at subsequent screens and double reading by radiologists.
[2] Included clinical mammography and biopsy for 12% of referred women, taking into account difference in cost for biopsy of palpable and non-palpable breast cancer.
[3] Consisted of breast- conserving therapy plus radiation, mastectomy, adjuvant therapy, treatment of stage IIIB/IV and cost of follow-up visits after treatment.

operation and Development), 1992). The categories of costs for the total programme with a duration of 27 years are listed for the basic scenario in table 3.3.

The costs and effects were presented using a discounting rate of 5%, representing a time preference. Based on all effects and costs in a screening situation compared to a non-screening situation, CE ratios were calculated. For policy decisions the differences between scenarios are often expressed in marginal CE ratios. This ratio represents the extra costs to gain one extra life-year compared to the reference scenario.

RESULTS

Screening policies for women aged 50-64 years

In Catalonia screening programme for women aged 50-64 years with an interval of 2 years (eight invitations), the detection rates predicted by the model were 3.3 per 1,000 examinations for the first screening and 2.2 per 1,000 for subsequent screening during the first 5 years (lower part of table 3.2). The mortality from breast cancer in the total female population is eventually expected to be reduced by 12%, which would mean that 157 deaths from breast cancer per year would be prevented in this scenario (table 3.4). The costs per life-year gained were estimated to be 730,000 Pts (table 3.4). If the screening is carried out with a 1 or 3 year interval, the mortality reductions would be

Table 3.4
Effects and costs of screening women aged 50-64 years, different policies

Scenario (age group, interval and number of invitations per woman)	Breast cancer mortality reduction (no discounting)[1]			Costs and cost-effectiveness (5% discounted)		
	%	N (per year)	Life-years gained	Total cost (*10^6 Pts)[3]	Life-years gained	CE ratio (Pts)
50(1)64 (15 invitations)[2]	15.8	207	104,505	21,121	22,864	923,800
50(2)64 (eight invitations)	12.0	157	78,593	12,451	17,049	730,300
50(3)64 (five invitations)	8.6	112	55,088	8,627	11,991	719,500
50(2)64 (sensitivity -12%)	10.9	143	71,107	12,395	15,479	800,700
50(2)64 (attendance 50%)	8.9	116	57,854	9,637	12,570	766,700

[1] Steady state after approximately 25 years from the start of the screening programme.
[2] 50(1)64 is a screening programme carried out for women aged 50-64 years with a 1 year interval.
[3] Refers to all costs due to breast cancer with screening minus all costs due to breast cancer without screening (see *table 3*).

16 and 9% and the CE ratios would be 924,000 and 719,000 Pts respectively. Intensifying the programme from a 2 year to a 1 year interval would result in an extra mortality reduction of only 33% [(16%-12%)/12%].

The scenarios were analysed with a 12% lower sensitivity or an attendance rate of 50% on average. The mortality reductions for these scenarios were 11 and 9% respectively and the CE ratios were 801,000 and 767,000 Pts respectively (table 3.4). These estimates show that a lower attendance rate would result in a proportionally lower reduction in mortality from breast cancer, but did not much influence the CE ratio.

Other screening policies

The estimates if the less intensive programme (age 50-64 years, interval 3 years and CE ratio 719,000 Pts) were to be extended to other age groups and/or intervals are summarized in table 3.5. A programme for women aged 50-69 years and a 3 yearly interval would have a CE ratio of 694,000 Pts and the marginal CE ratio would be 614,000 Pts indicating that the costs per life-year gained for women aged 65-69 years are less than for women of 50-64 years of age. Screening with a higher frequency (50-64 years and 2 year interval) would result in a marginal CE ratio of 756,000 Pts. Saving an extra life-year will cost almost the same amount of money if the screening interval for the age group 50-64 years is 2 years instead of 3 years (table 3.5). A comparison of the screening scenarios 50-64 years with an interval of 2 years and 50-69 years with an interval of 3 years showed an almost equal effectiveness (mortality reductions 12 and 12.3% respectively). The marginal CE ratio for the scenario 50-64 years with an interval of 2 years is, however, much higher; to save an extra life-year will cost over 1 million Pts (table 3.5).

Screening with an interval of 2 years until the age of 69 years would result in a mortality reduction of 15% and the costs per life-year gained would be 744,000 Pts, with a marginal CE ratio of only 844,000 Pts per life-year gained (table 3.5).

The first variant of the improvement in prognosis was based on the RR of dying from breast cancer of 0.90 (95% CI 0.65-1.24) from the Swedish overview by Nystrom et al (Nystrom *et al.*, 1993). In this scenario an extension to younger ages (45-64 years) did not much influence the mortality reduction (12.2% instead of 12.0%) and resulted in a higher CE ratio of 868,000 Pts. The marginal cost-effectiveness with the scenario 50-64 years with a 2 year interval as a reference shows that the costs per extra life-year saved are much less for an extension to the age group 65-69 years than for an extension to the age group 45-49 years (table 3.5). Screening from the age of 40 years would result in a high CE ratio of over 1 million Pts per life-year gained.

Table 3.5
Cost-effectiveness and marginal cost-effectiveness for an extension to other age groups

Scenario (age group, interval and number of invitations per woman)	Breast cancer mortality reduction[1]		Costs and cost-effectiveness (5% discounted)			
	%	N (per year)	Total costs (*10^6 Pts)	Life-years gained	CE-ratio (Pts)	MarginalCE ratio (Pts)
50(3)64 (five invitations)[2]	8.6	112	8,627	11,991	719,500	
50(3)69 (seven invitations)	12.3	161	10,926	15,734	694,400	614,200[3]
50(2)64 (eight invitations)	12.0	157	12,451	17,049	730,300	756,000[3]
50(3)69 (seven invitations)	12.3	161	10,926	15,734	694,400	
50(2)64 (eight invitations)	12.0	157	12,451	17,049	730,300	1,159,700[4]
50(2)64 (eight invitations)	12.0	157	12,451	17,049	730,300	
50(2)69 (ten invitations)	14.9	195	14,477	19,447	744,400	844,500[5]
45(2)64 (ten invitations)	12.2	159	15,240	17,559	867,900	5,468,600[5]
40(2)64 (13 invitations)	12.7	167	19,512	18,566	1,050,900	4,654,600[5]
Improvement in prognosis for women aged 40-49 years as high as for women aged 50-69 years (optimistic variant)						
45(2)64 (ten invitations)	13.8	180	14,947	20,438	731,400	736,500[5]
40(2)64 (13 invitations)	15.1	198	19,198	23,127	830,100	1,110,100[5]

[1] Steady state after approximately 25 years from the start of the screening programme.
[2] 50(3)64 is a screening programme carried out for women aged 50-64 years with a 3 year interval.
[3] With 50 (3) 64 as a reference.
[4] With 50 (3) 69 as a reference.
[5] With 50 (2) 64 as a reference.

In the other variant, women aged 40-49 years had the same improvement in prognosis after screen detection as women aged 50-69 years (table 3.5). The CE ratio of screening women aged 45-64 will be 731,000 Pts, which is almost the same value as for the age group 50-64 years (both with 2 year intervals). This is also represented by the marginal CE ratio (the difference is only 6000 Pts). For a screening programme for women aged 40-64 years the CE ratio becomes less favourable (830,000 Pts) which is also shown by the marginal CE ratio. The effectiveness based on mortality reduction would be 13.8 and 15.1% for both these programmes, which is more favourable when compared to 12.2 and 12.7% with the pessimistic assumptions in improvement in prognosis.

DISCUSSION

Due to uncertainty about the effectiveness of screening women aged 40-49 years resulting in controversy about introducing screening in that age group, we have carried out an optimistic and pessimistic scenario for the effectiveness under the age of 50 years. Extension to the age of 40 years resulted even with the high improvement in prognosis in a high CE ratio, because of a lower incidence rate of breast cancer and a lower assumed sensitivity of mammography in the younger ages.

The CE ratios for screening started at the age of 45 years are not so unfavourable compared to the CE ratio of a 50-64 years programme with an interval of 2 years. These relatively favourable CE ratios are, however, partially explained by the high incidence peak at ages 45-49 years in Catalonia (see figure 3.1), which is preceded by a high onset of pre-clinical disease and, thus, a high pre-clinical prevalence. Introducing screening for this age group would result in relatively many screen-detected cancers, although the sensitivity of mammography is lower for younger women. At age 50 years the incidence in Catalonia was relatively low, which is rather unfavourable for screening and the increase in incidence after the age of 50 years is less steep than in other European countries. These combined effects resulted in a CE ratio favourable for screening the age group 45-49 years in comparison with 50-64 years.

Under the optimistic assumptions of effectiveness under the age of 50, extending a 50-64 years programme and a 2-year interval to younger ages (45 years) is comparable to extending to older ages. Extension to older or younger ages was almost equally cost-effective, but the reduction in mortality from breast cancer is higher for extending to the age of 69 years (15 versus 14%). The number of life-years is, however, higher for an extension to younger ages implying that more years per life will be gained.

Women aged 64-69 years still have a long life expectancy and screening has been proven to be effective in the age group 60-69 years (Nystrom *et al.*, 1993). Due to the uncertainty of the effectiveness of screening younger women and the proven benefit for older women with a relatively favourable CE ratio an extension to higher ages should be preferred rather than an extension to ages under 50 years.

The mortality reductions we predicted should be regarded as estimates and some factors not taken into account in this analysis could affect these estimates. Future changes in the mortality or incidence in breast cancer were not accounted for. In the UK a decrease in mortality was shown before the screening could have had its effect on mortality (Quinn and Allen, 1995). The widespread use of tamoxifen during this period in the UK may be important. Another aspect not considered in this analysis is opportunistic screening, before introducing screening in the target population. These processes might also play a role in Catalonia and would probably result in less potential effectiveness of screening, although it is very difficult to quantify these effects. The numbers of the Catalonian pilot project were much too small to derive the quality of the screening programme. However, the prevalence:incidence ratio in Navarra was approximately 3.7 for the age group 45-64 years and the stage-distribution was as favourable as in the Dutch national screening (Ascunce *et al.*, 1994; de Koning *et al.*, 1995; van den Akker-van Marle *et al.*, 1997). The early indicators of efficacy in the Spanish setting seem not unfavourable.

The method used with regard to the costs may not represent the true value of resource use in the Spanish situation. Using the GDP-PPP and a correction for health care might be too approximate for the specialised field of breast cancer screening. Another aspect is that no correction was made for quality of life. From other analyses it is clear that the number of quality adjusted life-years (QALYs) lost due to screening on a national level was almost equal to the number of QALYs saved by screening (Beemsterboer *et al.*, 1994; de Koning *et al.*, 1991). Correction for quality of life would not change the conclusions of this analysis.

In countries where mortality from breast cancer is increasing (La Vecchia *et al.*, 1992; Sanchez *et al.*, 1994) and the introduction of breast cancer screening is a policy priority, the diffusion of opportunistic screening is usually rapid in younger age groups where the effectiveness is less clear (Almazan *et al.*, 1995). In that situation, the importance of the benefits of breast cancer screening in the older age group should be the incentive for policy making.

References

Almazan, C., Borras, J.M., de la Puente, M.L., Gimenez, A. and Granados, A., [Approach to the use of screening mammography in 2 health regions]. [Spanish]. *Aten Primaria*, 15, 452-6 (1995).

Ascunce, N., del Moral, A., Murillo, A., Alfaro, C., Apesteguia, L., Ros, J., Abascal, L., Aizcorbe, M., Dominguez, F., Inigo, J.J. and et, a., Early detection programme for breast cancer in Navarra, Spain. *Eur J Cancer Prev*, 1, 41-8 (1994).

Beemsterboer, P.M., de Koning, H.J., Warmerdam, P.G., Boer, R., Swart, E., Dierks, M.L. and Robra, B.P., Prediction of the effects and costs of breast-cancer screening in Germany. *Int J Cancer*, 58, 623-8 (1994).

Borras, J.A. and Galceran, J., Results from the cancer registry in Tarragona (in Spanish), Registre de cancer de Tarragona, Tarragona (1993).

Collette, H.J., Day, N.E., Rombach, J.J. and de Waard, F., Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *Lancet*, 1, 1224-6 (1984).

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M.M. and van der Maas, P.J., Quantitative interpretation of the age-specific mortality reductions from the Swedish breast cancer screening trials. *J Nat Canc Inst*, 87, 1217-1223 (1995).

de Koning, H.J., Fracheboud, J., Boer, R., Verbeek, A.L., Collette, H.J., Hendriks, J.H., van Ineveld, B.M., de Bruyn, A.E. and van der Maas, P.J., 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*, 60, 777-80 (1995).

de Koning, H.J., van Ineveld, B.M., de Haes, J.C., van Oortmarssen, G.J., Klijn, J.G. and van der Maas, P.J., Advanced breast cancer and its prevention by screening. *Br J Cancer*, 65, 950-5 (1992).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

Forrest, P., Breast Cancer Screening: Report to the Health Ministers of England, Wales, Scotland and Northern Ireland, Department of Health and Social Security, London (1987).

La Vecchia, C., Lucchini, F., Negri, E., Boyle, P., Maisonneuve, P. and Levi, F., Trends of cancer mortality in Europe, 1955-1989: III, Breast and genital sites. *Eur J Cancer*, 28A, 927-98 (1992).

Marzo, M., Analysis of breast cancer screening in our context (in Spanish), Autonomous University of Barcelona (1995).

Nystrom, L., Update of the statistical analysis of screening women aged 40-49 in the combined Swedish trials, *Congress Proceedings of the International Conference The Impact of Breast Cancer Screening with Mammography in Women Aged 40-49*, Falun, Sweden (1996).

Nystrom, L., Rutqvist, L.E., Wall, S., Lindgren, A., Lindqvist, M., Ryden, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J. and et, a., Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. *Lancet*, 341, 973-8 (1993).

OECD (Organisation for Economic Co-operation and Development), Health data file. *CREDES, OECD, Paris* (1992).

OECD (Organisation for Economic Co-operation and Development), *Health systems (vol II), The socio-economic environment & statistical references*, p. 59-68, CREDES, Paris (1993).

Quinn, M. and Allen, E., Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. United Kingdom Association of Cancer Registries [see comments]. *Bmj*, 311, 1391-5 (1995).

Sanchez, V., Borras, J.M. and Mingot, M., [Cancer mortality trends in Catalonia: 1975-1990]. [Spanish]. *Med Clin*, 102, 606-12 (1994).

Servei de programmes especials, Programme of breast cancer prevention in the Valencia region (in Spanish), Conselleria de Sanitat i Consum, Valencia (1993).

Servei d'Informacio Sanitaria, Analysis of mortality in Catalonia, 1991 (in Spanish), Departament de Sanitat i Seguretat Social, Barcelona (1993).

van den Akker-van Marle, M.E., Reep-van den Bergh, C.M., Boer, R., Del Moral, A., Ascunce, N. and de Koning, H.J., Breast cancer screening in Navarra: interpretation of a high detection rate at the first screening round and a low rate at the second round. *Int J Cancer*, 73, 464-9 (1997).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

Verbeek, A.L., Hendriks, J.H., Holland, R., Mravunac, M., Sturmans, F. and Day, N.E., Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet*, 1, 1222-4 (1984).

Viladiu, P., Izquierdo, A., Beltran, M. and Bosch, X., Epidemiology of breast and gynaecologic cancer, 1980-89 (in Spanish), Spanish cancer association, Girona (1994).

• Part II •

Quantification of secondary effects of the introduction of a national breast cancer screening programme

4

RADIATION RISK OF MAMMOGRAPHY RELATED TO BENEFIT IN SCREENING PROGRAMMES: A FAVOURABLE BALANCE?

Summary

Objectives: To estimate the number of breast-cancer deaths induced by low dose radiation in breast-cancer screening programmes compared to numbers prevented.
Methods: A computer simulation-model on the natural history of breast cancer was combined with a model from BEIR-V on induced breast-cancer mortality from low levels of radiation. The improvement in prognosis due to screening was based on the results of the Swedish overview of the randomised screening trials for breast cancer and the performance of screening in the Netherlands. Different scenarios (ages and intervals) were used to explore the objectives. Sensitivity analyses were carried out for latency period, dose of mammography, sensitivity of the screening test, early detection by screening of induced breast tumours and new 1996-risk estimates by Howe and McLaughlin.
Results: For a screening programme, age group 50-69, 2-year interval, 2 mGy per view, the balance between the number of deaths induced versus those prevented was favourable; 1:242. When screening is expanded to the age group 40-49 with a 1- or 2-year interval the results may be less favourable, i.e. 1:66 and 1:97. According to these scenarios and with the Dutch scenario as reference, 1 breast-cancer death from radiation may be expected to occur in order to save 8 extra deaths from breast cancer. If screening was equally effective in young women as in women aged 50-69, the marginal value was 1:±30. Assuming detection of induced cancers by screening could influence the ratios by about 30%, but did not substantially alter the conclusions. The new risk estimates by Howe and McLaughlin resulted in 5-8 times favourable ratios breast-cancer deaths induced to prevented. Besides age group of screening, dose of mammography is the other determinant of risk.
Conclusions: For screening under the age of 50, the balance between the number of breast-cancer deaths prevented by screening versus the number induced by radiation seems less favourable. Credibility-intervals were however wide, because of many uncertainties of radiation risk at very low doses.

Acknowledgements

We would like to thank G.J. van Oortmarssen and P.J. van der Maas for their comments.

INTRODUCTION

Different studies have demonstrated the risk of inducing breast-cancer deaths through high doses of radiation. In atomic bomb survivors the Relative Risk of dying from breast cancer for an organ-absorbed dose of 1 Gy was estimated at 2.19 (90%CI; 1.56-3.09) during the period following the explosion (1950 to 1985)(Shimizu *et al.*, 1990). Women treated for tuberculosis by fluoroscopy in Sweden and Canada as well as women treated for benign breast disease by radiation in the United States were found to have developed more breast cancers than expected (Hrubec *et al.*, 1989; Miller *et al.*, 1989; Shore *et al.*, 1986). Long term follow-up of radiation therapy for benign breast disease in Sweden has shown a statistically significant increase of breast cancer among women 40 years of age and older at first treatment (Rate Ratio 3.58 (95%CI; 2.77-4.63)(Mattsson *et al.*, 1993). The average breast dose for the total group of women was 5.8 Gy in that study.

These studies have shown a statistically significant effect of total dose, age at exposure and in some studies, time since exposure, on breast-cancer incidence and/or mortality. Radiation induced breast-cancer (death) at lower levels is a controversial issue ever since. A major problem with estimates of the effects of low-dose radiation (1 mGy-100 mGy) is that a very large follow-up study is required to show a statistically significant effect. The studies on the effects of low dose radiation that have so far been carried out are not yet conclusive (Anonymous, 1994; Boice *et al.*, 1995), so for the time being risk estimates are based on extrapolations of high doses. In a large case-control study among radiologic technologists (Boice *et al.*, 1995) it was concluded that the contribution to the risk of developing breast cancer of prolonged exposure to relatively low doses of ionising agents was too small to be detectable (RR=1.13; 95%CI: 0.79-1.64).

In a screening programme for women aged 50-69 with a screening-interval of 2 years, women having mammography would receive an expected lifetime dose of 100-300 times less than the doses for which a relation with breast-cancer risk was shown. In a screening programme, however, many women are exposed to very low doses of radiation. Radiation risks were assumed to be negligible in the national screening programmes in the Netherlands and the UK (Netherlands, 1987; Vessey, 1991). It appeared however that in the Netherlands larger doses were being administered (1-2 mGy per film) (Law, 1991; Young and Ramsdale, 1993) than previously assumed (0.5 mGy per examination). In the UK a study was conducted in which the number of cancers

induced was compared to the number of cancers detected, taking into account breast thickness and number of films needed (Law, 1995).

In this study the number of breast-cancer deaths induced by radiation due to mammography in a screening programme was directly related to the number of breast-cancer deaths prevented by that programme. Different scenarios (ages and intervals) within a breast-cancer screening programme might account for a different balance between the number of breast-cancer deaths induced and breast-cancer deaths prevented by the programme. The risk of induced breast-cancer deaths has been shown to be dependent on the age at exposure, the time since exposure and on the total accumulated dose that might become especially important if screening programmes are extended to women under the age of 50.

METHODS

Risk estimate

The model developed for breast cancer in the BEIR-V-report was used to estimate the risk of dying from breast cancer due to low dose mammography (BEIR(V), 1990). The BEIR-V Committee is a committee of the National Research Council that carried out extensive research to advise the US-government on the health consequences of radiation exposure. The BEIR-model is based on parallel analysis of mortality data of two cohorts: the Canadian Tuberculosis Fluoroscopy study and a subcohort of the Radiation Effects Research Foundation (RERF) Life Span Study for which doses based on the most up to date dosimetry system (DS86) were available (Miller *et al.*, 1989; Shimizu *et al.*, 1987). We gave preference to the BEIR-model, firstly because the effect estimates were based on two data sets with different ethnic populations and secondly because the dataset of the patients treated for tuberculosis included older women who had received fractionated doses which is more applicable in the quantification of the effects of screening. The BEIR-model is a Relative Risk model, in which the extra breast-cancer mortality is a percentage of the level of the underlying breast-cancer mortality. The important modifying factors are "time since exposure" and "age at exposure". In the model, the excess risk decreases with increasing age at exposure and also varies with time since exposure, with a maximum of approximately 20 years after exposure (BEIR(V), 1990). The latency period used was 10 years. Therefore the excess risk $\gamma(d)$ of dying from breast cancer after a single exposure to dose d is:

$$\gamma(d) = \gamma_0(1 + \alpha_1 d\, g(\beta))$$

with γ_0, age-specific risk of dying from breast cancer without radiation

and

$$g(\beta) = \exp(\beta_1 \ln(T/20) + \beta_2 \ln^2(T/20) + \beta_3 (E-15)) \text{ for } E > 15$$

with E, age at exposure
T, years since exposure and for $T < 10$, $g(\beta) = 0$
d, absorbed glandular dose in Sievert (Sv)

The estimated parameters (with standard errors between brackets) are: α_1 = 1.220 (0.610), β_1 =-0.104 (0.804), β_2 =-2.212 (1.376), β_3 =-0.0628 (0.0321). The unit of the dose is Sievert (Sv), but for Roentgen, which is the type of radiation in mammography, the weight factor equals 1 and thus 1 Sv = 1 Gy.

Benefit of screening programmes

The computer simulation MISCAN-model (MIcrosimulating SCreening ANalysis) on the natural history of breast cancer was used to predict the number of breast-cancer deaths prevented (van Oortmarssen, 1995; van Oortmarssen *et al.*, 1990). In the model, breast cancer has 4 invasive screen-detectable preclinical stages according to T-stage of the TNM classification of breast cancer and one non-invasive stage (≤0.5cm, 0.5-1cm, 1-2cm, >2cm and dCIS). The natural history of breast cancer is modelled as a progression through these stages. The duration in the different stages follows an exponential distribution. Key parameters of the performance of screening are the duration of the screen-detectable preclinical disease, sensitivity of mammography and improvement in prognosis. The model-assumption on sensitivity and mean duration of the screen-detectable preclinical stages had been validated with data from the Dutch pilot screening-projects in Utrecht and Nijmegen (Collette *et al.*, 1984; Verbeek *et al.*, 1984). Furthermore the model had been shown to adequately predict the early effects of national breast-cancer screening in the Netherlands, e.g. age-specific detection rates and stage distribution (de Koning *et al.*, 1995). Demography in the Netherlands and the epidemiology of breast cancer (clinical stage-distribution, incidence, mortality, survival) and the characteristics of the screening programme (attendance, interval, age groups) were taken into account (table 4.1).

The improvement in prognosis for screen-detected cases in the model was based on the results of the Swedish overview of the randomised trials. Based on the Relative Risks (RR) this was estimated to be less for women aged 40-49 than for women 50 years and older (table 4.1)(de Koning *et al.*, 1995). The mortality reduction for screening

women aged 40-49 at entry was 10% (RR=0.90 (95%CI;0.65-1.24). For women aged 50-59 and 60-69 at entry, respectively it was 28% (RR=0.72 (95%CI;0.58-0.90) and 31% RR=0.69 (95%CI;0.54-0.88)) (Nystrom *et al.*, 1993). We also made predictions for women 40-49 years, assuming the same effect of breast-cancer screening on mortality from breast cancer as for older women (50-69), since there is controversy over screening younger women and the CI in the Swedish overview was wide.

Dose

As part of an extensive quality control system in the Netherlands, the dose of the 40 mammography devices used in the national screening-programme was assessed daily with a Plexiglas phantom of 5 cm thickness. The average surface dose was 13.88 mGy for 28 kV (kiloVoltage) (Thijssen, 1993) which is the usual power for a mammography in the Netherlands. With the power (kV), the Half Value Layer (HVL) and the surface

Table 4.1
Incidence, mortality of breast cancer and characteristics of screening in the Netherlands

Crude Incidence (per 100.000)	106,1 [1]	
European Standardized Incidence (per 100.000)	100,6 [1]	
European Standardized Mortality (per 100.000)	39,0 [2]	
Attendance age 50-69	70% (decreasing from 75% to 65%)	
age 40-49	75%	
Sensitivity of screening (50-69)[3,4]		
DCIS	0.40	
invasive tumor ≤ 0.5 cm	0.65	
invasive tumor 0.5-1cm	0.80	
invasive tumor 1-2cm	0.90	
invasive tumor >2cm	0.95	
Improvement of prognosis due to screen detection[5]	*age 40-49*	*age 50-69*
DCIS	1.000	1.000
invasive tumor ≤ 0.5 cm	0.310	0.892
invasive tumor 0.5-1cm	0.230	0.814
invasive tumor 1-2cm	0.070	0.567
invasive tumor >2cm	0.050	0.395

[1] Source: Incidence of cancer in the Netherlands, 1989 (first report of the Netherlands cancer registry)
[2] Source: Central Bureau of Statistics, 1989
[3] Sensitivity: probability of positive screen result when screening a woman with screen-detectable preclinical cancer
[4] For women aged 40-44 and 45-59 these values were 60% and 80% of the values for women aged 50-69, resp.
[5] Reduction in risk of dying from breast cancer compared to a situation without screening, detected in that stage

dose (in Roentgen), the absorbed glandular dose (in mGy) in the breast can be calculated using values for the Normalized Glandular Dose (Dgn, in mGy/Roentgen). The Dutch national guidelines for the HVL recommend values between 28mm and 38mm. The accompanying Dgn values established by Wu et al. were 1.12 to 1.48 for a mammograph with power of 27-29 kV (Wu *et al.*, 1991). The average dose per one-view mammography was calculated as being between 13.88/8.73*1.12=1.78 mGy and 13.88/8.73*1.48=2.35 mGy in the Netherlands (1 Roentgen = 8.73 mGy in air). In our calculations, an average glandular dose of 2 mGy on the breasts for one-view mammography is assumed. At the first screening, a two-view mammography was carried out and a one-view mammography at a subsequent screening.

Analyses

In this study the number of breast-cancer deaths induced by radiation was directly related to the number of breast-cancer deaths prevented by screening in a dynamic population. The number of breast-cancer deaths induced by a screening programme was calculated by multiplying the age-specific life-time risks (BEIR-model) and the radiation dose of screening mammography with the number of women receiving a mammographic examination per 5-year age-category (MISCAN-model). In the BEIR-V report Geometric Standard Deviations (GSDs) due to sampling variation, model misspecification and non-model factors like population differences and the dosimetry-system were reported, resulting in 95% credibility intervals (95%CI) around the predicted number of induced breast-cancer deaths. The precision of the number of deaths prevented is in accordance with the 95% Confidence Intervals of the Relative Risks from the Swedish trials (see also under 'BENEFITS OF SCREENING PROGRAMMES').

The screening scenario implemented at a national level in the Netherlands, i.e. screening women aged 50-69 with an interval of 2 years was defined as the baseline scenario. This was also the reference for calculating marginal results. The results of the scenarios were presented for a screening programme introduced in 1990 and carried out during a 27-year period. Effects resulting from this programme were also accumulated beyond the 27-year period.

The effects of different factors influencing the balance of the number of breast-cancer deaths prevented and induced were analysed. First the effects of age group and intervals in screening programmes were predicted. Another variant was analysed in which stage specific values for sensitivity were all lowered by 12%, based on screening data from Germany (Beemsterboer *et al.*, 1994). The risk of radiation was proportional to the dose in the BEIR-model (an example is shown). The latency period in the BEIR-model as well as the effectiveness of screening women under age 50 was varied in

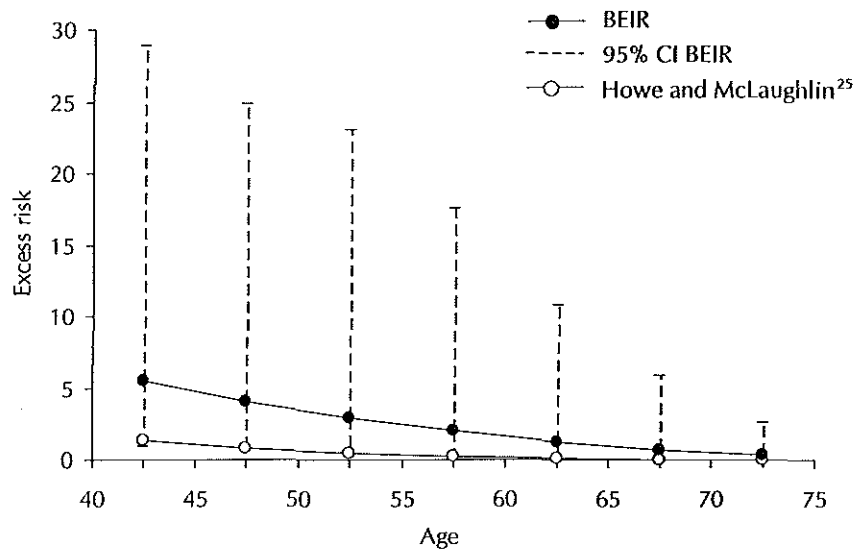


Figure 4.1
Expected lifetime risk of breast cancer death by age at screening per 1 million women exposed to 1 mGy (95% Credibility Intervals), BEIR-V (1990) and Howe and McLaughlin (1996)

other sensitivity analyses. Recently a study was published on radiation risk with a 7-years longer follow-up of the Canadian Tuberculosis Fluoroscopy study (one of the 2 cohorts as used by the BEIR-Committee) (Howe and McLaughlin, 1996). We also showed a scenario, based on the model proposed in that study

Women with radiation induced cancers can only partly benefit from the same screening programme. The reason for this is that a fair amount of these induced cancers will not be detected by screening, either because they emerge at older ages, or they are diagnosed between screening examinations. These factors taking into account, a linearly decreasing mortality reduction of 30% for cancers induced by exposure at age 40 to 0% for cancers induced by exposure at age 70 (upper age-limit of screening) was assumed. The results were shown in a sensitivity analysis.

RESULTS

The relationship between the risk of breast-cancer deaths induced by radiation and age is shown in figure 4.1. Based on the BEIR-model and the Dutch demography and mortality from breast cancer, women aged 40-44 who are exposed to 1 mGy, have a life-

time excess risk of dying from breast cancer of 5 per million. These lifetime risks decrease with increasing age at exposure. Using the recently published update by Howe and McLaughlin resulted in much lower age-specific radiation risks (Howe and McLaughlin, 1996).

In the Netherlands, the total expected number of breast-cancer deaths prevented through a 27-year screening programme for women aged 50-69 and at an interval of 2 years is 19,000, that is approximately 700 breast-cancer deaths prevented each year in a steady state situation. The reduction in breast-cancer mortality in the total female population is expected to be 17% (de Koning *et al.*, 1995). With the same programme, 79 breast-cancer deaths were expected to be induced by radiation, yielding a ratio of breast-cancer deaths induced and prevented of 1:242 (table 4.2). The 95% credibility intervals of the number of breast-cancer deaths induced, based on the GSDs of the BEIR-model were however wide.

Table 4.2
Number of breast cancer deaths induced and prevented in breast cancer screening programs for different age groups, intervals and improvement in prognosis

Scenario		Total no. deaths induced (95% CI)[1]	No. deaths induced /10^6 screens	Ratio (induced: prevented)	Marginal (induced: prevented)[2,3]
Screening ages and intervals					
BASELINE SCENARIO	50-69, interval 2 years	79 (10-627)	5.1	1:242	
	40-69, interval 2 years	208 (34-1,318)	7.9	1:97	1:8
	40-69 40-49, interval 1 year 50-69, interval 2 years	322 (55-1,942)	8.6	1:66	1:8
	40-69 interval 1 year	387 (63-2,468)	7.3	1:73	1:30
Scenarios with the improvement in prognosis for under age 50 as high as for 50-69					
	40-69 40-49, interval 1 year 50-69, interval 2 years	322 (55-1,942)	8.6	1:80	1:27

[1] 95%CI = 95% Credibility Interval, based on the Geometric Standard Deviations by BEIR-V (1990).
[2] With the scenario of screening women aged 50-69 with a 2-year interval as a reference
[3] Marginal: the extra number of breast cancer deaths prevented divided by the extra number of breast cancer deaths induced

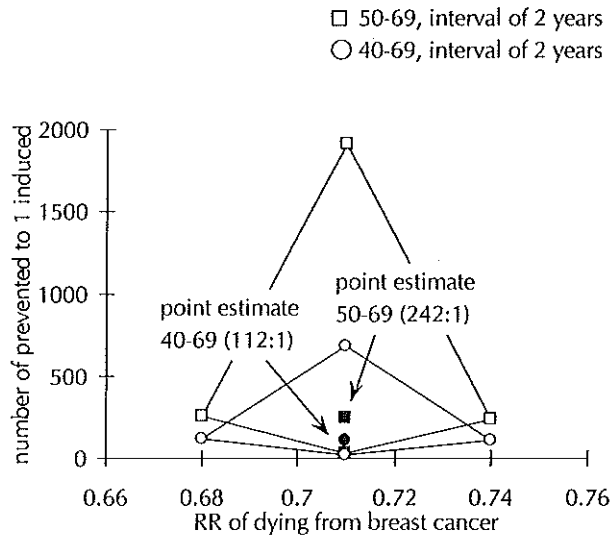


Figure 4.2
Number of prevented breast cancer deaths to 1 induced with 95% Credibility Intervals of the BEIR-V report and the 95% Confidence Intervals of the mortality reduction in the Swedish trials for 2 scenarios (baseline scenario and the scenario of screening women aged 40-69 with an interval of 2 years)

According to the study design of the trial for women under age 50 in the UK the effects of a scenario of screening women aged 40 to 49 with a 1-year interval and for the age group 50-69 with a 2-year interval were calculated. The ratio of the number of breast-cancer deaths induced and breast-cancer deaths prevented was about 4 times worse than the baseline scenario (table 4.2). We also assumed the same improvement in prognosis for women under 50 as for women aged 50-69. The number of breast cancers induced did not change but the number of women who benefited from screening increased, resulting in a more favourable ratio of breast cancer deaths induced to prevented (table 4.2). If all women aged 40-69 were screened with a 2-year interval this ratio improved even more, while a 1-year interval would lead to a ratio of 1:73 (table 2). The results of the marginal differences between the scenario of screening women aged 50-69 with a 2-year interval and the scenarios with an extension of these programmes to younger women with a 1- or 2-year interval are of special interest. In order to prevent 8 extra deaths from breast cancer, 1 breast-cancer death from radiation may be expected to occur in both scenarios (table 4.2). In the scenario with better prognosis for women aged 40-49 this marginal ratio was much more favourable for a 1-year and a 2-year interval for women under 50 (table 4.2).

Table 4.3
Number of breast cancer deaths induced and prevented in breast cancer screening programs for different sensitivity analyses

Scenario	Total no. deaths induced (95% CI)[1]	No. deaths induced /10^6 screens	Ratio (induced: prevented)
Scenarios with prevention of dying from radiation-induced breast cancer by screening			
50-69 interval 2 years	65 (8-519)	4.2	1:294
40-69[2] 40-49, interval 1 year 50-69, interval 2 years	237 (40-1,451)	6.3	1:108
Scenario with a 12% lower sensitivity			
50-69 interval 2 years	79 (10-630)	5.1	1:220
Scenario with a higher dose, 4mGy per view			
50-69 interval 2 years	158 (20-1,254)	10.1	1:121
Scenarios without a latency period (a) and latency period of 15 years(b)			
(a) 50-69 interval 2 years	81 (10-647)	5.2	1:236
40-69 [2] 40-49, interval 1 year 50-69, interval 2 years	329 (56-1,985)	8.8	1:78
(b) 50-69 interval 2 years	66 (9-522)	4.2	1:289
40-69 [2] 40-49, interval 1 year 50-69, interval 2 years	283 (49-1,686)	7.6	1:91
Scenarios with the model based on Howe and McLaughlin [3]			
50-69 interval 2 years	10	0.7	1:1912
40-69 [2] 40-49, interval 1 year 50-69, interval 2 years	67	1.8	1:317

[1] 95%CI = 95% Credibility Interval, based on the Geometric Standard Deviations by BEIR-V (1990)
[2] Improvement in prognosis for under age 50 as high as for women aged 50-69
[3] 95%CI could not be assessed, because no overall standard error of the model was given in the article of Howe and McLaughlin (1996)

With the 95% confidence limits of the RR of the reduction in mortality for the age group 50-69 from the Swedish trials, we calculated the range of the number of breast-cancer deaths prevented. The results of 2 scenarios (baseline scenario and age-group 40-69, interval of 2-years) are presented in figure 4.2, together with the 95% credibility intervals of the number of induced cancer deaths. The same confidence boundaries of the RR of mortality reduction were used for both scenarios, because the improvement in prognosis for women under age 50 was assumed as high as for women 50-69 of age. The boundaries of the RR of mortality reduction did not greatly influence the ratio induced to prevented, while the estimate of induced cancer deaths did. Due to large credibility intervals the worst ratio induced to prevented was 1:31 while the most favourable estimate was 1:1910 in the baseline scenario. The ratio prevented to induced is more favourable for the baseline scenario, but the absolute number of induced cancer deaths is however higher for the scenario for women aged 40-69 resulting in less wide credibility boundaries and thus a smaller area in figure 4.2.

The ratios between the induced and prevented breast-cancer deaths taking into account prevention of dying from radiation induced breast cancer by screening are shown in table 4.3. These values did not deviate very much from the values where screen detection was not taken into account (tables 4.3 and 4.2).

With lower sensitivity (12%) the number of induced breast-cancer deaths will not change because approximately the same number of women is screened. The number of breast-cancer deaths prevented will however be lower, since less breast cancers are detected, resulting in less cancers prevented per one induced.

The number of breast cancers induced in the BEIR-model is proportional to the dose of mammography resulting in twice as many induced cancers with a double dose (4mGy per view, table 4.3). The effect of variation in the latency period did not greatly influence the results (table 4.3). The scenarios with the estimate in radiation risk based on the study by Howe and McLaughlin resulted in much lower (up to 5 times) numbers of cancers induced (table 4.3).

DISCUSSION

Screening women under 50 is still a controversial topic, also according to the result of the NIH Consensus Panel. They recently concluded that the data currently available do not warrant a universal recommendation for mammography for all women in their forties (National Institutes of Health Consensus Development Panel, 1997). Women should be informed and decide for themselves about the balance of effects and risks.

The balance between the number of breast-cancer deaths induced by radiation and those prevented by early detection might be significantly more unfavourable for this age group.

Compared to the baseline scenario of screening women aged 50-69 with an interval of 2 years (ratio induced:prevented = 1:242), it was shown that the ratio between the number of breast-cancer deaths induced and those prevented became less favourable when screening was extended to women aged 40-49 with a 2- or 1-year interval (1:97 and 1:66, respectively). The marginal ratio for both these scenarios is even much worse, i.e. 1:8 (reference: baseline scenario).

There are a number of reasons for this unfavourable marginal ratio. Firstly we assumed there would be less improvement in prognosis for women aged 40-49 than for women aged 50-69. This assumption was based on the initial Swedish trial results (de Koning *et al.*, 1995). If the improvement in prognosis for women under 50 was assumed to be as high as for women of 50-69 years of age, the balance became more favourable, but was still 2-3 times worse than for women aged 50-69 years (table 2). "Age at exposure" is an important determinant of the risk of radiation with higher lifetime risks of radiation for younger women (BEIR(V), 1990). Another reason is that the absolute effects of screening under 50 years are less because the mortality and underlying incidence is lower. In this age group more examinations have to be carried out to detect one breast cancer than in older women. Therefore more women are exposed to radiation to prevent one breast-cancer death. Finally a lower sensitivity of mammography among women aged 40-49 was assumed which also has a negative effect on the balance.

The recently published model by Howe and McLaughlin (Howe and McLaughlin, 1996) resulted in a ratio of cancer deaths induced to prevented that was 5-8 times less than the results attained using the BEIR-model. In this new model time since exposure was not taken into account. On the basis of biological plausibility with respect to time since exposure and the fact that 7 years longer follow-up of only 1 of the cohorts is unlikely to change the risk estimates by a factor 5-8, the risk estimates in the present study seem the most reasonable approach until the new BEIR-report will emerge. Furthermore, both the BEIR-Committee and Howe and McLaughlin have decided not to include the Nova Scotia cohort, who had been exposed to much higher doses, resulting in relatively more excess deaths. Including these women for the risk estimates would have resulted in higher risk estimates, in spite of the relatively small study population.

As with all studies on risk of low dose radiation there is no direct evidence of risk of doses within the mammographic range. To quantify these risks nonetheless, extrapolations from high to low doses should be made. These extrapolations also include as-

sumptions on fractionation versus one dose, linear versus quadratic and extrapolations across populations.

The assumptions in the BEIR-model have not been disproved in recent literature. In a study on direct estimates on cancer mortality due to low doses of ionising radiation among nuclear industry workers (Anonymous, 1994), it was concluded that the risk estimates obtained by linear extrapolation from high to low doses from the studies of atomic bomb survivors were unlikely to be substantially in error. Additionally they concluded that there was little evidence for heterogeneity of risk across study populations. This was confirmed by the analysis of the BEIR-committee in which no significant difference was found in the Relative Risk for mortality of breast cancer between atomic-bomb survivors and patients treated for tuberculosis. Furthermore, the absolute excess risks were approximately equal (BEIR(V), 1990). Comparisons of the risk estimates of atomic bomb survivors with data from patient populations who received x-ray exposures in a few fractions and multiple fractions suggest that dose-fractionation may not be an important modifier of risk (Land *et al.*, 1980; UNSCEAR, 1994). In the BEIR-V model, a linear relationship is assumed between risk and low doses of radiation which is in agreement with radiobiological target theories (UNSCEAR, 1993) and is in accordance with other literature on this subject (Land, 1995; Mattsson *et al.*, 1995; Tokunaga *et al.*, 1994).

Some uncertainty has occurred in estimating the glandular dose. In this study we did not use real estimates of the surface dose but used the values measured with a phantom. It could be questioned whether a phantom (computable to a composition of 50% fat and 50% glandular tissue), is a good approximation of a breast. In a study by Cross the breast composition varied between 10% gland and 90% fat and 90% glandular tissue and 10% fat. The most frequent composition was 30% glandular and 70% fatty tissue. In their study of 212 women the calculated dose to the breast was between 1.5 and 2.5 mGy, using the normalised dose values of the gland by Wu et al. (Cross, 1994). Young et al. found a mean glandular dose of 1.24 mGy to the standard breast for 258 mammography systems in the UK screening programme (Young and Ramsdale, 1993). Recently they have suggested a higher target film density, which, with the same power (kiloVoltage) will result in higher doses to the breast (Young *et al.*, 1994). These reported values for doses are in the range of those we assumed in this study.

Dose of mammography is an important determinant because the induction of breast-cancer deaths is proportional with dose (table 4.3). This study illustrates the importance of good documentation of the entrance exposure to the breast. The main reason for an increase in dose during the last decade is a trade off between dose and image quality. By improving image quality (contrast) more and smaller cancers can be detected (Thijssen, 1993) which is expected to result in a higher reduction in mortality

from breast cancer. Therefore the dose should not be lowered if the image quality is negatively influenced. Based on extensive research carried out in the Netherlands on the quality control network, we assumed an optimal image quality with a dose of 2 mGy (Thijssen, 1993). Further research is needed to determine the doses given during screening mammography in relation to the thickness of the breast and the age of the women. This will enable better assessment of the risks to certain subgroups. Law has already done this in some detail with respect to breast thickness and number of films per view (Law, 1995).

Sub-optimal technical quality of mammography might also result in much larger quantities of radiation dose per examination or a lower sensitivity of mammography than assumed for the Netherlands. This affects the balance between the number of breast-cancer deaths prevented and induced substantially and is of special importance for countries where breast-cancer screening has not yet been introduced. Quality assurance measurements have to be carried out before and during the introduction of new breast-cancer screening programmes to maximise the image quality and sensitivity with a minimum dose to the breasts.

Feig has published on the same topic and has used the estimates of the BEIR-V Committee to calculate the extra number of breast-cancer deaths (Feig and Hendrick, 1993). In our calculations we directly estimated the risk of radiation and the benefit from the screening programme in the total population by using the MISCAN-model. The estimate of the mortality reduction by Feig was based on a complete screening programme while the effects of radiation were based on only one screening examination at a certain age. It is arbitrary and too optimistic to assume a benefit in life expectancy for a woman aged 40-44 with screen detected cancer of at least 20% due to a single screening (Feig and Hendrick, 1993).

Although we assume a mortality reduction of screen detected breast cancers of over 50% in our model (van Oortmarssen *et al.*, 1990), the 30% reduction of radiation-induced cancers as used in this study should be judged as an upper limit. Firstly, due to the 10-year latency period and the shape of the risk function many fatal-cancers originate after the upper age limit of screening. This proportion increases with increasing age at exposure. After the age of 60, 100% of all radiation induced cancer deaths occur after the age of 70. Moreover not all induced cancers under the age of 70 will be detected, either because they are missed by screening or because fast-growing tumours appear in the interval between screens. Another reason is that not all women will attend the screening every time.

Taking into account detection of radiation induced cancers by screening, the ratio of breast cancers induced to those prevented is still 1:108 for women aged 40-69 screened with a 1- and, from age 50 with a 2-year interval. This is not substantially dif-

ferent from the situation in which the induced cancers were not detected by screening (ratio 1:80). This relatively small difference in the ratio of breast cancers induced to those prevented illustrates that the radiation risk should not be neglected by simply assuming that the women with induced cancers will be prevented from dying by screening.

The results of a study by Law et al. were not completely comparable, because the outcome measure they used was the number of cancers induced versus detected and they used other (higher) risk estimates, but the conclusions were however in agreement (Law, 1995).

The most important determinants of radiation risk in breast cancer screening programmes are dose and age group of the screening. Despite the uncertainties in risk estimation, it is clear that effectiveness of screening is not the only criterion for the introduction of a screening programme. Although the risk of radiation is very low, the radiation dose in any breast cancer screening programme should not be ignored.

References

Anonymous, Direct estimates of cancer mortality due to low doses of ionising radiation: an international study. IARC Study Group on Cancer Risk among Nuclear Industry Workers. *Lancet*, 344, 1039-43 (1994).

Beemsterboer, P.M., de Koning, H.J., Warmerdam, P.G., Boer, R., Swart, E., Dierks, M.L. and Robra, B.P., Prediction of the effects and costs of breast-cancer screening in Germany. *Int J Cancer*, 58, 623-8 (1994).

BEIR(V), *National Research Council Committee on the Biological Effects of Ionizing Radiations. Health effects of exposure to low levels of ionizing radiation*, National Academy Press, Washington D.C. (1990).

Boice, J., Jr., Mandel, J.S. and Doody, M.M., Breast cancer among radiologic technologists. *Jama*, 274, 394-401 (1995).

Collette, H.J., Day, N.E., Rombach, J.J. and de Waard, F., Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *Lancet*, 1, 1224-6 (1984).

Cross, P., Doses in mammography: from the phantom to the patient. *Australas Radiol*, 38, 20-3 (1994).

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M.M. and van der Maas, P.J., Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst*, 87, 1217-23 (1995).

de Koning, H.J., Fracheboud, J., Boer, R., Verbeek, A.L., Collette, H.J., Hendriks, J.H., van Ineveld, B.M., de Bruyn, A.E. and van der Maas, P.J., 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*, 60, 777-80 (1995).

Feig, S.A. and Hendrick, R.E., Risk, benefit, and controversies in mammographic screening, *RSNA Categorial Course in Physics*, pp. 119-135 (1993).

Howe, G.R. and McLaughlin, J., Breast cancer mortality between 1950 and 1987 after exposure to fractionated moderate-dose-rate ionizing radiation in the Canadian fluoroscopy cohort study and a comparison with breast cancer mortality in the atomic bomb survivors study. *Radiat Res*, 145, 694-707 (1996).

Hrubec, Z., Boice, J., Jr., Monson, R.R. and Rosenstein, M., Breast cancer after multiple chest fluoroscopies: second follow-up of Massachusetts women with tuberculosis. *Cancer Res*, 49, 229-34 (1989).

Land, C.E., Studies of cancer and radiation dose among atomic bomb survivors. The example of breast cancer [see comments]. *Jama*, 274, 402-7 (1995).

Land, C.E., Boice, J.D.J., Shore, R.E., Norman, J.E. and Tokunaga, M., Breast cancer risk from low-dose exposures to ionizing radiation: results of a parallel analysis of three exposed populations of women. *J Natl Cancer Inst*, 65, 353-376 (1980).

Law, J., Patient dose and risk in mammography. *Br J Radiol*, 64, 360-5 (1991).

Law, J., Risk and benefit associated with radiation dose in breast screening programmes–an update. *Br J Radiol*, 68, 870-6 (1995).

Mattsson, A., Ruden, B.I., Hall, P., Wilking, N. and Rutqvist, L.E., Radiation-induced breast cancer: long-term follow-up of radiation therapy for benign breast disease [see comments]. *J Natl Cancer Inst*, 85, 1679-85 (1993).

Mattsson, A., Ruden, B.I., Palmgren, J. and Rutqvist, L.E., Dose- and time-response for breast cancer risk after radiation therapy for benign breast disease. *Br J Cancer*, 72, 1054-61 (1995).

Miller, A.B., Howe, G.R., Sherman, G.J., Lindsay, J.P., Yaffe, M.J., Dinner, P.J., Risch, H.A. and Preston, D.L., Mortality from breast cancer after irradiation during fluoroscopic examinations in patients being treated for tuberculosis. *N Engl J Med*, 321, 1285-9 (1989).

National Institutes of Health Consensus Development Panel, National institute of health consensus development conference statement: breast cancer screening for women ages 40-49, January 21-23, 1997. *J Natl Cancer Inst*, 89, 1015-1025 (1997).

The early detection of breast cancer (in Dutch), Health Council of the Netherlands, The Hague (1987).

Nystrom, L., Rutqvist, L.E., Wall, S., Lindgren, A., Lindqvist, M., Ryden, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J. and et, a., Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. *Lancet*, 341, 973-8 (1993).

Shimizu, Y., Kato, H. and Schull, W.J., Studies of the mortality of A-bomb survivors. 9. Mortality, 1950-1985: Part 2. Cancer mortality based on the recently revised doses (DS86). *Radiat Res*, 121, 120-41 (1990).

Shimizu, Y., Kato, H., Schull, W.J., Preston, D.L., Fujita, S. and Pierce, D.A., Life span study report 11, part 1. Comparison of risk coefficients for site-specific cancer mortality based on the DS86 and T65DR shielded kerma and organ doses., Hiroshima radiation effects research foundation (1987).

Shore, R.E., Hildreth, N., Woodard, E., Dvoretsky, P., Hempelmann, L. and Pasternack, B., Breast cancer among women given X-ray therapy for acute postpartum mastitis. *J Natl Cancer Inst*, 77, 689-96 (1986).

Thijssen, M.A.O., The assessment and control of image quality in diagnostic radiology (in Dutch) [Thesis], Katholieke Universiteit Nijmegen, the Netherlands (1993).

Tokunaga, M., Land, C.E., Tokuoka, S., Nishimori, I., Soda, M. and Akiba, S., Incidence of female breast cancer among atomic bomb survivors, 1950-1985. *Radiat Res*, 138, 209-23 (1994).

UNSCEAR, *United Nations Scientific Committee on the Effects of Atomic Radiation. Sources and effects of ionizing radiation.*, United Nations, New York (1993).

UNSCEAR, *United Nations Scientific Committee on the Effects of Atomic Radiation. sources and effects of ionizing radiation.*, United Nations, New York (1994).

van Oortmarssen, G.J., Evaluation of mass screening for cancer: a model-based approach, Erasmus University Rotterdam, the Netherlands (1995).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

Verbeek, A.L., Hendriks, J.H., Holland, R., Mravunac, M., Sturmans, F. and Day, N.E., Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet*, 1, 1222-4 (1984).

Vessey, M., *Breast cancer screening 1991: evidence and experience since the Forrest report. A report of the Department of Health, Advisory Committee. NHS Breast Screening Programme*, NHSBSP Publications, Sheffield (1991).

Wu, X., Barnes, G.T. and Tucker, D.M., Spectral dependence of glandular tissue dose in screen-film mammography. *Radiology*, 179, 143-8 (1991).

Young, K.C. and Ramsdale, M.L., Performance of mammography equipment in the UK breast screening programme, *8th European Congress of Radiology*, pp. 89, Springer International (1993).

Young, K.C., Wallis, M.G. and Ramsdale, M.L., Mammographic film density and detection of small breast cancers. *Clin Radiol*, 49, 461-5 (1994).

5

MAMMOGRAPHY REQUESTS IN GENERAL PRACTICE DURING THE INTRODUCTION OF NATIONWIDE BREAST CANCER SCREENING, 1988–1995

Summary

Introducing an organised breast cancer screening programme for certain age groups in a population might induce opportunistic screening in adjacent (non-invited) age groups and influence health behaviour in the target population. We analysed the effect of the start of the Dutch national screening programme on the number of mammographies requested by 43-45 general practices for the age groups 30-39, 40-49, 50-69 and 70+ years, using logistic regression analysis. In all age groups an immediate increase was observed in the number of mammography requests after the start of the screening, which was largest and statistically significant in the target-population of the screening programme (age 50-69 years). More than two years after the start of screening however the number of mammography requests in all age groups had decreased to the level of before the start and in the age group 50-69 years the number of mammographies was significantly lower than before the screening started. The unexpected increase in mammographies after the start of the breast cancer screening programme might be related to registry problems or to the process of building up the screening programme. Eventually there was a decrease in the number of mammographies in the target-population, probably an effect of the introduction of the national screening programme. Opportunistic screening was not clearly demonstrated in adjacent age groups.

INTRODUCTION

The Dutch national breast cancer screening programme started in 1989. Since then, women aged 50-69 years have been actively invited for biennial screening with mammography. The participation rate is 80% (de Koning *et al.*, 1995; Fracheboud *et al.*, 1998). Before screening was introduced, a cost-effectiveness analysis estimated a 16% reduction of breast cancer mortality in the total female population (de Koning *et al.*, 1991). Early effectiveness indicators of the screening programme are encouraging (Fracheboud *et al.*, 1998). The detection rate at first screening was approximately 6.4 per 1000 screens, in accordance with expectations. The stage distribution of the screen-detected cancers was much more favourable than for breast cancer clinically diagnosed before the start of the screening programme (de Koning *et al.*, 1995; Fracheboud *et al.*, 1998).

The expected changes in diagnostic procedures were also published (de Koning *et al.*, 1990). One hypothesis was that the number of mammographies undertaken outside the screening programme in the target-population would decrease, because of the introduction of screening. However, it was considered that the introduction of screening might induce more mammographies in adjacent (non-invited) age groups. Opportunistic screening in the target population could negatively influence the effectiveness of screening and costs of health care. Induced opportunistic screening in adjacent age groups, where the balance between favourable and adverse effects of screening is considered to be worse, could be seen as a negative effect from the public health perspective. Quantifying these mechanisms is important for a complete evaluation of the national screening programme.

In the Netherlands, the general practitioner functions as a gatekeeper to health care. All women who have symptoms of breast cancer or who are concerned about their breasts have to visit a general practitioner in order to be referred to a radiologist or surgeon for mammography. In an existing registration system of 43-45 general practices, a specific item was introduced that related to mammography practice during the build-up period of the screening programme (1988-1995). Using this registry we examined the effect of the start of the screening programme on the number of mammographies requested in general practice.

MATERIAL AND METHODS

Registration by general practices

A group of 43-45 general practices (sentinel practices) annually register consultations about specific health problems, which are collected and processed by the NIVEL (Netherlands Institute of Primary Health Care). The population covered by this registration is approximately 1% of the total population in the Netherlands. These practices are heterogeneous with regard to the degree of urbanisation and geographic area and are considered representative of the total population in the Netherlands (Bartelds, 1996). The population covered by the general practice is classified in 5-year age groups and is reassessed every 2 years.

The national screening programme was introduced gradually and complete national coverage was reached at the end of 1997. Between 1989 and 1997 screening took place in some parts of the Netherlands, but not in others.

Since 1988, the number of mammographic examinations requested by general practitioners has been recorded. The present analysis involves all mammographies for which women were referred by their general practitioner and covers the period 1988 until 1995. These mammographies cover the ones made for preventive and for clinical motives (on the basis of complaints or symptoms). These requests were linked to information on the start of the national screening programme (by month and year) in the municipalities of the general practitioners participating in the registry.

Statistical analysis

All analyses were carried out using the SPSS-package. Logistic regression analysis was used to model the chance of mammography, depending on whether or not the screening programme had started at that time. Separate logistic regression models were fitted for the age groups studied (30-39, 40-49, 50-69 and 70+ years). In the logistic model, calendar time (in months) was included as a continuous variable. General practice was also included to adjust for different levels of mammography requests.

To describe the effect of the start of the screening programme in the model, the total period covered in each general practice was divided into three intervals. The first interval ended at the start of the screening programme in each practice, the second interval covered the 2 years after the start of the programme and the last interval was from 2 years after the start until the end of the observation period. Each interval was modelled, using a separate linear trend with calendar time (termed TREND). An instant leap to a new value was allowed at the start of the screening programme, that is at the

transition from the first to the second interval (termed START). The transition from the second to the third interval however, was continuous (termed REBOUND). The difference between the level before the start of the national screening and more than 2 years afterwards was modelled by DIFF. See figure 5.1 for the general pattern of mammography.

The logistic regression model of the start of the screening was somewhat complex. We also fitted models with a 1- and 3-year interval after the start of the screening, but this did not result in a better fit. We tested the underlying trend before and after the start of the screening, but this was not statistically significant different.

RESULTS

The number of general practices participating in this study ranged from 43 to 45 per year (average duration of participation from 1988 to 1995: 6,4 years). Three practices were in municipalities where experimental screening projects had already been carried out before the start of the national screening programme in 1989. In 1995, in 5 practices, the national screening programme had not yet begun. The coverage of the national screening programme accorded well with the percentage of general practices where the screening had already started, emphasising the representativeness of the sample of the population covered by the general practitioners (data not shown).

During the whole period mammography was requested each year for an average of 2.7% of the women in the age group 40-49 years and, for 1.8% in the age groups 30-39 years and 50-69 years. For the oldest age group (70+ years) this was 0.5%. The number of mammographies by age and year showed a fluctuating pattern, but overall (and age standardised) a gradual increase occurred from 1988 to 1992 with a gradual decrease from 1993 onwards. The age group 30-39 years and 40-49 years showed fewer fluctuations than older ages. For the number of mammographies by start of the national screening programme overall (and age standardised) a higher level was observed 0-2 years after the start than before and more than 2 years after the start (table 5.1).

In table 5.2, the average number of mammographies requested by general practitioners is presented, stratified by start of the screening. Since screening had already started before 1988 in some municipalities of the general practitioners and in some the screening did not start in period studied, table 5.2 compares the results from these two groups. In all age groups, the rate of mammography requests was significantly higher if

Table 5.1
Number of participating general practices and number of mammographies by age and start of screening per year.

	'88	'89	'90	'91	'92	'93	'94	'95	'88-'95
participating general practices	45	45	44	43	43	45	45	45	
% of general practices where screening started[1]	-	4%	16%	9%	21%	24%	22%	11%	
number of mammography per 10,000 women)									
age 30-39	162	164	147	166	186	192	220	199	180
age 40-49	200	233	231	295	277	275	323	300	269
age 50-69	103	120	158	197	151	323	218	192	184
age 70+	24	25	39	55	41	87	65	59	50
before start (age stand.)	129	139	150	170	177	156	190	169	157
0-2 years after start (age standardised)	-	200	212	276	172	376	259	194	267
> 2 years after start (age standardised)	105	164	110	203	138	148	198	207	175
total (age standardised)	127	141	151	186	170	235	216	195	179
absolute numbers	471	514	542	646	572	708	718	663	4834

[1]in 3 practices the screening had already started as result of experimental screening project in '74-'75

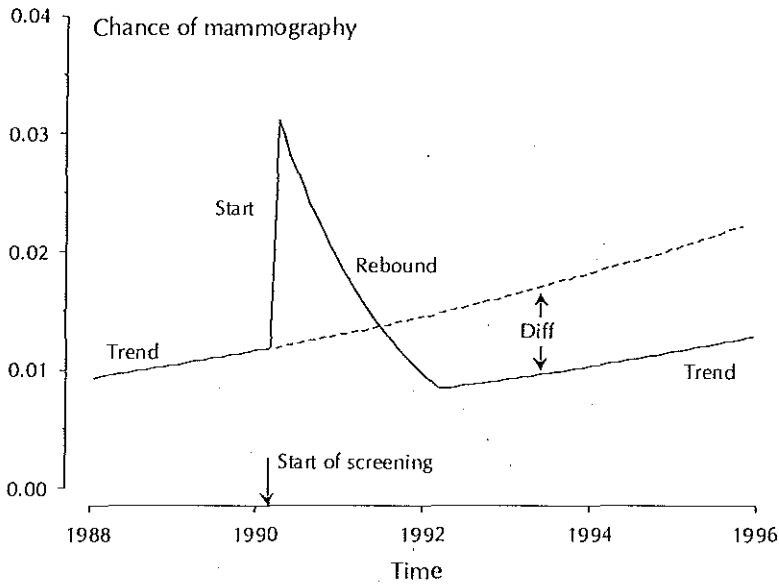


Figure 5.1
General pattern of mammography requests, age group 50–69

the screening programme had started, compared with practices where screening had not yet started. This could still, however, be only a reflection of differences by general practices or area without reference to the introduction of the screening programme.

In the logistic regression model general practice, calendar time and 3 parameters for the start of the screening programme (see Material and Methods) were included. The logistic regression model contained the same variables for all age groups in order to achieve a consistent presentation. The general pattern that emerged from the analyses is shown in figure 5.1.

Table 5.2
Number of mammographies per 10.000 women by age and start of the screening, period 1988-1995

	age group			
screening:	30-39 rate (95%CI)	40-49 rate (95%CI)	50-69 rate (95%CI)	70+ rate (95%CI)
not (yet) started	166 (156-176)	237 (224-250)	151 (142-160)	36 (30-42)
started	208 (192-223)	329 (309-350)	238 (224-253)	70 (60-80)

Table 5.3
Relation between start of the screening, age and mammography requests with adjustment for calendar time and general practice (separate age-models, see also figure 5.1)

	age group			
	30-39 OR (95%CI)	40-49 OR (95%CI)	50-69 OR (95%CI)	70+ OR (95%CI)
underlying TREND (per year)	0.99 (0.96-1.02)	1.03 (1.00-1.06)	1.12 (1.09-1.16)	1.14 (1.05-1.24)
START	1.12 (0.90-1.38)	1.29 (1.08-1.54)	2.79 (2.40-3.24)	1.74 (1.13-2.67)
REBOUND (0-2 years; trend per year)	1.05 (0.92-1.20)	0.88 (0.79-0.98)	0.50 (0.46-0.55)	0.84 (0.66-1.07)
DIFF (level before START - level after REBOUND)	1.26 (1.01-1.56)	0.93 (0.77-1.13)	0.56 (0.46-0.69)	0.94 (0.57-1.56)

For the youngest age group (30-39 years), almost none of the parameters corresponding to intervals of the start of screening were statistically significant. For all age groups, an increase after the start of the screening was observed, but it was most prominent for the age groups 50-69 years. The overall time trend in mammography requests was also largest for the age group 50-69 years and for the 70+ group (12% and 14% increase per year, table 5.3).

During the third interval (more than 2 years after start of screening), the level of mammography was lower in all age groups, except the youngest. This reduction in mammography examinations was largest and statistically significant for the target population (age 50-69 years) of the national screening programme (odds ratio (OR) 0.56, 95% confidence interval (CI) 0.46-0.69), but not significant in the adjacent age groups (table 5.3). It should be noted that this is over and above the increasing background trend (see figure 5.1).

DISCUSSION

The introduction of the national screening programme has significantly influenced the number of mammography requests by general practitioners, except in the age group 30-39 years. The number of requests first increased after the start and then decreased. As expected the number of mammography requests was significantly lower more than 2 years after the start of the programme in the target age group of the national screening programme (50-69 years) than before. In the adjacent age groups, the number of mammography was also lower more than 2 years after the start of the screening programme, but the difference was small and not statistically significant.

Increased opportunistic screening under the age of 50 years after the introduction of the screening programme could not be demonstrated in this data. The temporary increase after the start of the national screening programme was less prominent in this age group than for women aged 50-69 years. We would have expected a permanent increase in the number of mammographies that might last more than 2 years after screening had started in the case of opportunistic screening. In some European countries (Sweden, Iceland) screening is carried out also for women under the age of 50 years. Screening women under 50 years is still a controversial topic, according to the results of the NIH Consensus Panel. This panel recently concluded that the data currently available do not warrant a universal recommendation for mammography for all women in their forties (Anonymous, 1997).

Until 1993, women aged 70 years and older were allowed to participate in the screening programme. From 1994 onwards, these women were excluded from the pro-

gramme. This could have resulted in more mammographies via the general practitioner. Furthermore, these measures have directed attention to screening older women that could itself have resulted in more requests for mammography.

Although it has to be noted that some of the mammographies as requested by the general practitioner were for preventive motives and some for diagnostic motives, we conclude that opportunistic screening in adjacent age groups did not increase after the introduction of the national screening programme for women aged 50-69 years.

During the period of introducing screening for women aged 50-69 years, a continual increase in mammographies was however observed in those aged 40 years and over. This could be caused by more breast awareness generated by the programme and other factors as was suggested in part to be the cause of the decreased mortality from breast cancer in the UK, predating the effect of screening (Stockton *et al.*, 1997). In the period before the introduction of the screening programme the total number of mammographies (radiology departments included) had however already shown an increase in the Netherlands, implying that other factors also act on this trend. Furthermore this increasing background trend was very small in the age group 40-49 years, who would be expected most sensitive to breast awareness resulting from the screening programme.

The temporary rise (followed by a decline) in the number of mammography request 0-2 years after the local introduction of the national screening programme is somewhat surprising for the age group 50-69 years. When asked about this result, the participating general practitioners said they were confident that only mammographies outside the national screening programme had been registered, even in this age group. These were, however, self-reported data (data not shown).

The increase could be an artefact caused by extra alertness of the general practitioners to register mammographies after the start of screening. This would imply that the real number of mammography requests before the start and possibly also more than 2 years after screening was introduced was underestimated. Another explanation of these results might be that women who had not yet received an invitation, were encouraged by publicity surrounding the introduction of screening in their municipality to have a mammography via their general practitioner. Furthermore, women who already have breast problems or symptoms might not wait for the screening invitation but consult their GP immediately. It is also possible that the general practitioners refer, for a clinical mammography, women who did not attend screening. A small loss in attendance at the population-level could have a relatively high impact on mammography numbers in general practice. Literature regarding these issues is scarce. Garstin and colleagues observed a 42% increase in mammographies after the start of the national screening programme in the UK, which was mainly caused by referral by menopausal

clinics and general practitioners (Garstin *et al.*, 1993). In that study, concern was raised that double screening might take place or a false sense of security would occur.

The decrease in mammography requests by the general practitioner in the target population of the Dutch national screening programme can be interpreted as an effect of the introduction of screening. Part of this reduction may be due to false reassurance after a negative screen result. Still, a considerable number of breast cancer cases is diagnosed in the interval between 2 screening examinations (Day *et al.*, 1995; Tabar *et al.*, 1992). False reassurance might result in fewer cancers detected in the interval and, thus, an undesirable delay of diagnosis and treatment. Additional (qualitative) research is needed to unravel further the occurrence of false reassurance and to further interpret the sudden rise (and decline) as observed in this study.

The population covered by the general practices of the NIVEL is representative for the total Dutch population with respect to degree of urbanisation and geographic spread (Bartelds, 1996). The general practitioners may not be representative of all general practitioners. Participating in the registration is on a voluntary basis and it could be argued that this implies some selection. These general practitioners could, for example, be more restrictive in referring for mammography. Such selection would influence the overall level of mammography referrals, but is unlikely to lead to other conclusions. Some general practices contributed more to the time trend before the start of the screening programme whilst others contributed more to the trend after the start. By including general practice as a confounding variable, we adjusted for this.

Only 1.8% of all women have a mammography via their general practitioner each year, whilst 80% of all women aged 50-69 years attend the biennial screening in the Netherlands. The relative changes, as represented by the OR's, observed in this study were quite large, but have to be interpreted within this perspective.

In this study, the start of a national screening programme had a large temporary effect on the level of mammography requests by general practitioners. We have no information on whether this is caused by a change in requests by the women themselves, a change in policy of the general practitioner, or both. A decrease of mammography requests in the target population of the screening (age 50-69 years) was observed after some time. It may reflect the increased uptake of breast cancers and true negative results by the national screening programme and certainly supports the hypothesis that a screening programme reduces the number of clinical mammographies. Our study shows that at the same time it seems not to result in a permanent increase in opportunistic screening in women from adjacent age groups.

References

Anonymous, National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40-49, January 21-23, 1997. National Institutes of Health Consensus Development Panel. *J Natl Cancer Inst*, 89, 1015-26 (1997).

Bartelds, A.I.M., Continue morbiditeits registratie Nederland, 1995, NIVEL (Netherlands Institute of Primary Health Care),, Utrecht (1996).

Day, N., McCann, J., Camilleri-Ferrante, C., Britton, P., Hurst, G., Cush, S. and Duffy, S., 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*, 2, 180-5 (1995).

de Koning, H.J., Fracheboud, J., Boer, R., Verbeek, A.L., Collette, H.J., Hendriks, J.H., van Ineveld, B.M., de Bruyn, A.E. and van der Maas, P.J., 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*, 60, 777-80 (1995).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

de Koning, H.J., van Oortmarssen, G.J., van Ineveld, B.M. and van der Maas, P.J., Breast cancer screening: its impact on clinical medicine. *Br J Cancer*, 61, 292-7 (1990).

Fracheboud, J., de Koning, H.J., Beemsterboer, P.M.M., Boer, R., Hendriks, J.H.C.L., Verbeek, A.L.M., de Bruyn, A.E. and P.J., v.d.M., Nation-wide breast cancer screening in the Netherlands: results of initial and subsequent screening 1990-1995. *Int J Cancer*, 75, 694-698 (1998).

Garstin, I.W., Kaufman, Z., Michell, M.J., Rodway, A., Ebbs, S.R. and Baum, M., Side-effects of screening. *Eur J Cancer*, 29A, 2150-2 (1993).

Stockton, D., Davies, T., Day, N. and McCann, J., Retrospective study of reasons for improved survival in patients with breast cancer in east Anglia: earlier diagnosis or better treatment [published erratum appears in BMJ 1997 Mar 8;314(7082):721] [see comments]. *Bmj*, 314, 472-5 (1997).

Tabar, L., Fagerberg, G., Duffy, S.W., Day, N.E., Gad, A. and Grontoft, O., Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am*, 30, 187-210 (1992).

• Part III •

Quantification of factors influencing the performance of a prostate cancer screening trial or a future programme

6

CHANGING ROLE OF THREE SCREENING MODALITIES IN THE EUROPEAN RANDOMISED STUDY OF SCREENING FOR PROSTATE CANCER (ROTTERDAM)

Summary

Objectives: In Europe a randomised screening trial was started to show the effect of early detection of prostate cancer on mortality (European Study on Screening of Prostate Cancer). In one centre (Rotterdam), the screening protocol initially consisted of 3 screening tests: Prostate Specific Antigen (PSA), digital rectal examination (DRE) and transrectal ultrasonography (TRUS). A PSA value of ≥ 4 ng/ml and/or an abnormality on the DRE and/or the TRUS were used to indicate that a biopsy was required. In this study we examined the possibilities for a more efficient screening protocol.
Methods: A logistic regression model was used to predict the number of cancers for PSA < 4 ng/ml if all men were biopsied (Predictive Index). Effects of a change in PSA cut-off on the outcomes of the screening were explored. Weights were applied to procedures and cancers to explore the possibility of expressing the differences between scenarios in one overall figure.
Results: Biopsies in men with a PSA of < 1 ng/ml and a positive DRE or TRUS were very inefficient. Applying a DRE and a TRUS only in the PSA-range 1.5-3.9 and 2-3.9 ng/ml to indicate that a biopsy was required, would result in a decrease of biopsies by 29%-36% and a decrease of 5%-8% of cancers. However DRE and TRUS were difficult to reproduce. A protocol with only PSA ≥ 3 ng/ml as a direct biopsy indication resulted in a decrease of detected cancers by 7.6% and of biopsies by 12% and a much more simple screening procedure. With the use of the Predictive Index more efficient protocols could be achieved, but this should be viewed as a preliminary finding with the disadvantage of necessitating many additional screening visits.
Conclusion: Given the fact that DRE and TRUS appeared difficult to reproduce, a change in protocol towards PSA ≥ 3 ng/ml seems acceptable. If screening proofs to be effective, a final judgement about an optimal combination of screening tests can be made.

Acknowledgements

The 'European study of Screening for Prostate Cancer ERSPC' is supported by 'Europe Against Cancer', a programme of the European Union, and Hybritech Incorporated. The Rotterdam section is financed by the Dutch Cancer Society and the Dutch Prevention Fund.

INTRODUCTION

Prostate cancer is the second leading cause of cancer deaths in men in many industrialised countries. Tests have been developed to detect prostate cancer at an early stage. In retrospective studies, Prostate Specific Antigen (PSA) has been shown to advance the diagnosis of prostate cancer (Parkes *et al.*, 1995; Stenman *et al.*, 1994). These studies cannot however address the effectiveness of screening in the community.

The goal of the European Randomised Study of Screening for Prostate Cancer (ERSPC) is to test if a significant (and relevant) mortality reduction from prostate cancer can be achieved by screening (Auvinen *et al.*, 1996). This is expected to be effected through detection of cancer at earlier stages followed by effective treatment. In the Rotterdam and Antwerp section of the ERSPC a combination of 3 screening tests were initially chosen, partly because these (combined) tests had not yet been evaluated in a population-based setting. A screening interval of 4 years was based on the ratio of prevalence in the first pilot screening to the incidence in the general population (McWhorter *et al.*, 1992; Schröder *et al.*, 1996) and the knowledge that the disease has a slow growth rate, on average. Pilot projects in Rotterdam had shown a high detection rate of 3.5%, but the percentage of biopsies was also high, with a high percentage of false positive biopsy indications (Schröder *et al.*, 1996). In order to be acceptable the positive effects of screening in terms of mortality reduction should obviously outweigh the negative effects (e.g. negative biopsies, costs and side effects of treatment). One difficulty in assessing this is the large time lag between the initial negative effects and any possible future positive effects. It is too early at present to show whether definite mortality reduction can be demonstrated, nevertheless it is important to investigate whether the combination of screening tests or cut-off levels of PSA can be changed and optimized. In this study, we analysed the balance between disadvantages (screening tests and biopsies not resulting in the detection of cancer) and the benefits (cancers detected) from a public health point of view using results from the first 8,600 men screened.

MATERIAL AND METHODS

The ERSPC in Rotterdam started after pilot projects had been undertaken in Antwerp and Rotterdam in June 1994 (Schröder *et al.*, 1995). In the Rotterdam area, all men in

the age group 55-74 were identified by means of the population registry and received a letter asking them to participate. Men who responded by returning the intake questionnaire and who provided informed consent were randomised.

Initially all men in the screening arm received 3 screening tests consisting of a PSA determination in the serum (Hybritech-assay), a digital rectal examination (DRE) and a transrectal ultrasonography (TRUS) by a resident or staff urologist or trained paramedic. Prostate volume was determined by a TRUS using 5-mm step-section planimetry. The DRE and TRUS were carried out blind with respect to the PSA-level of the participants. If one or more of the 3 tests was positive or suspicious (PSA $\geq$ 4 ng/ml, hypoechoic lesions for the TRUS or induration for the DRE) the participant was referred for transrectal biopsy. Biopsies were carried out with a TRUS-guided needle biopsy gun (Manan pro-mag® 2.2) and 18 gauge Bard® biopsy needle, after taking prophylactic antibiotics at least 1 hour before. Sextant biopsies were taken augmented by one biopsy of the hypoechoic lesion, if applicable. In cases of positive histology, men were seen by their general practitioner and referred for treatment to one of the hospitals in the region. The results presented in this study were obtained from the start of the study (June 1994) until January 1997. There was some delay between the final diagnosis and biopsies. As this would have influenced the calculations we therefore used a cut-off date when almost all biopsy results were known (N=11 biopsies were pending (1.0%), which was neglected).

From June 1994 until January 1996 the participants received all the screening tests, i.e. DRE, TRUS and PSA determination (INITIAL PROTOCOL). It became clear that below a PSA of 1 ng/ml the Positive Predictive Value (PPV) was very unfavourable, resulting in the omission of DRE and TRUS below a PSA of 1 ng/ml from February 1996 onwards. This change was supported by the Medical Ethical Committee and the Data Monitoring Committee of the trial (MODIFIED PROTOCOL). In this study we examined grounds for a further change in the screening protocol. Three options to improve the screening process were investigated. The first option was a cut-off value of PSA as a single and direct indication for biopsy. The second option was a DRE and TRUS below a certain PSA cut-off (resulting in additional screening visits for these men). Test results above this cut-off were viewed as a direct indication for biopsy. Third, we evaluated the use of the Predictive Index (PI, see below) in combination with a PSA cut-off as a biopsy indication, implying additional screening visits for men with PSA below the cut-off to determine the PI.

Statistical analyses and Predictive Index

For the statistical analysis, the SPSS-PC package for windows was used. Logistic regression analysis was performed to model the Predictive Index (PI) based on the screening-test results (DRE, TRUS, log PSA) and log prostate volume (Kranse *et al.*, in press). The PI was defined as the chance of detecting a cancer at biopsy in its dependency on log PSA, log prostate volume and the outcome of a TRUS and DRE. Above PSA ≥ 4 ng/ml the prediction of the number of cancers based on the logistic regression analysis, corresponded well with the number of cancers detected by the screening. Therefore we applied the logistic regression model to all men screened. Summarising the resulting probabilities mounted to the number of potentially detected cancers, if biopsies were performed for all men with a PSA < 4 ng/ml (Kranse *et al.*, in press).

Applying weights to procedures/events

Since the non-detection of a cancer is presumed to have more negative impact (lower chance of cancer mortality reduction) than performing an additional biopsy (impact on Quality of Life; negligible chance of mortality induced), we applied weights to the screening procedures, the biopsies and the screen detected cancers. These weights were used to explore the possibility of expressing the differences between scenarios in one overall figure. For the first visit for a PSA test we applied a weight of -1. An additional visit for the screen tests was weighted as -5. Biopsy was assumed to have a weight of -25, while detecting a cancer was weighted as +250. In this way, a higher weight indicates more positive effects for the screened population. Because a screen detected cancer could also be a cancer that would never have been diagnosed without screening, the weight was chosen quite conservatively with regard to screening. Some sensitivity analyses were performed to evaluate the influence of a change in weights on the preference between scenarios.

RESULTS

June 1994-January 1996, *initial protocol*

In table 6.1 the screen results, the number of biopsies and number of cancers detected are presented for different PSA-levels in the first 4190 men screened. Below a PSA of 4 ng/ml, most biopsies were indicated by a positive TRUS either alone or in combination

Table 6.1
Number of screening examinations, DRE and TRUS results, number of biopsies and cancers detected and positive predictive value (PPV) for different PSA-values, June 1994-January 1996 (INITIAL PROTOCOL)

		PSA (ng/ml)								
		0-0.9	1-1.9	2-2.9	3-3.9	4-5.9	6-7.9	8-9.9	≥10	total
A	number screened	1451	1339	541	298	293	119	60	89	4190
	%	35%	32%	13%	7%	7%	3%	1%	2%	100%
B	% of biopsies among screened	12%	15%	17%	15%	96%	89%	97%	92%	25%
	indicated by: DRE+,TRUS+	18%	22%	18%	18%	6%	10%	14%	23%	15%
	DRE-,TRUS+	40%	45%	44%	53%	8%	9%	7%	6%	26%
	DRE+, TRUS-	42%	33%	38%	29%	7%	10%	3%	13%	22%
	PSA only					78%	71%	76%	57%	37%
C	number of cancers	4	10	13	8	50	28	23	47	183
D	PPV of biopsy (all)	2%	5%	14%	18%	18%	26%	40%	57%	18%
	indicated by: DRE+,TRUS+	6%	7%	13%	25%	44%	100%	88%	100%	35%
	DRE-,TRUS +	0%	3%	18%	8%	30%	22%	75%	100%	11%
	DRE+,TRUS-	3%	6%	12%	31%	35%	46%	100%	55%	15%
	PSA only					13%	13%	25%	36%	17%
E	% of cancers detected among screened	0.3%	0.7%	2.4%	2.7%	17.1%	23.5%	38.3%	52.8%	4.4%

Table 6.2
Changes in number of screening visits, biopsies and cancers detected for different changes to less intensive protocols compared to the INITIAL PROTOCOL

Screening Protocol	Biopsy indication	# screening visits		# DRE & TRUS	# biopsies	# cancers	% false positive biopsies	cancer: biopsy (incremental)
		first visit	additional visit					
PSA, DRE and TRUS (INITIAL PROTOCOL)	PSA ≥ 4 or DRE/TRUS+	4190	-	4190	1039	183	20%	1:45
CHANGES IN PROTOCOL Screening Protocol	Biopsy indication	# screening visits first visit (1)	 additional visit (2)	# DRE & TRUS less (screening) (%)	# biopsies less (%	# cancers less (%)	% false positive biopsies (3)	cancer: biopsy (incremental) (4)
PSA, and for ≥ 1-3.9 ng/ml: DRE and TRUS	PSA ≥ 4 or DRE/TRUS+	4190	2178	-2022 (-48%)	-180 (-17%)	-4 (-2%)	16%	1:24
PSA, and for ≥ 1.5-3.9 ng/ml: DRE and TRUS	PSA ≥ 4 or DRE/TRUS+	4190	1396	-2794 (-67%)	-299 (-29%)	-9 (-5%)	14%	1:16
PSA, and for ≥ 2-3.9 ng/ml: DRE and TRUS	PSA ≥ 4 or DRE/TRUS+	4190	839	-3351 (-80%)	-377 (-36%)	-14 (-8%)	12%	1:7
PSA, and for ≥ 3-3.9 ng/ml: DRE and TRUS	PSA ≥ 4 or DRE/TRUS+	4190	298	-3892 (-93%)	-467 (-45%)	-27 (-15%)	10%	1: 6
PSA	PSA ≥ 4	4190	-	-4190 (-100%)	-512 (-49%)	-35 (-19%)	9%	

(1) PSA-test
(2) DRE and TRUS for men with PSA between cut-off value and 4 ng/ml
(3) $\frac{\text{\# of biopsies - \# of cancers}}{\text{\# of participants}}$
(4) reference is less intensive protocol

Table 6.3
Number of screening examinations, DRE and TRUS results, number of biopsies and cancers detected and positive predictive value (PPV) for different PSA-values, January 1996-January 1997 (MODIFIED PROTOCOL)

		PSA (ng/ml)								
		0-0.9	1-1.9	2-2.9	3-3.9	4-5.9	6-7.9	8-9.9	≥10	total
A	number screened	1594	1324	552	344	307	141	60	109	4431
		36%	30%	13%	8%	7%	3%	1%	2%	100%
B	% of biopsy among screened	-	21%	20%	33%	91%	90%	93%	95%	24%
	indicated by: DRE+,TRUS+	-	18%	14%	22%	14%	18%	13%	35%	18%
	DRE-,TRUS+		37%	34%	32%	16%	16%	11%	15%	25%
	DRE+, TRUS-		46%	52%	47%	18%	17%	25%	9%	31%
	PSA only					52%	50%	52%	42%	26%
C	number of cancers	-	28	10	33	64	35	13	59	242
D	PPV of biopsy (all)	-	10%	9%	29%	23%	28%	23%	58%	23%
	indicated by: DRE+,TRUS+		17%	0%	60%	65%	52%	71%	89%	50%
	DRE-,TRUS+		9%	10%	19%	13%	20%	33%	73%	16%
	DRE+, TRUS-		9%	11%	21%	27%	19%	29%	44%	16%
	PSA only					13%	24%	7%	27%	17%
E	detection rate among screened	-	2.1%	1.8%	9.6%	20.8%	24.8%	21.7%	54.1%	5.5%

with a positive DRE, while for PSA ≥ 4 ng/ml, most biopsies were indicated by PSA only. The overall Positive Predictive Value (PPV) increased with increasing PSA-levels, but for PSA categories < 4 ng/ml the PPVs were very low. The highest PPV (80%) was observed in men with three positive screening tests. The number of cancers divided by the number of men screened increases with increasing PSA-levels. An increase at the transition of PSA of 3-3.9 ng/ml to 4 ng/ml and above is reflects the policy that biopsies were available for everyone from PSA 4 ng/ml and upwards.

Implications of changes in protocol

The INITIAL PROTOCOL seemed clearly quite inefficient with regard to the use of a DRE and a TRUS for some categories of PSA. In table 6.2 we estimated the consequences of less intensive screening protocols, comparing the number of biopsies and screening tests, the resulting number of cancers detected and the false positive rates (FP%). Both a DRE and a TRUS have approximately the same test characteristics (table 6.1) and therefore omitting them would result in approximately the same decrease in the number of biopsies at the expense of about the same number of cancers. For both the participants and the logistics of the trial it would have been preferable to have both tests omitted below certain PSA-levels. A change in protocol with no DRE and TRUS for certain levels of PSA implies that a portion of the participants having to attend twice for the screening would have to have an additional DRE and TRUS.

With increasing cut-off levels of PSA, less additional visits would be needed, resulting in less biopsies and less cancers detected (table 6.2). All changes in the protocol would lead to a lower percentage of false positive biopsies than the INITIAL PROTOCOL, at the expense of missing screen-detected cancers (2-19%). The trade-off between both is reflected in the incremental cancer/biopsy rate, calculated as the less number of cancers detected, divided by the less number of biopsies, when the PSA cut-off is increased by one step (last column). This ratio improved with increasing cut-off levels of PSA, showing that the number of men biopsied decreased relatively less than the number of cancers detected.

February 1996-January 1997; *modified protocol*

From February 1996, no DRE and TRUS were performed in men with a PSA < 1 ng/ml on the basis of the results from table 6.2. In table 6.3, the results of the next 4431 screens are presented for the MODIFIED PROTOCOL. An absolute increase in the number of biopsies compared to the INITIAL PROTOCOL was found for PSA-levels below 4 ng/ml.

Table 6.4
Consequences of different protocols including the Predictive Index (PI) for number of screening examinations, biopsies, cancers detected (expected) and test-characteristics (FUTURE PROTOCOL (1))

Screening Protocol	Biopsy indication	# screening visits		cancers predicted/ detected		cancers de-tected	change in # of biopsies (%)	change in # of cancers (%)	PPV	% FP biop-sies(4)	total weight(5)
		first visit (2)	additional visit (3)	# biopsies	PSA <4 ng/ml	PSA ≥4 ng/ml					
PSA, and for 1-3.9 ng/ml: DRE and TRUS (MODIFIED PROTOCOL)	PSA ≥ 4 or DRE/TRUS+	8621	4398	1923	102	319	-	-	22%	17%	26564
PSA	PSA ≥ 2	8621	-	2686	141	319	+763 (+40%)	+39 (+9.3%)	17%	26%	39229
PSA	PSA ≥ 3	8621	-	1683	70	319	-240 (-12%)	-32 (-7.6%)	23%	15%	46554
PSA	PSA ≥ 4	8621	-	1094	0	319	-829 (-43%)	-102 (-24.2%)	29%	9%	43779
PSA	PSA ≥ 6	8621	-	533	0	205	-1390 (-72%)	-216 (-51.3%)	39%	4%	29304
PSA, and for 2-2.9 ng/ml: DRE and TRUS	PSA ≥ 3 or DRE/TRUS+	8621	1093	1885	93	319	-38 (-2%)	-9 (-2.1%)	22%	17%	41789
PSA, and for 2-3.9 ng/ml: DRE and TRUS	PSA ≥ 4 or DRE/TRUS+	8621	1735	1455	64	319	-468 (-24%)	-38 (-9.0%)	26%	12%	42079
PSA, and for ≥ 1 ng/ml: DRE and TRUS	PI ≥ 0.08	8621	5576	1704	118	305	-219 (-11%)	+2 (+0.5%)	25%	15%	26654
PSA, and for ≥ 1 ng/ml: DRE and TRUS	PI ≥ 0.15	8621	5576	927	64	270	-996 (-52%)	-87 (-20.7%)	36%	7%	23829
PSA, and for ≥ 1 ng/ml: DRE and TRUS	PI ≥ 0.10	8621	5576	1409	99	298	-514 (-27%)	-23 (-5.7%)	28%	12%	27529
PSA, and for ≥ 1.5 ng/ml: DRE and TRUS	PI ≥ 0.10	8621	3989	1366	92	298	-557 (-29%)	-30 (-7.4%)	29%	11%	34789
PSA, and for ≥ 2 ng/ml: DRE and TRUS	PI ≥ 0.10	8621	2913	1290	80	298	-633 (-33%)	-42 (-10.2%)	29%	11%	39069

PSA, and for ≥ 3 ng/ml: DRE and TRUS	PI ≥ 0.10	8621	1820	1109	48	298	-814 (-42%)	-75 (-17.8%)	31%	9%	41054
PSA, and for ≥ 4 ng/ml: DRE and TRUS	PI ≥ 0.10	8621	1178	877	0	298	-1046 (-54%)	-123 (-29.2%)	34%	7%	38064

(1) the percentage of biopsy advice not resulting in a biopsies, because of drug use (anti coagulants), refusal or loss to follow-up (6%)were projected the same in FUTURE PROTOCOLS
(2) PSA-test
(3) DRE and TRUS
(4) $\frac{\text{number of biopsies - number of cancers}}{\text{number of participants}}$
(5) WEIGHTS: first visit -1, 2nd visit -5, biopsy -25, cancer +250, the higher the total weight the higher the positive effect for the population screened

For PSA ≥ 4 ng/ml the observed differences between the protocols were smaller, reflecting the less fluctuating test-characteristics of PSA as a biopsy indication.

Although we expected a decrease of biopsies by approximately 17% (table 6.2), the proportion of biopsies of the MODIFIED PROTOCOL was as high as before (25% and 24%, respectively). The PSA distributions at the screening were approximately the same in both protocols, but the number of suspicious DRE and TRUS was significantly higher in the MODIFIED PROTOCOL. The percentage of suspicious DRE among the screened men (PSA ≥ 1 ng/ml) had increased from 10% to 18% while the percentage of positive TRUS had increased from 12% to 16%. The number of cancers detected also increased in these biopsies, resulting in higher PPVs and an increase in overall detection rate from 4.4% to 5.5%. In all 5-year age groups an increase in detection-rate was observed, but the largest increase was in men aged 70-74 (from 5.3% to 9.7%, data not shown).

Future protocol

In table 6.4, ways were investigated to further increase the PPV and decrease the false-positive rate with the smallest loss in cancers detected. All FUTURE PROTOCOLS were compared to THE MODIFIED PROTOCOL. In the first part of table IV, scenarios with PSA as a direct biopsy indication are shown. The second part consists of protocols in which additional visits are needed depending of the initial PSA-level and biopsy is indicated by both a positive DRE/TRUS or PSA cut-off. In the last part, screening consisted of 2 visits for a portion of men dependent on their PSA level to establish their PI, which also determined the indication of biopsy.

The total weight (last column) was most favourable for the scenario with a PSA ≥ 3 ng/ml as a biopsy indication. The scenarios where DRE and TRUS were carried out for a certain cut-off of PSA were rather favourable with respect to their total weight and the number of cancers missed. The disadvantage of these scenarios was that a portion of men (13-20%) have to attend the screening twice. The scenarios in which a PSA cut-off was combined with a PI of ≥ 0.10 as a biopsy indication were rather favourable with respect to both the lower numbers of biopsies required and fewer numbers of cancers detected. For cut-offs of 2 ng/ml and higher, the total weight was also rather high, but relatively many additional visits would be needed to determine the PI.

Sensitivity analysis for the weighing of screening tests, biopsies and screen detected cancers

Lowering the weight for detecting a cancer from 250 to 200 would favour the scenario with PSA ≥ 4 ng/ml as a direct biopsy indication in table 6.4. From a weight of 400 and

higher, the scenario with PSA ≥ 2 ng/ml as a direct biopsy indication would have the most positive weight. A negative weight of biopsy of -35 or more would favour the screening policy of PSA ≥ 4 ng/ml as a direct biopsy indication and vice versa. Changing the weight of the additional visit to -2 resulted in the highest total weight for the scenario using a PSA cut-off of 2 ng/ml in combination with the PI of ≥ 0.10 as a biopsy indication. Other scenarios would then have almost the same total weight. From these sensitivity analyses it was clear that the scenario using PSA ≥ 3 ng/ml as a direct biopsy indication was rather stable with regard to total weight, compared to other scenarios.

DISCUSSION

The results of this study show a very unfavourable balance between the number of biopsies performed and the number of prostate cancers detected under a PSA cut-off of 2 ng/ml, according to very low PPVs between 2% and 5%. A few cases of severe complications occur due to the biopsy (sepsis 0,18%) and a large number of moderate to mild complications (e.g. fever, pain, haematuria and haematospermia, 64.6% in total) (Rietbergen *et al.*, 1997). Because of the high percentage of biopsies in this screening-study, reducing this number is important. The value of the detected cancers in terms of mortality reduction is yet unknown, while side effects of treatment occur early after screen detection. The quality of life study which is being conducted alongside this screening-trial will help to derive better weights for the balance between the positive and negative side effects (Essink-Bot *et al.*, 1998).

The MODIFIED PROTOCOL (i.e. no DRE and TRUS for PSA below 1 ng/ml) was expected to result in a small decrease in detection rate (2%) and biopsies (16%). Different results were observed, which are most likely caused by a change in personnel. This change was introduced simultaneously with a training scheme with supervision. Screening using a DRE and TRUS are clearly examiner-dependent and not reproducible as a standardised procedure. Smith observed that the reproducibility of a DRE among urologists was only fair (Smith and Catalona, 1995). Therefore, extrapolations of the results of this study should be performed with caution.

The base-line characteristics such as age-distribution, rating of own health, the WHO prostate symptom score, familial prostate cancer and previous screen history did not show large differences between both protocols. The participation rate in the study was also almost equal for the INITIAL and MODIFIED PROTOCOL (45% and 48%). The compliance (attending screening after randomisation) did not differ either. These factors could not give a sufficient contribution to explain the unexpected differences.

From the data presented in this study, choosing an option that recommends a change in the protocol is rather difficult, because none of the scenarios was favourable with respect to all parameters studied. Weights were applied to explore the possibility of expressing the differences between scenarios. Choosing these weights was quite arbitrary and subjective. Conservative weights have been used here, compared to values that have been applied in breast-cancer screening, based on empirical studies (de Koning *et al.*, 1991).

Some scenarios reached high positive weights with the weights as used in this study. The protocol with PSA ≥ 3 ng/ml as a direct biopsy indication has major practical advantages, because the screening is limited to PSA testing. This test is the least invasive and a decrease in the number of cancers detected of approximately 8% (of a total detection rate of about 5%) seems acceptable.

The protocols where a PSA cut-off was combined with a DRE and TRUS for certain values of PSA also had quite positive weights. These protocols may not be preferred however, because of the poor reproducibility of DRE and/or TRUS. With the use of the Predictive Index (PI) representing the probability of detecting a cancer based on the screening results and the volume of the prostate in combination with a PSA cut-off, biopsies could be performed more efficiently, i.e. with a higher PPV and without missing many cancers. It is however judged as a preliminary finding also because not all parameters seemed reproducible and standardised. The PI should be tested prospectively before implementation. Eventually, the PI is expected to be able to increase the specificity of the screening test with only a small loss of sensitivity (Kranse *et al.*, in press).

Catalona et al. concluded that free serum PSA measurements may reduce the number of biopsies in lower PSA cut-off (Catalona *et al.*, 1997). We did not study the effects of free PSA on the sensitivity and specificity of the screening. In decision-making about screening as a health care service free serum PSA should also be evaluated as an optional screening test.

Lodding et al. showed that 25% of all screen-detected cancers had PSA values between 3 and 4 ng/ml (Lodding *et al.*, 1998). With the PI we would estimate 22% of all cancers being within this PSA range, another indication that the PI might be usable in the future. Lodding et al also concluded that a majority of these cancers were clearly significant and suitable for curative treatment, therefore providing another argument to implement a screening protocol with a PSA cut-off of 3 ng/ml (Lodding *et al.*, 1998).

As with any protocol utilising random biopsies the results may be affected by random sampling. Because it is plausible that this would affect all studied protocols in the same way it has not altered our conclusions.

Changing a protocol where cancers are not detected when previously they were, could in theory result in missing those cancers with the largest potential to contribute to

the reduction in mortality from the disease. In a study by Hoedemaeker et al. the group of tumours detected by a DRE and/or TRUS below a PSA of 4 ng/ml had the lowest pathological stages with a considerable fraction (43%) fitting criteria for 'minimal tumour' and 86% with a tumour volume smaller than 0.5 ml (Hoedemaeker *et al.*, 1997). A minimal tumour was defined as a tumour smaller than 0.5 ml, lacking a Gleason pattern 4 or 5 and being confined to the prostate (stage pT2). Men with low PSA levels are therefore most likely to harbour clinically insignificant tumours (Hoedemaeker *et al.*, 1997). The numbers studied were however small (N = 14). These data indicate that a change in protocol where a DRE and TRUS are omitted in lower PSA-ranges may not result in significant loss in potential mortality reduction, if re-screening is performed at an adequate time interval.

By using the very intensive protocol in the first instance, more is known about the DRE and TRUS in very low PSA-ranges. It was shown that these procedures are not (yet) reproducible and have made a small contribution to the number of cancers detected for PSA $<$ 2 ng/ml. It seems very difficult to change a trial protocol if the effects of these changes in terms of prevention of prostate cancer deaths are not yet known. On the basis of the data presented in this study, combined with the knowledge of the tumours in low PSA-ranges (Hoedemaeker *et al.*, 1997) all men in the ERSPC Rotterdam with a PSA $\geq$ 3 ng/ml have received a biopsy from May 1997 onwards. This change implies that biopsies were no longer driven by a DRE and/or TRUS. In the future, we will hopefully be capable of defining PSA ranges, combined with other screening tests or e.g. a Predictive Index to discriminate between the slowly growing non-life threatening cancers and the aggressive cancers. Changes in mortality resulting from randomised trials will give the only definitive answer regarding the effectiveness of prostate cancer screening.

References

Auvinen, A., Rietbergen, J.B., Denis, L.J., Schroder, F.H. and Prorok, P.C., Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. *J Med Screen*, 3, 97-104 (1996).

Catalona, W.J., Smith, D.S. and Ornstein, D.K., Prostate cancer detection in men with serum PSA concentrations of 2.6 to 4.0 ng/mL and benign prostate examination. Enhancement of specificity with free PSA measurements [see comments]. *Jama*, 277, 1452-5 (1997).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy

alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

Essink-Bot, M.L., de Koning, H.J., Nijs, H.G., Kirkels, W.J., van der Maas, P.J. and Schroder, F.H., Short-term effects of population-based screening for prostate cancer on health-related quality of life. *J Natl Cancer Inst*, 90, 925-31 (1998).

Hoedemaeker, R.F., Rietbergen, J.B.W., Kranse, R., van der Kwast, T.H. and Schröder, F.H., Comparison of pathologic characteristics of T1c and non-T1c cancers detected in a population-based screening study, the European Randomized Study of Screening for Prostate Cancer. *World J Urol*, 15, 339-345 (1997).

Kranse, R., Beemsterboer, P.M.M., Rietbergen, J.B.W., Habbema, J.D.F., Hugosson, J. and Schröder, F.H., Predictors for biopsy outcome in the European Randomized Study of Screening for Prostate Cancer (Rotterdam region). *the Prostate* (in press).

Lodding, P., Aus, G., Bergdahl, S., Frosing, R., Lilja, H., Pihl, C.G. and Hugosson, J., Characteristics of screening detected prostate cancer in men 50 to 66 years old with 3 to 4 ng./ml. Prostate specific antigen [see comments]. *J Urol*, 159, 899-903 (1998).

McWhorter, W.P., Hernandez, A.D., Meikle, A.W., Terreros, D.A., Smith, J.A., Jr., Skolnick, M.H., Cannon-Albright, L.A. and Eyre, H.J., A screening study of prostate cancer in high risk families. *J Urol*, 148, 826-8 (1992).

Parkes, C., Wald, N.J., Murphy, P., George, L., Watt, H.C., Kirby, R., Knekt, P., Helzlsouer, K.J. and Tuomilehto, J., Prospective observational study to assess value of prostate specific antigen as screening test for prostate cancer [see comments]. *Bmj*, 311, 1340-3 (1995).

Rietbergen, J.B., Kruger, A.E., Kranse, R. and Schröder, F.H., Complications of transrectal ultrasound-guided systematic sextant biopsies of the prostate: evaluation of complication rates and risk factors within a population-based screening program. *Urology*, 49, 875-80 (1997).

Schröder, F.H., Damhuis, R.A.M., Kirkels, W.J., de Koning, H.J., Kranse, R., Nijs, H.G.T. and Blijenberg, B.G., European randomised study of screening for prostate cancer, the Rotterdam pilot studies. *Int. J. Cancer*, 1996, 145-151 (1996).

Schröder, F.H., Denis, L.J., Kirkels, W., de Koning, H.J. and Standaert, B., European randomized study of screening for prostate cancer. Progress report of Antwerp and Rotterdam pilot studies. *Cancer*, 76, 129-34 (1995).

Smith, D.S. and Catalona, W.J., Interexaminer variability of digital rectal examination in detecting prostate cancer. *Urology*, 45, 70-4 (1995).

Stenman, U.H., Hakama, M., Knekt, P., Aromaa, A., Teppo, L. and Leinonen, J., Serum concentrations of prostate specific antigen and its complex with alpha 1-antichymotrypsin before diagnosis of prostate cancer. *Lancet*, 344, 1594-8 (1994).

7

PSA-TESTING AND USE OF DRE BEFORE AND DURING A RANDOMISED TRIAL FOR SCREENING OF PROSTATE CANCER (ERSPC, ROTTERDAM)

Summary

Determination of prostate specific antigen (PSA) is a simple test that might influence the outcome of the European Randomised study on Screening for Prostate Cancer (ERSPC), if frequently utilised. In this study PSA and digital rectal examination (DRE) before and during the screening trial in Rotterdam and in the general population is quantified.

The data from the intake questionnaire regarding PSA-testing and DRE were analysed to evaluate the use of these tests before participation in the screening study. Data on PSA from the Laboratory of General Practice were linked to information from participants in the ERSPC. Different sources were used to quantify PSA-tests and DREs in the general population in a situation where no screening exists.

On average, 45% of the men had had a DRE and 13% reported that they had been PSA tested before participating in the trial. Both these percentages increased with age. No statistically significant effects of former PSA-testing or DREs on the cancer detection rate could be demonstrated. The rate of PSA determinations was approximately twice as high in the control arm than in the screen arm during the trial (76 and 33 per 1000 person-years, respectively). After randomisation, the number of PSA determinations first decreased in the screening arm, but after some time an increase was observed. The number of PSA determinations increased in the control arm after randomisation. The number of PSA determinations in the general population of the same age was estimated at 45/1000 person-years.

The use of PSA-tests in the control arm was moderate, but if different men undergo this test every year the contamination rate in the control arm might be rather high during a screening interval of 4 years. In the final analysis on mortality, PSA-testing should be taken into account. The motives and outcomes of PSA-testing are useful for further interpretation.

Acknowledgements

We would like to thank Mr. Marquinie and Mr Dersjant for providing the data of the health insurance company (DSW) and Mr Bartelds for providing the data of the sentinel practices. We also thank Rob Boer for his assistance with the analysis. The 'European study of Screening for Prostate Cancer ERSPC' is supported by 'Europe Against Cancer', a programme of the European Union, and Hybritech Incorporated. The Rotterdam section is financed by the Dutch Cancer Society and the Dutch Prevention Fund.

INTRODUCTION

A PSA determination in serum is a very simple test that has been shown to be useful in following the clinical course of men treated for prostate cancer (Catalona, 1996). Its value for population screening has not yet been demonstrated, but widespread use of the test has been reported (Jacobsen *et al.*, 1995). In the United States an exponential increase in Prostate Specific Antigen (PSA) testing has been seen over the period 1988-1991, which was assumed to be the most plausible explanation for the increase in prostate cancer incidence (Potosky *et al.*, 1995). Effects of PSA screening on prostate cancer mortality however have not (yet) been shown (Wingo *et al.*, 1998). Different authors disagree about the effectiveness of screening on the basis of data about trends in incidence, stage distribution and mortality (Smart, 1997; Weyler, 1998). The value of prostate cancer screening can only be reliably demonstrated in randomised trials. In Europe and the United States large randomised trials have begun to study the effect of early detection on the mortality of this disease (Auvinen *et al.*, 1996; Gohagan *et al.*, 1994; Schröder *et al.*, 1995).

In the Netherlands, there is also evidence that the increased incidence of prostate cancer is in part a result of increased PSA use (Post *et al.*, 1998). PSA-testing in the control and to a lesser extent in the screen arm (apart from PSA-testing in the trial) might bias mortality reduction estimates in the screening trial. Because of the simplicity of the test and the observed rapid dissemination in other countries it is very important to monitor PSA use in Rotterdam (one of the centres of the ERSPC). In this study, the frequency of PSA-testing before and during the screening trial is quantified, together with the use of Digital Rectal Examination (DRE) as well as the number of biopsies resulting from PSA-determinations outside screening.

MATERIAL AND METHODS

PSA-use before inclusion in the trial

The ERSPC in Rotterdam started in June 1994, after pilot projects had been carried out in Antwerp and Rotterdam (Schröder *et al.*, 1995). In the Rotterdam area, all men in the age-group 55-74 were identified by means of the population registry and were asked by letter to participate. Men who responded by returning the intake-questionnaire and a

signed informed consent form, were randomised. The men in the screen arm were invited to be screened by PSA-testing, DRE and TRUS until May 1997. From then, the screening consisted of PSA-testing only. Some intake-questions asked about prior prostate examination(s) using DRE or PSA-testing. For this study, all data on men randomised until March 3rd 1998 were used for the analysis.

PSA use after inclusion in the trial

All data from PSA-testing were obtained from the Laboratory of General Practice (GP-lab) for the period January 1996 to March 1998. This laboratory covers the total city of Rotterdam and some suburbs, comprising approximately 1 million inhabitants in total. The PSA determinations carried out by this laboratory were not performed within the context of the ERSPC, because the latter were collected and sampled in the Central Laboratory of the Academic Medical Hospital Dijkzigt. The data of the GP-lab were linked to individual data of men randomised in the screening study. The linkage was carried out by generating an identifier consisting of the first initial, the first and last character of the surname and the day, month and year of birth. This identifier was built into both data files and thereafter both files were linked. To calculate rates per 1000 person-years, all person-years at risk from randomisation until 3 March 1998 were calculated. Persons who were randomised before 1 January 1996 were at risk during the total period from 1 January until 3 March 1998, except in instances where their screening interval of 4 years ended before that date.

PSA use, referrals and biopsies outside the trial (actual practice)

The actual practice of PSA-testing, referrals and prostate biopsies by both general practitioners and urologists was extracted from 3 different sources. For each, the source population was known allowing the calculation of rates per 1000 person-years.

One source was a Health Insurance Company (HIC), covering the Rotterdam area. These data were from January 1995 until April 1997, a period during which screening had not taken place in that area. For each individual, the number of declarations of a PSA-test, a biopsy and a fee per person, representing the treatment by a urologist, were available in chronological order. It was also known whether the PSA was requested by a urologist or a general practitioner.

Two other sources of information were the GP-Lab and the sentinel practices of the NIVEL (Netherlands Institute of Primary Health Care) (Bartelds, 1998). In 1997, DRE, PSA requests and referral to a urologist were introduced into the registration sys-

tem of 45 sentinel practices. All participating general practitioners were asked to record DREs, PSA-tests and referrals to urologists with regard to prostate disease by age.

RESULTS

Table 7.1 presents the number of men participating in the Rotterdam part of the ERSPC trial, who reported having had a PSA-test or DRE before participation. 45% reported

Table 7.1
PSA-testing and DRE before randomisation in the ERSPC (N = 28,550)[1]

	DRE		PSA	
	(%)	N of men	(%)	N of men
Ever	44	12,474	13	3,618
Never	49	14,005	77	22,038
Unknown	7	2,071	10	2,894
Ever				
age 55-59	34	3,059	9	797
60-64	44	3,047	12	849
65-69	52	3,347	15	1,007
70-74	58	2,802	19	904
unknown (non-attenders)[2]	22	219	6	61
Ever, year of participation in trial				
1993-1994	39	1,142	10	278
1995	47	3,019	14	875
1996	44	3,412	12	958
1997	45	4,097	14	1,276
1998	45	585	13	170
unknown (non-attenders)[2]	22	219	6	61
Ever, time since inclusion	100	12,474	100	3,618
0-1 year before	24	2,975	58	2,103
1-2 years before	21	2,635	15	533
2-5 years before	26	3,242	16	594
> 5 years before	29	3,622	11	388
Ever, done by	100	12,372		
General Practitioner	39	4,802	n.a.[3]	
Urologist	46	5,750	n.a.	
Other	15	1,820	n.a.	

[1] includes both screen and control arm; no differences observed in both arms
[2] because of missing date of participation due to not attending the screening, age and year of participation could not be calculated in the screen arm
[3] not asked

ever having undergone an examination of the prostate and 13% a PSA-test. The non-attenders (randomised to the screen-arm, but not showing up for screening) reported significantly less PSA-tests (6%); the percentage of DREs was also lower, but approximately 50% of the answers were missing.

Year of participation in the trial did not show a relationship with the percentage of reporting having DREs or PSA-tests, while most previous PSA-tests were performed 0-1 year before participation. These results were also analysed for the age-group 55-69 separately, which did not change the conclusions.

The impact of former PSA-testing and prostate examinations on the detection rate of cancers in the trial is shown in table 7.2. The detection rates of those men having PSA-testing or DRE was higher than of men not tested before participation, but the difference was not statistically significant. The detection rates increased if the PSA-test or

Table 7.2
Detection rates (number of cancers per 100 men screened) by former DRE and PSA-testing

	Cancers (N)	Detection rate (/100)	P-value (chi-2)
Total	663	4.8	
No DRE before	323	4.6	
DRE before, missing [1]	26	3.6	P=0.20
DRE before	314	5.2	P=0.12
Time since participation in trial			
0-1 year before	66	4.7	test for trend P=0.17
1-3 years before	64	5.1	
3-5 years before	85	5.2	
> 5 years before	99	5.8	
Done by [2]			
General Practitioner	122	5.3	P=0.21
Urologist	135	5.0	P=0.44
Other	54	6.0	P=0.06
No PSA before	508	4.8	
PSA before, missing [1]	59	4.2	P=0.31
PSA before	96	5.4	P=0.26
Time since participation in trial			
0-1 year before	50	5.1	test for trend P=0.74
1-3 years before	16	5.6	
3-5 years before	23	7.2	
> 5 years before	7	3.8	

[1] missing, meaning that the question was not filled in
[2] from N=3, information is missing

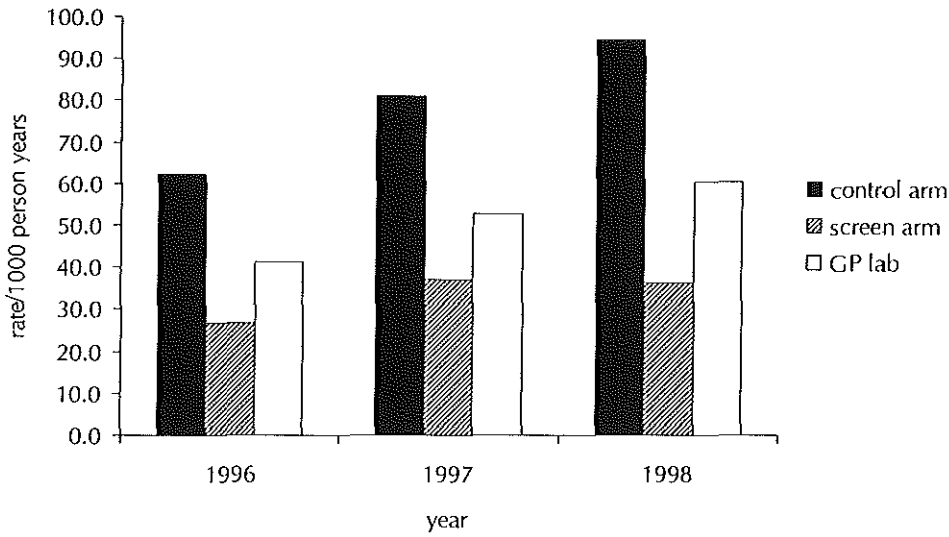


Figure 7.1
PSA-testing in screen and control arm of the trial and in GP-lab by year, age 55–69.

DRE had been performed longer ago, except for PSA, if it was performed more than 5 years before participation. These trends were not statistically significant either.

If the DRE had been carried out by a physician other than a urologist or general practitioner, the detection rate was higher and of borderline significance (P = 0.06). The most frequently reported other specialities (in order of occurrence) were internal medicine, surgery and occupational medicine. A combination of having both a PSA-test and a DRE did not show other effects on the detection rates.

Table 7.3
PSA use in screen and control arm of the trial by age at randomisation (data from GP lab), from 1/1/'96 until 3/3/'98

PSA	Screen arm	Control arm	Rate Ratio Control/Screen
Total PSA (N)	800	1629	
PSA rate (/1000 PY)	37.9	78.6	2.1
age 55-59	28.4	56.4	2.0
60-64	32.3	91.2	2.8
65-69	41.5	86.5	2.1
70+	58.2	88.6	1.5
age 55-69	33.4	76.3	2.3

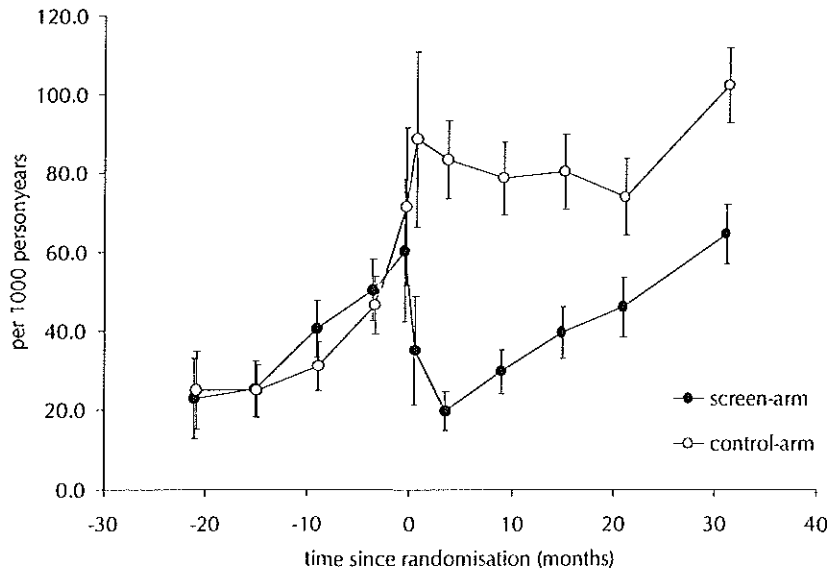


Figure 7.2
PSA use in screen and control arm before and after randomisation with 95% confidence intervals (per 1000 person-years)

In figure 7.1 and table 7.3 PSA-testing in both the screen and the control arm of the screening trial are presented. On average, the PSA rate in the control arm was more than twice as high as in the screen arm (Rate Ratio 2,1) while it was 60% higher in comparison with the GP-lab (figure 7.1). An increase of PSA-testing was observed in both the screen and control arm and the GP-lab in the period 1996-1998. The largest difference between screen and control arm, represented by a Rate Ratio of 2.8 was present in the age group 60-64 (table 7.3). The same age-pattern was seen in the separate years (data not shown).

Figure 7.2 shows that after participating in the screening trial, different patterns in PSA-testing occurred in the screening and control-arm. In the screening-arm the number first decreased, while it started to increase from 6 months after randomisation onwards. In the control arm however, an increase of PSA-testing through the general practitioner took place that continued during the total follow-up time.

In table 7.4 an overview is given of PSA-tests, DREs and biopsies in a non-screening situation. The rate of PSA examinations as requested by general practitioners obtained from 3 different sources shows quite large differences (table 4, first part). The number of examinations per 1000 men is twice as high in the GP-lab than in the Sentinel Practi-

Table 7.4
PSA, DRE and referrals by age and specialty, according to 3 different sources

	National Sentinel practices (GPs) 1997		Health insurance company 1995-1997		Laboratories of General Practice 1996-1998	
Coverage (N men, all ages)	62,802		100,152[1]		519,983[2]	
GENERAL PRACTITIONER						
PSA requested /done (N)	510		2,560		18,075	
PSA rate (/1000 PY)		8.1		11.3		16.0
age < 40		0.3		0.4		0.7
40-54		6.7		9.2		13.4
55-69		22.2		37.0		48.1
≥70		14.4[3]		60.0		85.3
DRE performed (N)	732					
DRE rate (/1000 PY)		11.6				
age < 40		1.6				
40-54		13.3				
55-69		26.2				
≥70		18,9[4]				
Referral (N)	59		445			
Referral rate (/1000 PY)		0.9		2.0		
age < 40		0		0.1		
40-54		0.2		1.2		
55-69		2.1		6.2		
≥70		3.0[5]		11.7		
UROLOGIST						
PSA (N)			3,833			
PSA rate (/1000 PY)				16.9		
age < 40				1.1		
40-54				9.4		
55-69				45.2		
≥70				111.1		
Biopsies (N)			255			
Biopsy rate (/1000 PY)				1.1		
age < 40				0		
40-54				0.2		
55-69				3.5		
≥70				8.0		
OTHER SPECIALIST						
PSA (N)			759			
PSA rate (/1000 PY)				3.5		
age < 40				0.0		
40-54				1.6		
55-69				9.8		
≥70				24.2		

[1] '96
[2] '97
[3] 70-74: 30,3/1000
[4] 70-74: 49,8/1000
[5] 70-74 8,5/1000

ces. In all 3 sources, a steep increase in the number of PSA-tests was reported from the age 55 onwards. This increase persisted in both the GP-Lab and the HIC, but a decrease was found for the age group 70 and over in the Sentinel practices. The number of DREs in the sentinel practices was on average 44% higher than the number of PSA-tests and this percentage decreased with age. A DRE resulted in more cases of PSA-testing from the age 55 and older, than in younger ages. On average, the referral rate to a urologist was only approximately 10% of the number of PSA-tests and DREs.

From the HIC data it was clear that on average approximately 20% of men who had a PSA-test were referred to a urologist and that about half of the referrals resulted in a biopsy. This latter rate also increased with increasing age. The number of PSA examinations requested by a urologist was higher than requested by a general practitioner and showed a steeper increase with age. The number of PSA-tests requested by other specialists was lower than by the general practitioner and urologist but still amounted to 11% of all PSA-tests requested.

Based on the weighted average of the 3 sources, we assumed a baseline PSA use of 15 per 1000 person-years in the total population and 45 per 1000 in the age-group 55-69 through the general practitioner in a situation without a screening programme.

DISCUSSION

The rate of PSA-testing was 76/1000 person-years in the control arm of the trial in the age group 55-69 and participation in the trial raised this level compared to the rate in the general population in the same age-group (45/1000). A rate of PSA-testing of 7.6% per year (assuming that different individuals undergo a PSA-test every year) would result in a contamination percentage of 30% in the control arm, during the first screening interval of 4 years. In the screen arm, however PSA-testing outside screening still also takes place, increasing with time since randomisation.

It can be assumed that the number of cancers detected by these PSA examinations will be much lower than in the trial. This is also in agreement with the standard of general practitioners operating in the Netherlands. General practitioners are advised to refer a patient on the basis of a PSA-test higher than 10 ng/ml, doubt about the benign character of the prostate after DRE, relevant comorbidity, haematuria without infection or acute urine retention (Klomp *et al.*, 1997). This standard will result in much fewer referrals and biopsies than in the trial. To clarify this, more insight is needed in the motives for PSA-testing. Furthermore the outcomes of PSA-testing (number of biopsies and cancers detected) should be evaluated. The final analysis of the trial should use data about PSA-testing within the screen arm and control arm to allow controlling for the

effect of contamination. Evaluation of the trial should also include analyses of trends in PSA use in the future.

Having had a PSA-test before participation in the trial without having prostate cancer diagnosed may result in healthy screenee bias. This is also supported by the fact that non-attenders (men who were randomised but never participated) had reported much lower rates of PSA-testing (4% versus 13%). This would result in a bias towards zero, meaning that the effect on prostate cancer mortality will be larger than can be assessed in the trial. The higher the proportion of men with a prior PSA-test the larger this bias will be. Fortunately the percentage of men receiving a PSA-test before participation is quite low (13%) and not clearly increasing in more recent years, although a widespread increase of PSA-testing is assumed between 1990 and 1995 (Post *et al.*, 1998). Healthy screenee bias is however partly contradicted, because men who had previously PSA tested or who had had a DRE showed higher cancer detection rates. Although these results were not statistically significant, this might represent a high-risk group.

The data obtained form the GP-Lab allowed us to link two data sources. By using the key-variable only 0,9% false matches occurred (which were deleted). By this way of linking we have of course no knowledge about the number of persons not linked because of missing initials or typing errors in initials. Our approach resulted in an underestimate of the real number of PSA-tests performed outside the trial, but it will not greatly influence the results of this study.

In some countries, a widespread use of PSA-testing has been reported especially related to the increase in incidence (Jacobsen *et al.*, 1995; Post *et al.*, 1998; Potosky *et al.*, 1995). The rates in the USA as reported by Jacobsen et al were much higher than those reported in this study. Their PSA rates in 1992 were between 150/1000 and 400/1000 for the age groups 55-69 (Jacobsen *et al.*, 1995). The effects on incidence are thus also expected to be less in the Netherlands. Post et al reported on the increase in incidence in the Netherlands, which showed the same patterns as in the US. The increase was approximately 40% between 1991 and 1995, while this increase was more than 100% between 1987 and 1992 in the US. The increase in incidence was merely attributed to more diagnoses of localised disease in both studies, although Post also demonstrated a decrease in the rate of locally advanced and metastasised disease.

Our data from the GP-lab are in agreement with the observation of Post et al that an increase in PSA-testing is occurring. In a study of PSA requests from general practitioners it was concluded that PSA-testing is not focused on higher risk groups and shows tendencies towards a screening approach. In that study, although screening was not recommended, rising levels of PSA-testing were observed, resulting in an increase of PSA results of lower than 4 ng/ml (McGing, 1998). Ward et al concluded in a study

among general practitioners that more than half of the PSA-tests ordered were screening tests (Ward *et al.*, 1998).

PSA-testing of a asymptomatic men should not be disseminated through the general population without proven benefit of screening. It has still to be established whether PSA-testing does more harm than good (de Koning and Schroder, 1998). Furthermore widespread dissemination might introduce very serious problems in interpreting the results of ongoing randomised trials. In that way the effectiveness of prostate cancer screening with PSA will only be based on surrogate endpoints, an undesirable situation.

References

Auvinen, A., Rietbergen, J.B., Denis, L.J., Schroder, F.H. and Prorok, P.C., Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. *J Med Screen*, 3, 97-104 (1996).

Bartelds, A.I.M., Continue morbiditeits registratie Nederland, 1997, NIVEL (Netherlands Institute of Primary Health Care),, Utrecht (1998).

Catalona, W.J., Clinical utility of measurements of free and total prostate-specific antigen (PSA): a review. *Prostate Suppl*, 7, 64-9 (1996).

de Koning, H.J. and Schroder, F.H., PSA screening for prostate cancer: the current controversy. *Annals of Oncology*, 9, 1293-1296 (1998).

Gohagan, J.K., Prorok, P.C., Kramer, B.S. and Cornett, J.E., Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute [see comments]. *J Urol*, 152, 1905-9 (1994).

Jacobsen, S.J., Katusic, S.K., Bergstralh, E.J., Oesterling, J.E., Ohrt, D., Klee, G.G., Chute, C.G. and Lieber, M.M., Incidence of prostate cancer diagnosis in the eras before and after serum prostate-specific antigen testing [see comments]. *Jama*, 274, 1445-9 (1995).

Klomp, M.F.L., Gercema, A.J., de Jonge-Wubben, J.G.M., Mulders, A.H.P.W., Romeijnders, A.C.M. and Rosmalen, C.F.H., NHG-Standaard bemoeilijkte mictie bij oudere mannen. Eerste herziening (in Dutch). *Huisarts Wet*, 40, 114-124 (1997).

McGing, P.G., A study of PSA requests from general practitioners received by one Dublin hospital. *Ir Med J*, 91, 61-2 (1998).

Post, P.N., Kil, P.J.M., Crommelin, M.A., Scapers, R.F.M. and Coebergh, J.-W.W., Trends in incidence and mortality rates for prostate cancer before and after prostate-specific antigen introduction. A registry-based study in Southeastern Netherlands, 1971-1995. *Eur J Cancer*, 34, 705-709 (1998).

Potosky, A.L., Miller, B.A., Albertsen, P.C. and Kramer, B.S., The role of increasing detection in the rising incidence of prostate cancer. *Jama*, 273, 548-52 (1995).

Schröder, F.H., Denis, L.J., Kirkels, W., de Koning, H.J. and Standaert, B., European randomized study of screening for prostate cancer. Progress report of Antwerp and Rotterdam pilot studies. *Cancer*, 76, 129-34 (1995).

Smart, C.R., Prostate cancer facts and fiction [editorial]. *J Surg Oncol*, 66, 223-9 (1997).

Ward, J.E., Gupta, L. and Taylor, N.J., Do general practitioners use prostate-specific antigen as a screening test for early prostate cancer? [see comments]. *Med J Aust*, 169, 29-31 (1998).

Weyler, J., Prostate cancer: screening or watchful waiting? [editorial]. *Ann Oncol*, 9, 9-11 (1998).

Wingo, P.A., Ries, L.A., Rosenberg, H.M., Miller, D.S. and Edwards, B.K., Cancer incidence and mortality, 1973-1995: a report card for the U.S. *Cancer*, 82, 1197-207 (1998).

8

ADVANCED PROSTATE CANCER; CARE AND COST IMPLICATIONS

Summary

Background: If prostate cancer screening proves to be effective, some cases will be prevented from reaching the advanced stage. In order to evaluate screening programs thoroughly, it is important to quantify course, care and accompanying costs of advanced disease.
Methods: We studied 70 files of patients in 2 hospitals, who had received a diagnosis of distant metastases of prostate cancer and had died in the years 1994-1998. The total healthcare received by these patients, including symptoms and complaints was recorded
Results: Most frequently reported symptoms were pain (42%), urogenital symptoms (25%) and malaise (20%). 89% of all patients were hormonally treated (either by orchidectomy and /or chemical castration) and 47% received one or more series of radiation therapy. 69% of all patients were treated with pain medication. The average duration of advanced disease in all patients was 24 months. Average costs of advanced disease were estimated at $11,182 over the total period. $1,547 (14%) was allocated to assessment and outpatient care and $9,635 (86%) to treatment and costs of hospital stay. Almost half of the total costs were determined by hospital stay.
Conclusions: These data give a better understanding of the course, care and costs of advanced prostate cancer. These estimates together with the effects of advanced prostate cancer on quality of life will be used for the evaluation of prostate cancer screening.

Acknowledgements

We would like to thank H. A. van Leiden for the assistance with data collection and K. el Naggar for the assistance with data collection and the causes of death evaluation. We also thank J. H. M. Blom for providing access to the patient-files of the SFG hospital.

INTRODUCTION

Screening for prostate cancer has been the subject of intense debate for many years. Still many questions have to be resolved before the value of screening can be established satisfactorily. International trials have been set up to assess the effect of prostate cancer screening and the early diagnosis and treatment, on prostate cancer-related mortality and morbidity (Auvinen *et al.*, 1996; Gohagan *et al.*, 1994). The final results of these trials are not expected until the year 2005-2010.

To support decision making, the effects of screening and early treatment of prostate cancer on long-term mortality and morbidity outcomes and costs should be assessed. In the European Randomized Study on Prostate Cancer (ERSPC) (Schröder *et al.*, 1995), a prostate cancer disease model was developed in which the course of prostate cancer through different disease states with different durations was simulated. This was similar to the model developed for breast cancer and cervical cancer (de Koning *et al.*, 1991; Habbema *et al.*, 1985; van Oortmarssen *et al.*, 1990).

If screening is effective, the advanced stages of prostate cancer will (in part) be prevented or postponed. Some patients will therefore die of other causes, without passing through the stage of advanced prostate cancer. A full evaluation of screening will thus not only cover the early stages of the disease and the short-term effect of early detection and treatment, but also the effects on the advanced stages.

In the literature, information on course of advanced prostate cancer is scarce and the populations studied were not curatively treated (Aus *et al.*, 1995; Otnes *et al.*, 1995). In the Netherlands and many other countries, localised prostate cancer is treated more aggressively. Furthermore, in these studies the patients had died between 1988-1991, while in the meantime new treatment options for advanced disease became available (Anonymous, 1995). If screening is effective in avoiding advanced prostate cancer it will also reduce the care and costs of advanced disease. In this study, the course, care and costs of advanced disease of prostate cancer were studied.

MATERIAL AND METHODS

Selection of Patient-Files

Patient-files were collected in the Academic Hospital Rotterdam and the residential hospital Sint Franciscus Gasthuis (SFG) Rotterdam. In each hospital, we used a different

procedure to identify eligible patients. In the Academic Hospital, we used the local patient flow information system. This system assigns a code to every outpatient on the basis of the extent of disease (localised, metastasised to the regional lymph nodes and distant metastases). Included were patients who had distant metastases from prostate cancer who had died between 1994 and 1997.

In the SFG hospital we checked the medical records of two groups of patients. The first group consisted of 213 patients with prostate cancer who had regularly visited the hospital in 1994. The second group were patients using LHRH-analogues, the conventional treatment option in the SFG for prostate cancer patients with lymph nodes or distant metastases. From these two groups, we only included patients with distant metastases from prostate cancer who had died between 1995 and 1998.

In the Academic Hospital, 50 patients were identified. We excluded patients whose medical records could not be traced (n = 6), those who had visited the hospital only for a second opinion (n = 6), or those who were lost to follow-up due to treatment in another hospital or a removal (n = 6). In the Academic Hospital, 32 patients were included and in the SFG hospital 38 patients.

Patients were distinguished between those who had distant metastases as a first diagnosis of prostate cancer (M_1-patients) and those who had either localised prostate cancer or positive lymph nodes as a first diagnosis, developing distant metastases in the course of time. This latter group is referred to as M_0M_1-patients.

Data Collection

Data were retrieved from the patient's medical records, from the diagnosis of distant metastases onwards. Data were recorded using a structured registration form. The following data were collected for eligible patients: resource use (e.g. outpatient visits, hospital admissions, use of analgesics, palliative surgery or radiation) and clinical data (stage at first diagnosis, location and course of distant metastases, symptoms etc.). All symptoms requiring treatment were recorded, including complaints and symptoms that were simultaneously present. The incidence of symptoms was defined as the number of periods in which the complaint or symptom occurred.

A diagnosis of distant metastases was extracted directly from the medical records. The date of diagnosis of advanced disease was based on the date of the positive bone scan with additional radiological images (as judged by the urologist and radiologist) or pathology and/or CT scan in cases where the metastases were located elsewhere. Hospital admissions and assessment clearly related to other concomitant diseases were not used in the analysis. Place and date of death was recorded but only the information from the medical records was used.

Assessing the Cause of Death

The cause of each death was assessed independently by two medical doctors on the basis of the extracted information. In cases of disagreement a consensus was reached. Three criteria were used to establish the cause of death. First, presence of invalidating metastases directly preceding death (not more than 9 months), that had to be relieved by pain medication or were treated by abstinence. Second an increasing Prostate Specific Antigen (PSA) before death, based on at least 3 measurements. The third criterion consisted of the presence of direct/indirect impairment due to metastases, such as anaemia or treatment with corticosteroids. In case at least 2 criteria were fulfilled, death was categorised as 'definitely prostate cancer'. In cases of one positive criterion, death was rated as 'probably prostate cancer'. If none applied death was considered as 'other cause'.

Cost Calculations

Costs were calculated as actual resource use multiplied by the respective cost per resource unit taking into account the duration of treatment if applicable. In this study, only the hospital based resource use was considered. If available, costs per unit of resource use were based on actual economic costs, i.e. including fixed and indirect costs (Finkler, 1982). Estimates were obtained from internal reports (de Koning *et al.*, 1990; Krenning *et al.*, 1998), literature, guidelines, or calculated from data retrieved from the hospital's budgetary and financial control system. Financial charges were used if true economic costs were not available.

Costs were converted in U.S. $ by using gross-domestic product purchasing power parity of 1996 (1$ = Dfl 2.07). If health care purchasing power parities would have been used, the costs (in $) would become much higher, because of the relatively high costs of health care in U.S. compared to the Netherlands.

Analysis

Differences between M_1- and M_0M_1-patients were tested using Student's t-test or Chi2. Kaplan Meier analysis combined with the Log-rank test was used to analyze the survival-function after the diagnosis of distant metastases. In cases where censoring was applied, patients whose death was assessed as 'other cause' were censored.

RESULTS

The baseline characteristics of the men included in this study are presented in table 8.1. The differences between both groups of patients were not statistically significant and the absolute differences were mostly small. For patients who first were diagnosed as having localised disease (M_0M_1-patients), the time to metastases was 34 months on average (median 26). The average duration of advanced disease, defined as the period from the diagnosis of metastases until death, was for all patients was 24 months (median 19). The survival of advanced disease of both M_1- and M_0M_1-patients is shown in figure 8.1.

Table 8.1
General characteristics of advanced prostate cancer for M_0M_1 and M_1 patients, from two hospitals in the Netherlands.

	M_0M_1-patients (N = 35)	M_1-patients (N = 35)	Total (N = 70)
Age at first diagnosis of prostate cancer (years)			
mean	71.4	70.6	71.0
median	72.0	73.9	72.5
min-max	51.9-86.5	39.7-83.9	39.7-86.5
T stage at first diagnosis (%)			
T1-2	20	9	14
T3-4	63	80	71
unknown	17	11	14
Grade at first diagnosis (%)			
G1-2	57	40	48
G3-4	34	43	39
unknown	9	17	13
First treatment (not for M1) (%)			
-radiation therapy	37	n.a.[1]	n.a.[1]
-hormonal	29	n.a.[1]	n.a.[1]
-WW	17	n.a.[1]	n.a.[1]
-other/unknown (N = 1)	17	n.a.[1]	n.a.[1]*
Time to metastasis (months)			
mean	33.9	n.a.[1]	n.a.[1]
median	25.6	n.a.[1]	n.a.[1]
min-max	2.9-140	n.a.[1]	n.a.[1]
Age at diagnosis of advanced disease (years)			
mean	74.2	70.6	72.4
median	74.1	73.9	74.0
min-max	53.8-87.9	39.7-83.9	39.7-87.9
Location of distant metastasis (%)			
-bone	74	89	81
-other	26	11	19

Of all metastases 81% were located in the bone, which was more often the case in M_1-patients than M_0M_1-patients. Patients with distant metastases at other locations than bone had a statistically significant shorter survival (16 versus 26 months; M_1- and M_0M_1-patients grouped, $P=0.05$). Advanced disease was generally diagnosed on the basis of urogenital complaints and symptoms as a result of metastases (e.g. loss of weight, malaise, pain, fracture). The cause of death was judged probably to be due to prostate cancer in 75% of patients. The average duration of lost to follow-up was 21 days, representing the time between last contact with the urologist and the date of death. 31% died in the hospital.

Table 8.1 – continued

	M_0M_1-patients (N = 35)	M_1-patients (N = 35)	Total (N = 70)
Event leading to diagnosis of M1 (%)			
-urogenital symptoms	20	29	24
-complaints/symptoms as a result of M1 other than urogenital	40	51	46
-screening/follow-up visit	23	3	13
-other/unknown	17	17	17
Duration of M+ until death (months)(censored)[2]			
mean	21.1 (26.9)	27.8 (33.3)	24.2 (30.6
median	14.9 (19.2)	25.6 (26.6)	18.9 (26.3)
min-max	0.1-65 (0.6-65)	0.6-127 (0.6-127)	0.1-127 (0.6-127)
Age at death (years)			
mean	76.0	72.8	74.4
median	75.1	76.3	75.7
min-max	53.8-90.9	41.9-85.1	41.9-90.9
Cause of death (%)			
-definitely prostate cancer	66	63	64
(PC)	9	14	11
-probably PC	26	23	24
-other cause			
Time between last contact and death (days) (N = 69)[3]	17	24	21
Place of decease (%)			
-hospital	31	31	31
-at home	20	9	14
-other/unknown	49	60	54

[1] not applicable
[2] men who died from other causes were censored
[3] information missing from 1 patient

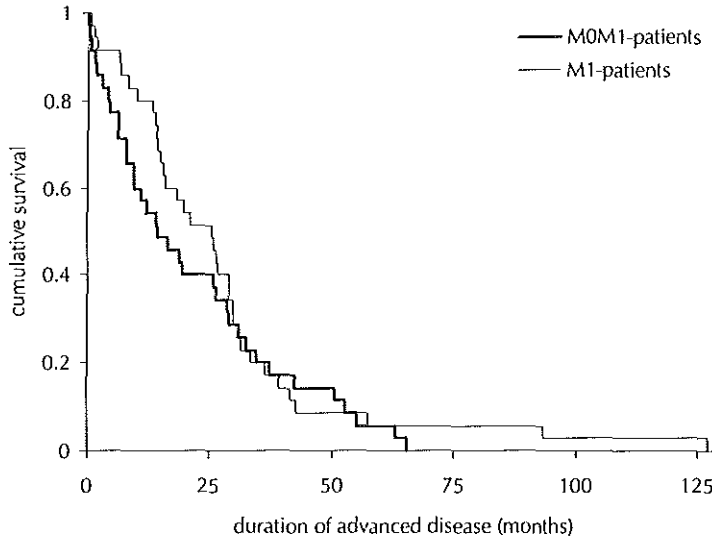


Figure 8.1
Survival of M_0M_1 and M_1 patients after diagnosis of distant metastases

Table 8.2
Incidence of symptoms in advanced prostate cancer in 70 patients (average duration 24 months)

symptoms	N (%)	(%) of total
pain	135	**(42%)**
-bone		
lower spine, SI, sacrum, pelvis	(16%)	
other locations	(27%)	
-other	(58%)	
urogenital problems	81	**(25%)**
-haematuria	(17%)	
-urine retention	(17%)	
-other	(65%)	
malaise	63	**(20%)**
-anaemia /blood loss	(41%)	
-other	(59%)	
neurological	27	**(8%)**
-decrease of sensibility/power	(78%)	
-other	(12%)	
other/unknown	9/5	**(4%)**
Total	320	**(100%)**

Table 8.2 presents the symptoms related to advanced prostate cancer. The most frequently registered symptoms were pain (42%) and urogenital symptoms (25%). Pain was most frequently treated with radiation therapy and pain medication. For urological symptoms, catheters were most often applied, while anaemia associated with malaise was mostly treated by blood transfusions. Neurological symptoms were generally relieved by radiation or laminectomy and pain medication.

Table 8.3
Assessment procedures and costs in all advanced disease patients (N = 70)

	Average per patient	% of patients with at least one diagnostic tool	Costs per diagnostic tool ($)[2]	Proportion of total diagnostic costs
Outpatient visits	16[1]	97%	22	22%
DRE	3.3	80%	0[3]	
TRUS	0.4	33%	45	1%
Prostate biopsy	0.4	36%	48	1%
Pathology	0.2	17%	48	1%
PSA	8.2	100%	17	9%
Other lab	6.6	93%	10[4]	4%
CT-scan	1.0	50%	181	12%
MRI	0.4	23%	580	13%
Bone scans	1.9	91%	167	21%
X-ray				
- thorax	1.3	57%	22	2%
- bone	2.0	71%	22	3%
IVP (intra venous pyelogram)	0.2	9%	65[4]	1%
Ultrasonography				
- abdomen	0.8	57%	41	2%
- kidney	0.4	20%	41	1%
- prostate	0.3	13%	41	1%
Other	1.8	63%	49[5]	6%

[1] including 1 visit of another specialist during hospital stay
[2] based on cost calculations, unless stated otherwise
[3] cost of DRE are included in cost of outpatient visit
[4] based on tariffs (COTG 1997)
[5] weighted average

The assessment procedures during the course of advanced disease and as a result of symptoms/complaints are presented in Table 8.3. DRE and PSA were most frequently used to monitor the course of advanced disease. Every patient had on average 2 bone-scans, one to establish the presence of bone metastases and an additional one to localise progression for treatment with palliative radiation therapy.

Other imaging techniques were also used to localise the (bone) metastases. To diagnose local obstruction or local extension of the cancer, intra venous pyelography (IVP) and ultrasonography were applied. On average 15 outpatient visits were recorded over a period of 24 months and 1 consultation with another specialist during hospital stay.

Table 8.4
Treatment and cost of treatment for all advanced disease patients (N = 70)

	Total number (%)		% of patients with at least one such treatment	Average number of hospital days (per patient)	Costs per treatment (hosp days incl.)[1] ($)
Hormonal (total)	110	(24%)	89%	3.4 (15%)	
- orchiedectomy	(27%)				1696
- LHRH[2]	(40%)				2464
- other (non-LHRH or combinations)[2]	(33%)				2047
Radiation therapy (total)	84	(19%)	47%	4.2 (18%)	
- external radiation[3]	(89%)				1769
- metastron	(11%)				3106
Palliative TURP	11	(2%)	14%	1.2 (5%)	2275
Chemotherapy	5	(1%)	6%	0.2 (1%)	1025
Pain medication (total)[4]	92	(20%)	69%	3.2 (14%)	622
- non-opioids	(57%)				
- opioids	(43%)				
Other treatment (total)	161	(34%)	80%	10.3 (45%)	
- CAD/PCN[5]	(32%)				815
- blood transfusions	(27%)				795
- other	(41%)				1736
TOTAL	463	(100%)		22.6 (100%)	1457

[1] costs of hospital days (all out) $ 242
[2] duration based on date of diagnosis until date of hormone refractory disease
[3] average number of sessions (external radiation) 7.4
[4] assumed duration 1 month
[5] CAD = catheter a demeure / PCN = percutaneous nephrostomy

Forty-three percent of the costs of assessment consisted of the costs of outpatient visits (22%) and the costs of bonescans (21%). PSA was performed 8 times on average per patient (every 3 months) and amounted to 9% of the costs of assessment.

Table 8.4 shows the total number of treatments for advanced disease (6.6 on average per patient). Hormonal treatment, radiation therapy and pain medication were most frequently applied. The category 'other treatment' consisted of, in addition to the use of catheters and blood transfusions, heterogeneous types of treatment. Hospitalisation for fractures, treatment by abstinence (before death), further (neurological) analysis and laminectomy, represented almost two-third of the hospital days in 14 patients. Of all hospital days, 45% were related to category 'other treatment', followed by hospital days related to radiation (18%). 14% of all costs of hospital days were attributed to the pain treatment with medicine.

The total costs (table 8.5) of advanced disease were $11,182 and consisted for the mostly of hospital days, followed by cost of treatment (49% and 37% respectively). The cost of outpatient visits and the cost of assessment accounted for only a small part of the total cost.

Table 8.5
Average total costs of advanced disease of prostate cancer per patient

	Cost ($)		%	
outpatient visits	348		3%	
assessment	1199		11%	
treatment	4173		37%	
- hormonal (chemical)		*2339*		*56%*
- radiation therapy		*1306*		*31%*
- surgery (orchidectomy, turp)		*225*		*5%*
- other		*301*		*7%*
hospital days	5462		49%	
TOTAL	11182		100%	

DISCUSSION

The average total costs per patient in the advanced disease stage of prostate cancer is $11,182 of which $1,547 (14%) relates to assessment (including outpatient visits) and $9,635 (86%) to treatment (including hospital days). If screening proves to be effective and in part prevents advanced disease, then our cost estimates can be used to calculate savings proportionately to the number of men prevented from dying of prostate cancer. This reduction in costs of advanced disease will partly balance costs of screening, diagnosis and early treatment.

We had to enrol a representative group of patients with distant metastases of prostate cancer. Because an academic hospital might select more serious patients we analysed the baseline characteristics for the hospitals separately (data not shown). The patient characteristics did not show statistically significant differences between both hospitals (although the numbers are small). Furthermore, in the academic hospital patients with a first diagnosis of advanced disease (M_1-patients) were not overrepresented. Exclusion of patients that were referred for a second opinion has also enhanced comparability by preventing overrepresentation of the more serious cases and problems of loss to follow-up. The survival of patients in both hospitals was also not statistically different. We therefore conclude that no serious selection effects were observed in this study.

The stopping of medication (e.g. hormonal, pain) was incompletely registered in the medical files, so we had to make assumptions about duration of these episodes. Because the average period of hormonal resistant disease was known (6 months), the duration of hormonal treatment could be estimated. The costs of other medication contributed relatively little, so the uncertainty of these estimates will have little effect on the reliability of the total cost estimates.

Information bias could have occurred as a result of incomplete registration of all care received by the patients, resulting in an underestimate of the care and thus the costs of advanced disease. It is unknown to what extent this was the case, but the files were documented in detail in both hospitals. With the approach used in this study it was not possible to accurately register nursing home care and home care. In other studies concerning advanced disease of breast and cervical cancer, the costs of nursing home care covered 7-8% of the total costs (de Koning *et al.*, 1992; van Ballegooijen *et al.*, 1992).

The burden of disease of prostate cancer was also investigated in the Nordic countries (Aus *et al.*, 1995; Borre *et al.*, 1997; Otnes *et al.*, 1995). However in those countries, prostate cancer was not curatively treated and no division was made between M_0M_1- and M_1-patients. We compared some outcomes of our patients

presenting with distant metastases at first diagnosis with these studies. The median crude survival in the studies by Aus and Otnes were 19 and 23 months, which was slightly less than our survival of M_1-patients (26 months). Otnes et al. found that the period from progression until death was 11 months, while the period was 6 months in our study. Compared to Aus et al. and Otnes et al., we found less palliative TURP (25% and 30% versus 14%) and more palliative radiotherapy (31% and 16% (all patients) versus 47%). This difference in the number of TURPs might also explain why both studies found a higher mean hospital stay (35 days and 1 month, respectively). Causes that might contribute to these differences are improved palliative care during recent years and differences between countries. Borre et al studied the human and economic burden of prostate cancer, but this was carried out irrespective of treatment strategy and thus stage of disease, which makes it impossible to compare the results (Borre *et al.*, 1997).

The survival of M_0M_1-patients was somewhat worse than of M_1-patients during the first 30 months of follow-up. This might be explained by the higher percentage of distant metastases in other locations than the bone in these patients. Metastases in other locations were associated with a lower survival.

Taplin also studied the health care costs of prostate cancer (Taplin *et al.*, 1995). Although a different approach was used, the net cancer costs for distant prostate cancer were higher than in our study (about $17,776), but not totally different. This difference could in part be caused by the relatively expensive health care in the U.S. compared to the Netherlands.

Advanced disease has been also evaluated for screening programs of breast cancer and cervical cancer. For breast cancer screening, the largest saving of both costs and quality of life was made through the prevention of advanced disease. Approximately 40% of the extra costs induced by a screening program for women aged 50-70 with an interval of 2 years are compensated for by the prevention of advanced disease (de Koning *et al.*, 1992). For cervical cancer, the saving in regard to advanced disease was approximately 10% (van Ballegooijen *et al.*, 1992). Studying the care and cost of breast cancer revealed that almost all women with advanced breast cancer die of their disease. For prostate cancer we found that the percentage was 75%.

The absolute costs of advanced disease of prostate cancer were lower than the costs of both advanced breast cancer and cervical cancer, while the duration of advanced disease of prostate cancer was longer than of both breast and cervical cancer (average duration 21 and 11 months, respectively). The costs were fl 34,200 and 28,200, respectively. The average number of hospital days (45) for breast cancer was higher than for cervical cancer (27 days) which was similar to prostate cancer (23 days).

The difference in costs between advanced breast cancer and prostate cancer was mainly caused by the difference in numbers of hospital days.

Future developments in assessment and treatment of prostate cancer might change the conclusions of this study. The widespread use and increasing knowledge about PSA, might result in a decrease of the number of bonescans and MRI for the assessment of distant metastases in the future. New therapy trials, such as immediate versus delayed treatment and continuous versus intermittent therapy are ongoing to improve the treatment of advanced disease. Especially new therapeutic options for patients with hormone refractory advanced disease have potency. The improvement in drug treatment will merely consist of developing (combinations of) drugs with less side effects, which in themselves could help to improve the quality of life in this last phase of the disease. These new developments are likely to involve extra costs, resulting in higher costs of the treatment of advanced disease.

This study described the symptoms, complications and accompanying treatment and costs from advanced prostate cancer. Combined with additional information about quality of life, these data give a better understanding of advanced disease and the possible health gains and savings due to early detection and effective treatment.

References

Anonymous, Maximum androgen blockade in advanced prostate cancer: an overview of 22 randomised trials with 3283 deaths in 5710 patients. Prostate Cancer Trialists' Collaborative Group [see comments]. *Lancet*, 346, 265-9 (1995).

Aus, G., Hugosson, J. and Norlen, L., Need for hospital care and palliative treatment for prostate cancer treated with noncurative intent. *J Urol*, 154, 466-9 (1995).

Auvinen, A., Rietbergen, J.B., Denis, L.J., Schroder, F.H. and Prorok, P.C., Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. *J Med Screen*, 3, 97-104 (1996).

Borre, M., Nerstrom, B. and Overgaard, J., The dilemma of prostate cancer–a growing human and economic burden irrespective of treatment strategies. *Acta Oncol*, 36, 681-7 (1997).

de Koning, H.J., van Ineveld, B.M., de Haes, J.C., van Oortmarssen, G.J., Klijn, J.G. and van der Maas, P.J., Advanced breast cancer and its prevention by screening. *Br J Cancer*, 65, 950-5 (1992).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., Boer, R., Collette, H.J.A., Verbeek, A.L.M., Hendriks, J.H.C.L., de Bruyn, A.E., van Agt, H.M.E., de Haes, J.C.J.M., van der Zwan, L., Koopmanschap, M.A., Habbema, J.D.F. and van der Maas, P.J., The costs and effects of mass screening for breast cancer (in Dutch), Department of Public Heath, Erasmus University Rotterdam, Rotterdam (1990).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

Finkler, S.A., The distinction between cost and charges. *Ann Intern Med*, 96, 102-9 (1982).

Gohagan, J.K., Prorok, P.C., Kramer, B.S. and Cornett, J.E., Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute [see comments]. *J Urol*, 152, 1905-9 (1994).

Habbema, J.D., van Oortmarssen, G.J., Lubbe, J.T. and van der Maas, P.J., The MISCAN simulation program for the evaluation of screening for disease. *Comput Methods Programs Biomed*, 20, 79-93 (1985).

Krenning, E.P., Lowenberg, B., Lugtenburg, P.J., Lamberts, S.W.J., van Putten, W.L.J., Habbema, J.D.F., Polder, J.J. and Eijkemans, M.J.C., Diagnostic value of Somatostatine Receptor Scintigraphy for staging malign lymphomas (in Dutch), Academic Hospital Rotterdam and Erasmus University Rotterdam, Rotterdam (1998).

Otnes, B., Harvei, S. and Fossa, S.D., The burden of prostate cancer from diagnosis until death. *Br J Urol*, 76, 587-94 (1995).

Schröder, F.H., Denis, L.J., Kirkels, W., de Koning, H.J. and Standaert, B., European randomized study of screening for prostate cancer. Progress report of Antwerp and Rotterdam pilot studies. *Cancer*, 76, 129-34 (1995).

Taplin, S.H., Barlow, W., Urban, N., Mandelson, M.T., Timlin, D.J., Ichikawa, L. and Nefcy, P., Stage, age, comorbidity, and direct costs of colon, prostate, and breast cancer care [see comments]. *J Natl Cancer Inst*, 87, 417-26 (1995).

van Ballegooijen, M., Koopmanschap, M.A., Tjokrowardojo, A.J. and van Oortmarssen, G.J., Care and costs for advanced cervical cancer. *Eur J Cancer*, 28A, 1703-8 (1992).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

9

GENERAL DISCUSSION

INTRODUCTION

Screening is evaluated in different ways throughout the course of its development. The evaluation starts with the preparation of a randomised screening trial and continues after the implementation of a screening programme, ending only when a screening programme is finally abolished (table 9.1). During the course of preparing a trial and the process of deciding to introduce screening, many evaluation issues will emerge for consideration. Next to favourable effects, screening often also has important unfavourable effects. The latter in particular justify extensive and continuous evaluation.

A nation-wide screening programme has already been introduced in the Netherlands for breast cancer, which requires a different evaluation perspective than that for prostate cancer, for which a randomised screening trial was started in 1994. In this thesis, aspects of the evaluation of screening programmes and of trials were brought together. This discussion offers an overview of the results and discusses the future challenges for the evaluation of screening of both cancers. The discussion of these chal-

Table 9.1
Evaluation issues in different phases of programme development

EVALUATION OF SCREENING
1st phase: Trial
• Study design (internal validity)
• Screening effects (mortality reduction)
• Test-characteristics screen test
• Quantify factors influencing the main outcome
2nd phase: Implementation of a programme
• Demography
• Epidemiology (incidence, stage distribution, mortality, survival)
• External validity
• Test-characteristics screening programme
• (Changes in) treatment and assessment
• Cost and savings
• Quality of life
• Screening effects (improvement in prognosis)
3rd phase: Continuation of a programme
• Early outcomes (detection rates, stage distribution etc.) and late outcomes (mortality)
• Changes as a result of screening
• New scientific evidence: epidemiology, behavioural factors, therapy etc.

lenges highlights the different perspectives from which screening programmes may be evaluated, i.e. as a health care service or within the context of a trial.

SUMMARY OF THE RESULTS

First research question: cost effectiveness in other countries and settings

The first research purpose of this thesis was to identify and quantify factors that influence the cost-effectiveness of breast cancer screening programmes in different countries and thus health care settings (Part I). These results may also be of use for the evaluation of the cost-effectiveness of other cancer screening programmes.

This research question was addressed by two studies on breast cancer screening, in Germany and Spain. Both studies showed that the incidence and mortality of breast cancer were important determinants of cost-effectiveness. The incidence (I) and mortality (M) (European Standardised Rate) in Spain were 29% (I) and 38% (M) lower than in the Netherlands, while in Germany the rates were 16% (I) and 24% (M) lower. This negatively influenced the CE-ratio for both countries. The age distribution of the incidence and mortality is also important as was illustrated in Spain. Cost-effectiveness of screening in the age-group 40-50 was more favourable, due to a relatively high incidence in this age group compared to other European countries.

The effect of lower sensitivity and specificity was studied in Germany, resulting in higher CE-ratios. This was further elaborated in a later study by Warmerdam et al, showing that the costs of extra quality assurance can be counterbalanced by the savings obtained through improved sensitivity and specificity (Warmerdam *et al.*, 1997).

Attendance appeared to influence costs and effects more or less equally, resulting in almost no effect on the CE-ratio. At attendance rates ranging between 50% and 80%, the overhead costs of organisation and evaluation have little influence on the costs per life-year gained. With lower attendance rates, however, these overhead costs will markedly increase CE-ratios.

In the German study, the effect of health care setting was clearly demonstrated. Were the organisation of the screening programme to be decentralised, many mammographs would not be used at maximum capacity. While this may negatively effect cost-effectiveness, it is more importantly also very likely to have a negative effect on the quality of mammography. A decentralised approach would lead to less experience in reading mammograms by medical professionals. Mammographs would tend to be less often replaced by newer equipment, which is likely to negatively influence sensitivity

and specificity. A decentralised organisation thus has many disadvantages, with possibly as only benefit a more personal approach, as for some women mammography is done by their own physician.

Both in Germany and Spain, opportunistic screening, i.e. mammography performed for preventive purposes in the absence of an organised screening programme, occurs on a considerable scale (personal communication Robra, 1993 and Borras, 1995). Using the MIcro simulation SCreening ANalysis (MISCAN) model, the situations with screening and without screening were compared. An organised screening programme is expected to replace to some extent part of the opportunistic screening performed. A major problem is, however, that the effectiveness of screening outside screening programmes (opportunistic screening) is hard to quantify and therefore difficult to deal with in a cost-effectiveness analysis. We did not take opportunistic screening into account, but the savings in cost by the introduction of a screening programme, especially achievable where opportunistic screening is widespread could be considerable.

Conclusion

Costs and effects are determined by many factors that can differ between countries. The relative contribution of each factor separately is difficult to quantify, because many factors are mutually correlated. All factors should therefore be considered together. Despite these limitations, we have shown that standardised cost-effectiveness analyses for different countries are possible, and that these may result in very different cost-effectiveness ratios between countries.

Second research question: secondary effects of screening programmes

The second research question concerns the quantification of secondary effects of screening after the introduction of a nation-wide breast cancer screening programme (Part II). This research question was studied by quantifying the effects of radiation risk on breast cancer mortality and by quantifying screening outside the target population (opportunistic screening).

Radiation risk

The effects of radiation risk are quite difficult to assess because the dose used in breast cancer screening is only about 2 mGy per mammogram, whereas available risk estimates are based on doses that are 100-300 times higher. Extrapolation of the risk estimates to very low doses is thus needed to estimate the number of breast cancer deaths and the balance between deaths induced and prevented for different screening policies.

This balance was favourable in the Dutch screening programme (age-group 50-69, interval of 2 years), but this balance becomes less favourable in younger women for at least two reasons. The first is an increased radiation risk at younger ages, when the breast tissue is more sensitive to radiation. Secondly, the breast cancer incidence and mortality of women in their forties is lower, thus screening may prevent fewer breast cancer deaths. Although screening in women under 50 was assumed to be ineffective for a long time, recently published trial results suggest a smaller than for the age group 50-69, but statistically significant mortality reduction (Hendrick *et al.*, 1997; Larsson *et al.*, 1997). Assuming the same effectiveness of screening in women aged 40-49 as in women aged 50-69, the ratio of cases induced/prevented deaths remained favourable. Because of the many uncertainties in the risk estimates, the confidence limits of these estimates were wide. Recently, lower risk estimates than those we used were published, resulting in 3-6 times lower estimates for induced breast cancer. If screening was extended to include women in their forties, radiation risk should be considered, but its importance is mainly determined by the extent of effectiveness of screening in this age group.

Mammograms outside screening

The effect of the introduction of nation-wide screening on the number of mammograms requested by general practitioners for women outside the target ages (50-69) was studied. A steady increase in requests for mammograms was revealed for women older than 40 during the period 1988-1995. Taking into account this overall increase, in the target population a decrease in mammograms performed outside the scope of screening programmes was observed. This decrease had been expected at the time the cost-effectiveness analysis for breast cancer screening was done (de Koning *et al.*, 1991). It was assumed that the number of clinical mammograms and the breast cancer diagnostics would decrease proportionally with the number of cancers detected by screening instead of clinically.

Conclusion

During the implementation of a screening programme, the monitoring of effects and costs remains essential, as developments in health care and the epidemiology of the disease can influence the effects and costs of the programme. The factors studied in this thesis did not put the nation-wide screening programme in a different perspective, because both outcomes of the evaluation of radiation risk estimates and of 'opportunistic screening' were favourable. The section on the evaluation challenges (section 9.3) provides more examples of the need for a continuous evaluation of the programme.

Third research question: determinants of prostate cancer screening effectiveness

The third research question concerns the quantification of factors influencing the performance of the present prostate cancer screening trial and thus a possible future screening programme (Part III). This research question was addressed in 3 studies.

The first study was intended to optimise the screening protocol used in the European Randomised study on Screening of Prostate Cancer (ERSPC, Rotterdam). This was prompted by the objection from both researchers and the Health Council of the Netherlands to the high proportion of false positive biopsy indications (Health Council of the Netherlands, 1996). We looked at several combinations of screening tests (prostate specific antigen (PSA), digital rectal examination (DRE) and transrectal ultrasonography (TRUS) and estimated the resulting number of false negatives in each combination in an attempt, to reduce the number of false positive biopsy indications. By weighing screening and assessment procedures and the cancers detected, we concluded that screening with PSA alone would be the best option. Although about 8% of the cancers would consequently be missed, the advantages would be a reduction of the false positive biopsy indications and a very simple screening procedure. Since May 1997, screening has consisted of PSA only, followed by a biopsy if the PSA is ≥ 3ng/ml.

In the second study, PSA-testing before and during the ERSPC and the base-line use of assessment procedures such as DRE and biopsy for prostate disease was quantified for Rotterdam. PSA-testing is very simple and its use should be monitored, especially in the control arm of the trial, as it could influence the outcome of the trial. In the screen arm of the trial, 3.3% underwent a PSA annually outside the screening trial, while this was 7.6% in the control arm. An average of 45% of the men had had a DRE at some point, and 13% reported having a PSA test before participating in the trial. Undergoing PSA-testing and DRE before participation in the trial did not result in statistically significant differences in detection rates.

The 4-year screening interval yields a contamination level that must also be taken into account in the final evaluation. The results moreover underscore the importance of a continuous follow-up of further monitoring of PSA use in all trial participants.

In the third study, we studied the course of advanced prostate cancer and its health care costs. The most frequently reported symptoms were pain (42%), urogenital symptoms (25%) and malaise (20%). 89% of all patients were hormonally treated (either by orchidectomy or chemical castration or a combination) and 47% received one or more courses of radiation therapy. 69% of all patients were on pain medication.

The average costs of advanced disease were estimated at $11,182 over the total period of advanced disease of 24 months. 14% was allocated to assessment and outpatient care, 37% to treatment costs and 49% to costs of hospital stay. If screening proves to be effective in preventing advanced disease in some participants, this will result in a sizeable reduction of treatment costs for advanced disease. These estimates, together with the effects of advanced prostate cancer on quality of life, will be used for the evaluation of prostate cancer screening.

Conclusion

The evaluation of an ongoing screening trial requires continuous attention to directly related questions that can not be answered in the main screening trial. This is extremely important as the answers to these questions may determine the final judgement about the possible mortality reduction and the desirability of introducing screening.

EVALUATION CHALLENGES FOR BREAST CANCER

Breast cancer screening, a national health care service that has reached the 3rd phase (table 1), is facing two main evaluation challenges. The first challenge is to maintain an optimal breast cancer screening programme in the light of new scientific evidence on e.g. treatment, techniques, screening effectiveness and policy developments.

The second challenge regards the evaluation of the mortality reduction as a result of breast cancer screening. This issue is all the more interesting as standard methods used in evaluation trials can not be easily applied. A 'control arm' is not available for breast cancer screening as a health care service. A few aspects of these challenges are presented in this paragraph.

Maintenance of an optimal screening programme

Screening of women aged 40-49 and over the age of 70

Breast cancer screening for women aged 40-49 is still a controversial topic. Many questions concerning the screening of women in their forties could not be resolved, because the studies were not especially designed to produce age-specific results. Nevertheless, a recent meta-analysis of 8 randomised trials showed a statistically significant mortality reduction of 18% for women aged 40-49 at entry (Hendrick *et al.*, 1997). When all the data on women aged 40-49 at entry of 5 Swedish randomised trials were combined, a statistically significantly 29% mortality reduction was found among women invited for

screening (Hendrick *et al.*, 1997). Part of this mortality reduction could, however, be due to screen detection of cancers in these women above the age of 50, as was suggested by de Koning et al. (de Koning *et al.*, 1995). The magnitude of this effect was studied in four Swedish randomised trials, which concluded that almost the entire effect in the group aged 40-44 years at randomisation was due to screening before the age of 50 (Larsson *et al.*, 1997). Now that screening women in their forties may be as effective as for women in the age-group 50-69, the questions arises if screening should become available as a health care service for this group. The cost-effectiveness of screening women in their forties is less favourable, mainly because breast cancer incidence and mortality are much lower in this age-group in most countries. One study concluded that the cost-effectiveness of screening women in the age-group of 40-49 years was almost 5 times as high as in older women (age 50-69) (Salzmann *et al.*, 1997).

A new trial has already been set up to quantify the effectiveness of screening women aged 40-49 years (Moss, submitted). The first mortality analyses are expected around 2010-2012. After the effectiveness (if present) has been proved, additionally cost-effectiveness analyses will support policy decisions about the introduction of screening women in their forties.

In 1998, the upper age limit of the Dutch screening programme was moved from 69 to 74, resulting in 3 more invitations per woman. Political pressure has been brought to bear on the upper age limit, with age discrimination as an argument. The effectiveness of screening women older than 75, however, has never been addressed in randomised trials (Nystrom *et al.*, 1993). Case-control studies have indicated that screening until the age of 75 could reduce the mortality of breast cancer (van Dijck *et al.*, 1997; van Dijck *et al.*, 1996). The number of life-years gained per women prevented from dying of breast cancer decreases with increasing age. Due to the (assumed) longer preclinical period, more years with lead-time are generated by screening until an older age. At older ages, screening will thus result more often in diagnosis and treatment of breast cancer in women that would have died from another cause, before they themselves would have noticed a breast abnormality. The combination of these factors results in a less favourable balance between positive and negative effects of screening at older ages. Screening until the age of 75 may be as cost-effective as for the age group 50-69, but a further increase would show a markedly less favourable balance (Boer *et al.*, 1995). While a trial might be the best method to quantify the effect of screening women aged 75 years and older, very large numbers would be required in order to finally establish an optimal upper age limit.

Interval cancers

Evaluating the detection rates together with the interval cancer rates in the 2-year period after the screening, and the stage distribution of screen-detected and interval cancers (incidence rate of advanced cases) is necessary to judge the performance of the programme. The first results of the Dutch nation-wide screening programme have recently been published (Fracheboud *et al.*, 1998). Although the interval cancer rate was as high as estimated in advance (with the use of the MISCAN-model), these results also revealed a somewhat worse performance in comparison with screening in two regions of the UK. Despite a relatively high breast cancer incidence the Dutch screening programme showed a lower detection rate of invasive cancers, followed by somewhat higher interval cancer rates. Furthermore, the stage distribution of the screen-detected cases was less favourable than in some regions of the UK (Day *et al.*, 1995; Woodman *et al.*, 1995). The interval cancer rates were also almost twice as high as those reported from one Swedish trial (Tabar *et al.*, 1987). These results were one of the reasons for starting research on the possible improvement of sensitivity in the national screening programme (see under Optimisation).

New screening and assessment techniques

One of the most important new developments in mammographic screening is perhaps the use of Computed Assisted Diagnosis (CAD). With the use of CAD the computer sets a prompt on a digitised mammogram, preventing the radiologist from overlooking an abnormality. Several studies have shown that this method could improve the sensitivity of mammography reading (te Brake *et al.*, 1998; Thurfjell *et al.*, 1998). However, the CAD in itself had a very low specificity, which may need improvement before becoming useful in mammographic screening (Thurfjell *et al.*, 1998). Furthermore, issues such as developing optimal detection schemes for the computer, assessment of the costs involved, possibility of training etc. need to be addressed, before this technique can be applied in screening

As a result of screening, the number of non-palpable breast abnormalities has increased. The assessment of these lesions is time consuming and costly due to a complicated procedure (stereotactic surgical biopsy) demanding total anaesthesia and hospital stay of 1 or 2 days. A new study was started in the Netherlands to rate the value of stereotactic core needle biopsy (Anonymous, 1997). This new technique has the advantage that no hospital admission is needed, there is less mutilation and scar tissue and the interpretation of future mammograms (in the case no cancer is present) is less hampered. Moreover, this technique may prove less expensive, because the costs of a hospital stay are saved. If this new technique proves to have the same sensitivity it could

replace the existing procedure, reducing the burden for women diagnosed with a non-palpable breast lesion.

High risk groups

Hereditary risk of breast cancer has attracted more attention with the discovery of the BRCA1 and BRCA2 genes. Presence of the BRCA1 or BRCA2 mutations are supposed to give about 70% risk of developing breast cancer before the age of 70 on the basis of population based studies (Claus *et al.*, 1991; Whittemore *et al.*, 1997). Presence of this gene results in a psychological load and has already resulted in decisions for preventive amputation of both breasts (DudokdeWit *et al.*, 1997). In the Netherlands, genetic counselling is available for women who request this service. Women attending the counselling vary in breast cancer risk and the guidelines are not yet clear. An alternative to preventive amputation might be (frequent) screening from young ages onwards. In particular, development of a proper guideline for young women bearing a hereditary burden is a challenge.

Women with dense breast tissue seem to be at higher risk of developing breast cancer (van Gils, 1998). If these findings are confirmed in other studies, this might result in a separate screening programme (with other intervals or more views) for some women.

Chemoprevention

The first results of prevention of breast cancer with tamoxifen have recently become available (Fisher *et al.*, 1998). Although the incidence of breast cancer was significantly reduced by 49% in women aged 35 and over, many questions still have to be resolved. The increased incidence of cervical cancer, stroke, pulmonary embolism and deep vein thrombosis must be weighted against the reduction in the incidence of breast cancer. Furthermore, the incidence of breast cancer might only be postponed, which, among other long-term effects, will emerge with longer follow-up. Because the effects were also strong at younger ages (44% reduction in incidence for women aged 49 or younger), the developments in this field must be continuously monitored, especially if an expansion of screening towards younger ages is considered. These developments have to be followed carefully and may have a profound consequence for the present screening programme in the Netherlands.

Optimisation

Now that the first data about interval cancers and repeat screening examinations have become available, the performance of the nation-wide screening programme in the

Netherlands can be better assessed (Fracheboud *et al.*, 1998). The data suggest that the screening performance might be improved.

The Netherlands Evaluation Team of Breast cancer screening (NETB) and the National Expert and Training Centre for Breast Cancer Screening (LRCB) will start a study on a possible improvement of sensitivity in the national screening programme. Research questions concern the practice of independent double reading, two-view screening in subsequent screens for all women, minimal signs predictive for cancer and the effect of higher referral rates. Many studies have already demonstrated the improved detection by 5-15% with independent double reading (Anderson *et al.*, 1994; Anttinen *et al.*, 1993; Brown *et al.*, 1996; Ciatto *et al.*, 1995; Thurfjell, 1995). This increased detection was counterbalanced by an increased referral in most studies, resulting in a lower specificity. The extra detection has even been balanced against the incurred costs, concluding that a consensus double reading was more cost-effective than single reading policy (Brown *et al.*, 1996). Extrapolation of these results should however be done with caution, because the referral rates are much lower in the Dutch screening programme than in other programmes (Chamberlain *et al.*, 1993; Hakama *et al.*, 1995; Thurfjell, 1995).

The effect of two-view screening has been studied in the UK, because that programme initially had one-view screening. Wald et al showed a 24% increase in detection rates with two-view screening as compared with one-view (Wald *et al.*, 1995). The extra costs were balanced by the extra cancers detected, resulting in similar cost-effectiveness. Another study also showed an improved detection of small invasive cancers (< 15 mm) from 45% for first screens and 25% for subsequent screens by two-view mammography compared to one-view (Blanks *et al.*, 1997). In the Dutch situation, two-view mammography is carried out during first screens and the results of a previous screening round are used for the interpretation of the following mammogram. This different approach makes extrapolation of these results difficult and justifies additional research.

Minimal signs on screening mammograms have been often studied, especially in relation to review of previous mammograms (Ciatto *et al.*, 1995; Daly *et al.*, 1998; van Dijck *et al.*, 1993). The interpretation of these signs prospectively in relation to breast cancer detection is, however, unknown. Improved knowledge of the predictive value of certain types of minimal signs might improve the sensitivity of breast cancer screening by mammography in the future.

Assessing cancer mortality reduction

One of the most challenging topics in the future evaluation will be to assess the mortality reduction of national breast cancer screening. A first model-based analysis, comparing expected mortality due to the introduction of screening with the one observed, was done for the Netherlands and the UK for the period 1989-1996. This study showed that the effects of screening in the Netherlands are expected to be demonstrable from 1998 onwards (van den Akker-van Marle *et al.*, in press).

The effectiveness of the Finnish national screening programme could be evaluated because of the gradual invitation of different cohorts of women (Hakama *et al.*, 1997). They concluded that a breast screening programme can achieve a similar effect on mortality as achieved by the trials for breast cancer screening. This methodology may be used in The Netherlands, where the screening programme was gradually introduced by region. The number of screened women might, however, be not sufficient. The assessment of mortality reduction due to screening is a major methodological challenge. Trend analysis taking into account other developments alongside the implementation will be a method of analysis. Yet, inevitably, a comparison will have to be made between women having attended the screening and those having not attended. This may be done in a case-control study, which however always entails some problems with the interpretation of the results (Cronin *et al.*, 1998; Demissie *et al.*, 1998).

EVALUATION CHALLENGES FOR PROSTATE CANCER

The major challenge for prostate cancer screening evaluation involves establishing the effectiveness of screening in reducing deaths from prostate cancer, in the context of e.g. changes in therapy, increase in screening outside the screening trial, and perhaps the development of more specific screening tests. In the mean time, also because of the adverse side effects of screening, it is ethically unacceptable to offer a-symptomatic men screening for prostate cancer outside experimental randomised settings (de Koning and Schroder, 1998). After the value of prostate cancer screening has been established, the evaluation will be directed towards a decision about screening in the general population. In the next paragraphs these challenges are further discussed.

Establishing of the effectiveness of prostate cancer screening

Analysis of cancer mortality reduction

The largest challenge in the evaluation of prostate cancer screening is a reliable estimate of its effectiveness in reducing prostate cancer mortality, based on well-designed studies. The ERSPC and PLCO will give a large contribution by enrolling 200.000-250.000 men in randomised trials (Auvinen *et al.*, 1996; Gohagan *et al.*, 1994). Screening will in general result in a more favourable stage distribution of detected disease, which often entails a different distribution of treatment modalities applied. A difference in distribution in treatment modalities within a stage of prostate cancer between the screening and control arm of the study might hamper the interpretation of the results of the trial. There is to date no convincing evidence, based on a randomised trial, to show whether radiation therapy or prostatectomy should be preferred as the curative treatment modality for localised prostate cancer. If one of these should, in theory, be superior to the other and this treatment is more often applied to the screening arm than the control arm, a distortion of the results might occur.

Opportunistic 'screening', i.e. screening outside the trial in both screening and control arm, was reported in this thesis. The level of opportunistic screening in the control arm was about 8% per year and this phenomenon warrants further evaluation in the future. Registering both treatment and opportunistic screening for all persons in the study and control group, will enable adjustment for these factors to be made in the final mortality analysis.

Screening tools

The optimal (combination of) screening tools for the early detection of prostate cancer has not definitively been established. PSA test is very promising, but the test characteristics of this test are far from ideal (Rietbergen *et al.*, 1997). The sensitivity is likely to be quite high, but the specificity is moderate. This results in low positive predictive value of the biopsies done on the basis of some PSA cut-off values. Several methods have been suggested to reduce the number of false positive biopsy indications, like PSA density (PSA corrected for total prostate volume or transition zone volume) and free to total PSA ratio (Babaian *et al.*, 1992; Catalona *et al.*, 1995; Kalish *et al.*, 1994). Until now this has not resulted in a clear preference for one of these methods, because improvement in specificity resulted in a considerable loss in sensitivity (Bangma *et al.*, 1997; Rietbergen *et al.*, 1998). Within the different centres participating in ERSPC various cut-off levels of PSA and screening intervals are used, that will help to determine the optimal screening protocol. New developments, to improve the test-characteristics of PSA have to be taken into account to optimise the screening protocol.

Natural history

The natural history of prostate cancer is largely unknown. The screening trial will yield much knowledge, especially from the comparison of the screening and control arm. Differences in number of cancer diagnoses, stage distribution, grade and timing of the diagnosis will help to unravel the natural history. The incidence of interval cancers and timing since screening of these cancers will also contribute. The use of micro-simulation will help to estimate the mean preclinical-duration of prostate cancer (van Oortmarssen *et al.*, 1990).

Deciding about the introduction of prostate cancer screening

External validity

An important theme in future evaluation should be the external validity of the results obtained in the ERSPC, Rotterdam. The screening trial has been set up as an efficacy trial, because randomisation took place after informed consent. This will result in an estimate of the effect of screening of only 45% (attendance rate) of the target population. If people participating in the trial are selected according to e.g. disease prevalence or general health status, the result might not be valid for the general population, for which the screening is eventually intended (Nijs *et al.*, 1997). This issue has to be further addressed.

Cost-effectiveness analysis

In order to perform a thorough cost-effectiveness analysis, many issues remain to be quantified in the next few years. Some men will be diagnosed and treated for prostate cancer and die (of other disease) before the time they would have been diagnosed without screening. Those men have a diagnosis of prostate cancer without benefiting from screening. Due to the probable longer preclinical period and the higher age at which prostate cancer occurs than for breast cancer, this adverse effect will be more prominent than in breast cancer. Furthermore, with the present screening modality (PSA $\geq$ 3 ng/ml as a biopsy indication) a considerable percentage of men will still have a false positive biopsy indication. Therefore assessment of the quality of life effects of screening attendance, false positive test results and of prostate cancer during different stages of disease and stratified for different treatment modalities is indispensable. In Rotterdam quality of life studies are conducted alongside the screening trial (Essink-Bot *et al.*, 1998; Madalinska *et al.*, in preparation). The outcomes of these studies will help to weigh many aspects of prostate cancer.

In order to estimate the effects of population screening the present practices of diagnosis and treatment of prostate cancer should be documented. The costs of assessment, treatment, organisation and evaluation of screening must also be assessed. With the MISCAN model, stage-specific improvement of prognosis due to screening and therapy will have to be estimated. By combining all this information, the effects of different screening intervals and age groups can be analysed and will support the decision about introducing screening.

CONCLUDING REMARKS

Screening always has to strike a balance between favourable and adverse effects. This balance should always guide the implementation and monitoring a nation-wide screening programme. Changes in elements included in the cost-effectiveness analysis or new developments may lead to changes in the screening programme and in theory eventually the abolition of the screening programme. The results achieved in trials may differ from the results in a total population. An obvious reason for such a divergence may be differences in the quality of follow-up assessment and therapy. A profound evaluation has to include all these aspects, also because national cancer screening programmes are rather costly. The Dutch breast cancer screening programme costs, for example, 59 million guilders per year which is 0.1% of the total health care costs in the Netherlands.

The main question to be answered by the evaluation of a screening trial is the quantification of the cancer mortality reduction. However, many other factors must also be quantified to obtain a balanced judgement on the benefit/harm ratio.

It should be kept in mind that the benefit/harm ratio of screening for cancer is subject to continuous change, as a result of improvement in test properties and diagnostic and therapeutic improvement. This necessitates continuous alertness to adapt the programme in order to obtain optimal results. Obviously, the ideal will remain to find effective forms of primary prevention or therapy that will make screening superfluous.

References

Anderson, E.D., Muir, B.B., Walsh, J.S. and Kirkpatrick, A.E., The efficacy of double reading mammograms in breast screening [see comments]. *Clin Radiol*, 49, 248-51 (1994).

Anonymous, Core biopsy after radiologic localisation (Cobra), Academic Hospital Utrecht, Utrecht (1997).

Anttinen, I., Pamilo, M., Soiva, M. and Roiha, M., Double reading of mammography screening films--one radiologist or two? [see comments]. *Clin Radiol*, 48, 414-21 (1993).

Auvinen, A., Rietbergen, J.B., Denis, L.J., Schroder, F.H. and Prorok, P.C., Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group. *J Med Screen*, 3, 97-104 (1996).

Babaian, R.J., Miyashita, H., Evans, R.B. and Ramirez, E.I., The distribution of prostate specific antigen in men without clinical or pathological evidence of prostate cancer: relationship to gland volume and age. *J Urol*, 147, 837-40 (1992).

Bangma, C.H., Rietbergen, J.B., Kranse, R., Blijenberg, B.G., Petterson, K. and Schroder, F.H., The free-to-total prostate specific antigen ratio improves the specificity of prostate specific antigen in screening for prostate cancer in the general population [see comments]. *J Urol*, 157, 2191-6 (1997).

Blanks, R.G., Moss, S.M. and Wallis, M.G., Use of two view mammography compared with one view in the detection of small invasive cancers: further results from the National Health Service breast screening programme. *J Med Screen*, 4, 98-101 (1997).

Boer, R., de Koning, H.J., van Oortmarssen, G.J. and van der Maas, P.J., In search of the best upper age limit for breast cancer screening. *Eur J Cancer*, 31A, 2040-3 (1995).

Brown, J., Bryan, S. and Warren, R., Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms [see comments]. *Bmj*, 312, 809-12 (1996).

Catalona, W.J., Smith, D.S., Wolfert, R.L., Wang, T.J., Rittenhouse, H.G., Ratliff, T.L. and Nadler, R.B., Evaluation of percentage of free serum prostate-specific antigen to improve specificity of prostate cancer screening [see comments]. *Jama*, 274, 1214-20 (1995).

Chamberlain, J., Moss, S.M., Kirkpatrick, A.E., Michell, M. and Johns, L., National Health Service breast screening programme results for 1991-2 [published erratum appears in BMJ 1993 Aug 28;307(6903):543] [see comments]. *Bmj*, 307, 353-6 (1993).

Ciatto, S., Del Turco, M.R., Morrone, D., Catarzi, S., Ambrogetti, D., Cariddi, A. and Zappa, M., Independent double reading of screening mammograms. *J Med Screen*, 2, 99-101 (1995).

Ciatto, S., Rosselli Del Turco, M. and Zappa, M., The detectability of breast cancer by screening mammography. *Br J Cancer*, 71, 337-9 (1995).

Claus, E.B., Risch, N. and Thompson, W.D., Genetic analysis of breast cancer in the cancer and steroid hormone study. *Am J Hum Genet*, 48, 232-42 (1991).

Cronin, K.A., Weed, D.L., Connor, R.J. and Prorok, P.C., Case-control studies of cancer screening: theory and practice. *J Natl Cancer Inst*, 90, 498-504 (1998).

Daly, C.A., Apthorp, L. and Field, S., Second round cancers: how many were visible on the first round of the UK National Breast Screening Programme, three years earlier? *Clin Radiol*, 53, 25-8 (1998).

Day, N., McCann, J., Camilleri-Ferrante, C., Britton, P., Hurst, G., Cush, S. and Duffy, S., 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*, 2, 180-5 (1995).

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M.M. and van der Maas, P.J., Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst*, 87, 1217-23 (1995).

de Koning, H.J. and Schroder, F.H., PSA screening for prostate cancer: the current controversy. *Annals of Oncology*, 9, 1293-1296 (1998).

de Koning, H.J., van Ineveld, B.M., van Oortmarssen, G.J., de Haes, J.C., Collette, H.J., Hendriks, J.H. and van der Maas, P.J., Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer*, 49, 531-7 (1991).

Demissie, K., Mills, O.F. and Rhoads, G.G., Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J Clin Epidemiol*, 51, 81-91 (1998).

DudokdeWit, A.C., Tibben, A., Frets, P.G., Meijers-Heijboer, E.J., Devilee, P., Klijn, J.G., Oosterwijk, J.C. and Niermeijer, M.F., BRCA1 in the family: a case description of the psychological implications. *Am J Med Genet*, 71, 63-71 (1997).

Essink-Bot, M.L., de Koning, H.J., Nijs, H.G., Kirkels, W.J., van der Maas, P.J. and Schroder, F.H., Short-term effects of population-based screening for prostate cancer on health-related quality of life. *J Natl Cancer Inst*, 90, 925-31 (1998).

Fisher, B., Costantino, J.P., Wickerham, D.L., Redmond, C.K., Kavanah, M., Cronin, W.M., Vogel, V., Robidoux, A., Dimitrov, N., Atkins, J., Daly, M., Wieand, S., Tan-Chiu, E., Ford, L. and Wolmark, N., Tamoxifen for prevention of breast cancer: report of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *J Natl Cancer Inst*, 90, 1371-88 (1998).

Fracheboud, J., de Koning, H.J., Beemsterboer, P.M.M., Boer, R., Broeders, M.J.M., Hendriks, J.H.C.L., Verbeek, A.L.M., van Ineveld, B.M. and van der Maas, P.J., Interval cancers in the nationwide breast cancer screening programme in the Netherlands. *submitted* (1998).

Gohagan, J.K., Prorok, P.C., Kramer, B.S. and Cornett, J.E., Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute [see comments]. *J Urol*, 152, 1905-9 (1994).

Hakama, M., Holli, K., Isola, J., Kallioniemi, O.P., Karkkainen, A., Visakorpi, T., Pukkala, E., Saarenmaa, I., Geiger, U., Ikkala, J. and et al., Aggressiveness of screen-detected breast cancers [see comments]. *Lancet*, 345, 221-4 (1995).

Hakama, M., Pukkala, E., Heikkila, M. and Kallio, M., Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study [see comments]. *Bmj*, 314, 864-7 (1997).

Health Council of the Netherlands, Proeve toepassing Wet bevolkingsonderzoek: prostaatkanker (in Dutch), Gezondheidsraad, Den Haag (1996).

Hendrick, R.E., Smith, R.A., Rutledge, J.H., 3rd and Smart, C.R., Benefit of screening mammography in women aged 40-49: a new meta-analysis of randomized controlled trials. *Monogr Natl Cancer Inst*, 87-92 (1997).

Kalish, J., Cooner, W.H. and Graham, S.D., Jr., Serum PSA adjusted for volume of transition zone (PSAT) is more accurate than PSA adjusted for total gland volume (PSAD) in detecting adenocarcinoma of the prostate. *Urology*, 43, 601-6 (1994).

Larsson, L.G., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J., Tabar, L. and Nystrom, L., Updated overview of the Swedish Randomized Trials on Breast Cancer Screening with Mammography: age group 40-49 at randomization. *Monogr Natl Cancer Inst*, 57-61 (1997).

Madalinska, J.B., Essink-Bot, M.L., de Koning, H.J., Schroder, F.H. and van der Maas, P.J., Health related quality of life in patients with either screen-detected or clinically diagnosed prostate cancer before undergoing primary treatment. *(in preparation)* (1999).

Moss, S., (for the Trial Steering Group) A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40: description of method. *(submitted).*

Nijs, H.G., Tordoir, D.M., Schuurman, J.H., Kirkels, W.J. and Schroder, F.H., Randomised trial of prostate cancer screening in The Netherlands: assessment of acceptance and motives for attendance. *J Med Screen*, 4, 102-6 (1997).

Nystrom, L., Rutqvist, L.E., Wall, S., Lindgren, A., Lindqvist, M., Ryden, S., Andersson, I., Bjurstam, N., Fagerberg, G., Frisell, J. and et, a., Breast cancer screening with mammography: overview of Swedish randomised trials [see comments] [published erratum appears in Lancet 1993 Nov 27;342(8883):1372]. *Lancet*, 341, 973-8 (1993).

Rietbergen, J.B., Kranse, R., Hoedemaeker, R.F., Kruger, A.E., Bangma, C.H., Kirkels, W.J. and Schroder, F.H., Comparison of prostate-specific antigen corrected for total prostate volume and transition zone volume in a population-based screening study. *Urology*, 52, 237-46 (1998).

Rietbergen, J.B., Kranse, R., Kirkels, W.J., De Koning, H.J. and Schroder, F.H., Evaluation of prostate-specific antigen, digital rectal examination and transrectal ultrasonography in population-based screening for prostate cancer: improving the efficiency of early detection. *Br J Urol*, 79, 57-63 (1997).

Salzmann, P., Kerlikowske, K. and Phillips, K., Cost-effectiveness of extending screening mammography guidelines to include women 40 to 49 years of age [published erratum appears in Ann Intern Med 1998 May 15;128(10):878] [see comments]. *Ann Intern Med*, 127, 955-65 (1997).

Tabar, L., Faberberg, G., Day, N.E. and Holmberg, L., What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer*, 55, 547-51 (1987).

te Brake, G.M., Karssemeijer, N. and Hendriks, J.H., Automated detection of breast carcinomas not detected in a screening program. *Radiology*, 207, 465-71 (1998).

Thurfjell, E., Mammography screening methods and diagnostic results. *Acta Radiol Suppl*, 395, 1-22 (1995).

Thurfjell, E., Thurfjell, M.G., Egge, E. and Bjurstam, N., Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening. *Acta Radiol*, 39, 384-8 (1998).

van den Akker-van Marle, M.E., de Koning, H.J., Boer, R. and van der Maas, P.J., Reduction in breast cancer mortality due to the introduction of mass screening in the Netherlands; comparison with the UK. *J Med Scr (in press).*

van Dijck, J.A., Verbeek, A.L., Beex, L.V., Hendriks, J.H., Holland, R., Mravunac, M., Straatman, H. and Werre, J.M., Breast-cancer mortality in a non-randomized trial on mammographic screening in women over age 65. *Int J Cancer*, 70, 164-8 (1997).

van Dijck, J.A., Verbeek, A.L., Beex, L.V., Hendriks, J.H., Holland, R., Mravunac, M., Straatman, H. and Werre, J.M., Mammographic screening after the age of 65 years: evidence for a reduction in breast cancer mortality. *Int J Cancer*, 66, 727-31 (1996).

van Dijck, J.A., Verbeek, A.L., Hendriks, J.H. and Holland, R., The current detectability of breast cancer in a mammographic screening program. A review of the previous mammograms of interval and screen-detected cancers [see comments]. *Cancer*, 72, 1933-8 (1993).

van Gils, C.H., Mammographic density and breast cancer risk, *Department of epidemiology*, pp. 128, University of Nijmegen, Nijmegen (1998).

van Oortmarssen, G.J., Habbema, J.D., van der Maas, P.J., de Koning, H.J., Collette, H.J., Verbeek, A.L., Geerts, A.T. and Lubbe, K.T., A model for breast cancer screening. *Cancer*, 66, 1601-12 (1990).

Wald, N.J., Murphy, P., Major, P., Parkes, C., Townsend, J. and Frost, C., UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening [see comments]. *Bmj*, 311, 1189-93 (1995).

Warmerdam, P.G., de Koning, H.J., Boer, R., Beemsterboer, P.M., Dierks, M.L., Swart, E. and Robra, B.P., Quantitative estimates of the impact of sensitivity and specificity in mammographic screening in Germany [see comments]. *J Epidemiol Community Health*, 51, 180-6 (1997).

Whittemore, A.S., Gong, G. and Itnyre, J., Prevalence and contribution of BRCA1 mutations in breast cancer and ovarian cancer: results from three U.S. population-based case-control studies of ovarian cancer. *Am J Hum Genet*, 60, 496-504 (1997).

Woodman, C.B., Threlfall, A.G., Boggis, C.R. and Prior, P., Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's north western region [see comments]. *Bmj*, 310, 224-6 (1995).

SUMMARY

Screening for cancer may be evaluated from different perspectives, which relate to the phases of evaluation. During a randomised trial the evaluation of screening is aimed to demonstrate the effectiveness, but also to facilitate the interpretation of the trial outcomes. After effectiveness has been proofed, the possible introduction as a health-care service should be evaluated. During implementation continuous monitoring is also needed to compare the results with the expectations and the developments in clinical practice. Screening has, besides favourable effects often also important unfavourable effects. Especially the latter justify an extensive and continuous evaluation in these different phases. In this thesis parts, of the evaluation of both breast and prostate cancer are described.

Breast cancer screening is already introduced in the Netherlands and many other European countries. The effectiveness of prostate cancer screening has not yet been shown and therefore a large European randomised trial was started in 1994. This resulted in a different evaluation perspective for breast cancer than for prostate cancer. In this thesis the evaluation of screening of both breast cancer and prostate cancer is combined.

Although the effectiveness of breast cancer screening has been demonstrated in randomised trials and the cost-effectiveness was determined for the Netherlands, the balance between the positive and negative effects of screening might be different in other countries. Factors that might influence the cost-effectiveness of screening are for example, the incidence and mortality of breast cancer, the clinical stage distribution and the organisation of health care. In this thesis the effects and costs of the introduction of a breast cancer screening programme in Germany and Spain were studied

In chapter 2 the effects and costs of a national breast cancer screening programme in Germany were evaluated. Special attention was paid to the decentralised German health-care system and to the influence of attendance, interval and age group. The analysis shows that a programme providing for the screening of women aged 50-69 at

2-year intervals might be expected to result in a decrease in mortality from breast cancer estimated at 11% for the total German population, representing 2,100 deaths from breast cancer prevented each year. The predicted cost per life-year gained was 2 to 3 times less favourable than in the UK and the Netherlands. The sensitivity of mammography was estimated to be approximately 12% lower than in the Netherlands and the attendance rate was calculated at 47% on average (70-80% in the Netherlands). A greater effort to ensure the quality of the screening programme and to improve the invitation system to improve the attendance to the Dutch level might finally lead to a mortality reduction of 18%.

In Spain (Catalonia) the emphasis lay on the effects and costs of different policies for breast cancer screening, to give a basis for setting priorities and deciding on the introduction of a screening programme (chapter 3). The reduction in breast cancer mortality in the total female population due to a screening programme for the age group 50-64 years would be 16, 12 and 9%, with screening intervals of 1, 2 and 3 years respectively. The most cost-effective screening scenario is the one in which women aged 50-69 years are screened with an interval of 3 years with a mortality reduction of approximately 12% in the total female population. Screening until the age of 69 years (2-year interval) was almost as cost-effective as screening the age group 50-64 years with a 2-year interval, with a reduction in breast cancer mortality of 15%. Extension to under the age of 50 years resulted in diverging results depending on the assumptions for improvement in prognosis for younger women (40-49 years). If an extension of a 2 yearly screening programme for women aged 50-64 years is considered, extension to older women would be more advisable, based on proven benefits and costs, than extension to younger age groups.

Although the screening with an interval of 2 years has been introduced in the Netherlands as a national screening programme for all women aged 50-69 and since 1998 until the age of 75, the evaluation still continues. In that phase, secondary effects of the screening were also investigated. In this thesis the effect of radiation risk was evaluated and the use of mammography through the general practitioner, related to the start of the screening programme.

In chapter 4 the effects of radiation due to mammography were studied, to estimate the number of breast-cancer deaths induced by low dose radiation in breast-cancer screening programmes compared to numbers prevented. A computer simulation-model on the natural history of breast cancer was combined with a model on induced breast-cancer mortality from low levels of radiation. Different scenarios (ages and intervals) were used to show the effects on the balance between breast cancer deaths induced and prevented. For a screening programme, age group 50-69, 2-year interval, 2 mGy per view, the balance between the number of deaths induced versus those pre-

vented was favourable; 1:242. When screening is expanded to the age-group 40-49 the results may be less favourable, but if screening was equally effective in young women as in women aged 50-69, the balance would become more favourable again. The new risk estimates by Howe and McLaughlin resulted in 5 to 8 times more favourable balance. Besides age group, dose of mammography is the other determinant of radiation risk. It was concluded that for screening under the age of 50, the balance between the number of breast-cancer deaths prevented by screening versus the number induced by radiation seem less favourable. Credibility-intervals were however wide, because of many uncertainties of radiation risk at very low doses.

Another possibly secondary effect of the introduction of screening is the increase of screening in adjacent (non-invited) age groups and a change in health behaviour in the target population. In chapter 5, we analysed the effect of the start of the Dutch national screening programme on the number of mammograms requested by 43-45 general practices for the age-groups 30-39, 40-49, 50-69 and 70+. In all age-groups an immediate increase was observed in the number of mammography requests after the start of the screening of age-group 50-69 in that area, which was largest and statistically significant in the target-population of the screening programme. More than two years after the start of screening however the number of mammography requests in all age-groups had decreased to the level of before the start and in the age-group 50-69 the number of mammograms was significantly lower than before the screening started. The unexpected increase in mammograms after the start of the breast-cancer screening programme might be related to registry problems or to the process of building up the screening programme. Eventually there was a decrease in the number of mammograms in the target-population, probably as an effect of the introduction of the national screening programme. Screening outside the organised screening programme in adjacent age groups was not clearly demonstrated.

In this thesis some studies were conducted alongside the European Randomised Study of screening for Prostate Cancer (ERSPC) in Rotterdam. The screening protocol was investigated to optimise the combination of screening tests. Furthermore the use of PSA-testing and DRE before and during the trial was studied. Finally the care and costs of advanced disease were studied in this thesis.

For prostate cancer the combination of different screening tests to come to a more efficient screening protocol was studied in chapter 6. The screening protocol initially consisted of 3 screening tests: prostate specific antigen (PSA), digital rectal examination (DRE) and transrectal ultrasonography (TRUS). A PSA value of $\geq$ 4 ng/ml and/or an abnormality on DRE and/or TRUS were an indication for biopsy. Effects of a change in PSA cut-off on the outcomes of the screening were explored. A logistic regression

model was used to predict the number of cancers if all men would have been biopsied. Biopsies in men with a PSA of < 1 ng/ml and a positive DRE or TRUS were very inefficient. Increasing the PSA cut-off to 1.5 or 2 ng/ml to indicate a biopsy would result in 5%-8% cancers less and many (29%-36%) biopsies less. A protocol with PSA ≥ 3 ng/ml as a direct biopsy indication resulted in 7.6% cancers detected less and 12% biopsies less. Given these results and the fact that DRE and TRUS appeared difficult to reproduce, a change in protocol towards PSA ≥ 3 ng/ml seems acceptable. With this protocol DRE and TRUS could be omitted. If screening proofs to be effective in saving prostate cancer deaths, a final judgement about an optimal combination of screening tests should be made.

Determination of (PSA) is a simple test, possibly resulting in frequent usage. This might influence the outcome of the European Randomised study on Screening for Prostate Cancer (ERSPC). In chapter 7, PSA and digital rectal examination (DRE) before and during the screening trial in Rotterdam and in the general population were quantified. Data from the intake questionnaire regarding PSA testing and DRE were analysed to evaluate the use of these tests before participation in the screening study. Data on PSA from the Laboratory of General Practice were linked to information from participants in the screening study. Different sources were used to quantify PSA and DRE in the general population, in a situation without screening. On average, 45% of the men had had a DRE and 13% reported that they had been PSA tested before participating in the trial. Both these percentages increased with age. No statistically significant effects of former PSA-testing or DREs on the cancer detection rate could be demonstrated. The rate of PSA determinations was approximately twice as high in those not offered screening (control-arm) than in those offered screening (screening-arm), which was 76 and 33 per 1000 person-years, respectively. After inclusion in the trial the number of PSA determinations first decreased in the screening arm, but after 6 months an increase was observed. The number of PSA determinations increased in the control arm after inclusion in the trial. The number of PSA determinations in the general population was estimated on 45/1000 person-years. The use of PSA tests in the control arm was moderate, but if different men undergo this test every year the contamination rate in the control arm might be rather high during a screening interval of 4 years. In the final analysis on mortality, PSA-testing should be taken into account. The motives and outcomes of PSA testing would be useful for further interpretation.

If prostate cancer screening proves to be effective in preventing prostate cancer deaths, some cases will be prevented from reaching the advanced stage of prostate cancer. In order to evaluate screening programs thoroughly, it is important to quantify course, care and accompanying costs of advanced disease. In chapter 8, 70 files of patients in 2 hospitals, who had received a diagnosis of distant metastases of prostate can-

cer and had died in the years 1994-1998, were studied. The total healthcare received by these patients, including symptoms and complaints, was recorded. Most frequently reported symptoms were pain (42%), urogenital symptoms (25%) and malaise (20%). 89% of all patients were hormonally treated and 47% received one or more series of radiation therapy. 69% of all patients were treated with pain medication. The average duration of advanced disease was 24 months. Average costs of advanced disease were estimated at $ 11,182 over the total period of which 14% was allocated to assessment and outpatient care and 86% to treatment and costs of hospital stay. These data give a better understanding of the course and costs of advanced prostate cancer. These estimates can be used for the future evaluation of prostate cancer screening.

In chapter 9 some future evaluation challenges of breast cancer and prostate cancer screening are summarised. For breast cancer these challenges consist of new developments in diagnostics, primary prevention and the evaluation of the screening programme. These include digital mammography, fine needle biopsy and tamoxifen to prevent breast cancer. Evaluation challenges of the current breast cancer screening programme are the screening of younger (40-49) and older women (70+), the establishment of the mortality reduction as a result of screening and the question whether the performance of the present national screening programme is the maximum achievable.

The main question to be answered by the evaluation of a prostate screening trial is the quantification of the cancer mortality reduction. However, many other factors like quality of life, have to be quantified in order to give a balanced judgement about the benefit/harm ratio. Future evaluation challenges are the determination of the effectiveness of prostate cancer screening and making up the balance of favourable and unfavourable effects to support a decision to introduce screening.

Screening always has to strike a balance between favourable and adverse effects. This balance should be taken into account during the evaluation of a screening trial, during the possible introduction and during the execution of a screening programme. The benefit/harm ratio of screening for cancer is subject to continuous change, as a result of, for example, improvement in test properties and diagnostic and therapeutic developments. This necessitates continuous alertness to adapt the programme in order to obtain optimal results. Obviously, the ideal remains to find effective forms of primary prevention or therapy that will make screening superfluous.

SAMENVATTING

Evaluatie van kankerscreening kan met verschillende doelstellingen worden uitgevoerd. Deze doelstellingen hangen samen met de fase waarin de evaluatie zich bevindt. Het belangrijkste doel van evaluatie van een gerandomiseerde screeningstudie is de effectiviteit van de screening aan te tonen. De evaluatie in die fase is echter ook van belang om de interpretatie van de uitkomsten te vergemakkelijken. Als effectiviteit is aangetoond kan een mogelijke invoering van de screening als een gezondheidszorg-voorziening worden geëvalueerd. Een continue evaluatie tijdens de uitvoering van een programma is nodig om de resultaten af te zetten tegen de verwachtingen en veranderingen in de klinische praktijk. Screening kent naast gunstige effecten ook nadelige effecten. Vooral de nadelige effecten maken evaluatie in deze verschillende fases van belang. In dit proefschrift worden delen van de evaluatie van zowel borstkanker als prostaatkanker beschreven.

Borstkankerscreening is in Nederland en in sommige andere Europese landen reeds geïntroduceerd. In juni 1994 is een gerandomiseerd onderzoek gestart om de effectiviteit van screening op prostaatkanker aan te tonen. Dit resulteerde in een ander uitgangspunt voor de evaluatie van borstkankerscreening dan voor prostaatkanker.

Hoewel de effectiviteit van borstkankerscreening was aangetoond in gerandomiseerde studies en de kosten-effectiviteit was vastgesteld voor Nederland, kan de balans tussen positieve en negatieve effecten van de screening in andere landen anders uitpakken. Factoren die van invloed kunnen zijn op de kosten-effectiviteit van screening zijn bijvoorbeeld het niveau van incidentie en sterfte aan borstkanker, de klinische stadiumverdeling en de organisatie van de gezondheidszorg. In dit proefschrift zijn de effecten en kosten onderzocht van invoering van een screeningprogramma in Duitsland en Spanje.

In hoofdstuk 2 is dit voor Duitsland geëvalueerd. Speciale aandacht werd besteed aan opkomst, screeningsinterval, leeftijdsgroep en aan de organisatie van de gezondheidszorg in relatie tot screening, die in Duitsland decentraal is. De analyse liet zien dat een screeningprogramma voor vrouwen van 50-69 met een interval van 2 jaar zou

resulteren in een sterftereductie van 11%, hetgeen overeenkomt met 2100 sterfgevallen per jaar. De kosten per gewonnen levensjaar waren echter 2 tot 3 keer hoger dan in Nederland en Engeland. De sensitiviteit van de mammografie was 12% lager dan in Nederland en de verwachte opkomst bedroeg slechts 47% (70-80% in Nederland). Een grotere inspanning om de kwaliteit van de screening te garanderen en de opkomst te verhogen tot de Nederlandse waarden, zou kunnen leiden tot een sterftereductie van 18% in Duitsland.

De nadruk van de analyse van de effecten en kosten van screening in Spanje (Catalonië) lag op verschillende leeftijdsgroepen en screeningsintervallen (hoofdstuk 3). De uitkomsten dienden als basis voor een beslissing over de invoering van screening. Voor de leeftijdsgroep 50-64 zou een screeningsprogramma met een 1, 2 of 3 jaars interval resulteren in een sterftereductie van respectievelijk 16%, 12% en 9%. Het meest kosten-effectief was het programma waarbij de leeftijdsgroep 50-69 met een 3 jaars interval gescreend werd. Screening tot 69 jaar met een 2 jaars interval was bijna even kosten-effectief als voor de leeftijdsgroep 50-64 en resulteerde in een sterftereductie van 15%. Een uitbreiding naar onder de leeftijd 50 resulteerde in verschillende bevindingen afhankelijk van de aanname over effectiviteit van screening in deze leeftijdsgroep. Als een uitbreiding van screening in de leeftijd van 50-64 wordt overwogen is het wenselijker naar oudere leeftijdsgroepen uit te breiden dan naar jongere.

Hoewel screening met een 2 jaars interval in Nederland is geïntroduceerd voor vrouwen van 50-69 en sinds 1998 tot de leeftijd 75, wordt de evaluatie gecontinueerd. In die fase is, onder andere, de evaluatie van secundaire effecten van screening van belang. In dit proefschrift is het risico van door straling geïnduceerde sterfte onderzocht. Ook is het effect van het bevolkingsonderzoek in de leeftijdsgroep 50-69 op het gebruik van mammografie via de huisarts buiten deze leeftijdsgroep bestudeerd.

In hoofdstuk 4 is het effect van röntgenstraling van mammografie bestudeerd om het aantal verwachte voorkomen borstkanker sterfgevallen af te zetten tegen het aantal geïnduceerde gevallen ten gevolge van straling. Uit dit onderzoek bleek dat de balans gunstig uitviel voor de doelgroep van de screening (leeftijd 50-69, 2 jaar interval, 2 mGy per opname). Het aantal verwachte voorkomen sterfgevallen bedroeg 242 tegenover één geïnduceerd sterfgeval. Uitbreiding naar de leeftijdsgroep 40-49 resulteerde in een minder gunstige balans, maar als aangenomen werd dat screening van jongere vrouwen even effectief is als van vrouwen van 50-69 jaar, werd de balans weer gunstiger. Ook toepassing van de recent verschenen schattingen van stralingsrisico van Howe en Mclaughlin resulteerde in een 5 tot 8 keer gunstiger balans. Naast leeftijdsgroep is stralingsdosis een andere belangrijke determinant van stralingsrisico. Geconcludeerd werd dat voor jongere vrouwen de balans minder gunstig is dan voor

vrouwen van 50-69 jaar. De betrouwbaarheidsintervallen zijn echter groot, omdat er veel onzekerheden zijn over stralingsrisico bij deze lage doses.

Andere secundaire effecten van screening zijn de toename van mammografie buiten de doelgroep en een verandering in gedrag van de doelgroep. In hoofdstuk 5 is het effect van de invoering van het Nederlandse bevolkingsonderzoek op de hoeveelheid mammografie via de huisarts onderzocht in de leeftijdsgroepen 30-39, 40-49, 50-69 en 70+. In alle leeftijdsgroepen werd een toename in mammografie via de huisarts waargenomen na het begin van het bevolkingsonderzoek in de leeftijdsgroep 50-69. Twee jaar na de start, was het niveau van mammografie buiten de doelgroep echter weer gelijk aan het niveau van voor de start en in de doelgroep van de screening zelfs significant lager. De onverwachte toename zou kunnen worden verklaard door een probleem met de registratie of door de opbouw van de screening. De afname van mammografie via de huisarts in de doelgroep van de screening is waarschijnlijk een gevolg van het landelijke programma. Screening via de huisarts buiten de doelgroep werd niet duidelijk aangetoond.

In dit proefschrift is ook een aantal studies uitgevoerd in relatie tot het Rotterdamse deel van de Europese gerandomiseerde studie naar screening op prostaatkanker. Het screeningprotocol werd onderzocht om de combinatie van screeningtesten te optimaliseren. Ook werd het gebruik van een test op Prostaat Specifiek Antigen (PSA) en rectaal toucher voor en tijdens deelname aan de screeningstudie onderzocht. Het laatste onderzoek in dit proefschrift beschrijft het beloop, de zorg en de kosten van gemetastaseerd prostaatkanker.

De combinatie van de 3 screeningtesten (PSA-test, rectaal toucher en trans rectale ultrasonografie (TRUS)) werd bestudeerd in hoofdstuk 6. Een waarde voor PSA van meer dan 4 ng/ml of een afwijking bij rectaal toucher of TRUS waren een indicatie voor biopsie bij het begin van de studie. Ten eerste werd het effect van een verandering in de afkapwaarde van PSA bepaald. Bovendien werd een logistisch regressie model gebruikt om het aantal verwachte kankers te berekenen in het geval alle mannen zouden zijn gebiopteerd. Biopten onder een PSA van 1 ng/ml waren erg inefficiënt. Een toename in PSA afkapwaarde naar 1,5 ng/ml of 2 ng/ml zou resulteren in 5 tot 8% minder ontdekte tumoren, maar ook tot een afname van 29 tot 36% biopsieën. Een protocol waarbij iedereen (dus onafhankelijk van de bevindingen bij rectaal toucher en TRUS) werd gebiopteerd met een PSA waarde van meer dan 3 ng/ml zou resulteren in 7,6% minder kankers en 12% minder biopten. Gegeven deze resultaten en het feit dat rectaal toucher en TRUS moeilijk te reproduceren waren, leek een protocol met een biopsie voor iedereen met een PSA groter dan 3 ng/ml acceptabel. Bij dit protocol konden ook het rectaal toucher en de TRUS achterwege blijven. Als screening effectief

blijkt te zijn om sterfte aan prostaatkanker te verminderen, moet weer een beoordeling van de meest gunstige combinatie van screeningtesten plaatsvinden.

Een PSA bepaling is een zeer eenvoudige test, waardoor deze mogelijk op grote schaal zal worden toegepast. Dit kan de uitkomst van de Europese studie naar het effect van screening op prostaatkankersterfte beïnvloeden. In hoofdstuk 7 is het gebruik van testen, zoals rectaal toucher en PSA voor en tijdens de gerandomiseerde screeningstudie en in de algemene populatie in kaart gebracht. Op basis van de gegevens van de vragenlijst, die bij deelname was ingevuld, is het gebruik van deze testen voor deelname aan de screeningstudie gekwantificeerd. Data van de deelnemers in de studie werden gekoppeld aan data van het huisartsen laboratorium, waarmee het gebruik van PSA tijdens de screeningstudie, van zowel de screening- als de controlegroep, kon worden vastgesteld. Voor het vaststellen van PSA en rectaal toucher in de algemene populatie werden verschillende bronnen gebruikt, waaronder gegevens van een zorgverzekeraar. Gemiddeld had 45% van de mannen een rectaal toucher ondergaan voor deelname aan de screeningstudie en 13% een PSA test. Deze percentages namen toe met de leeftijd. Er kon echter geen effect van deze testen op de kankerdetectie tijdens de studie worden waargenomen. Het aantal PSA testen tijdens de studie was ongeveer twee keer zo hoog in de controlegroep als in de screeninggroep (76/1000 en 33/1000 per jaar). In de screeninggroep werd na deelname eerst een afname in het aantal PSA testen waargenomen, maar na ongeveer 6 maanden weer een toename. Het aantal PSA testen in de algemene bevolking werd geschat op 45/1000 per jaar. Geconcludeerd werd dat het gebruik van PSA testen tijdens de studie matig was, maar als ieder jaar andere mannen deze test ondergaan, zou het percentage contaminatie hoog kunnen zijn gedurende het interval van 4 jaar. In de definitieve sterfte-analyse moet contaminatie in beschouwing worden genomen. Inzicht in de motieven en uitkomsten van de PSA testen zou nuttig zijn voor aanvullende interpretatie.

Indien prostaatkankerscreening effectief blijkt in het voorkomen van sterfte ten gevolge van deze ziekte, zal een deel van de patiënten het gemetastaseerde stadium van prostaatkanker niet meer doormaken. Bij de evaluatie van screeningprogramma's moet dit in beschouwing worden genomen en is het van belang het beloop, de zorg en de daarmee gepaard gaande kosten in kaart te brengen. In hoofdstuk 8 zijn 70 statussen bestudeerd van patiënten met gemetastaseerd prostaatkanker, die waren gestorven in de jaren 1994-1998. De totale zorg inclusief symptomen en klachten werden geregistreerd. De meest frequent gerapporteerde symptomen waren pijn (42%), urogenitale klachten (25%) en malaise (20%). Van de patiënten werd 89% hormonaal behandeld en 47% onderging één of meerdere series van bestraling. 69% van de patiënten werd behandeld met pijnmedicatie. De gemiddelde duur van gemetastaseerd prostaatkanker was 24 maanden. De gemiddelde kosten van het gemetastaseerde

stadium van de ziekte bedroeg $ 11,182, waarvan 14% werd besteed aan diagnostiek en poliklinische zorg en 86% aan behandeling en ziekenhuisopnames. Deze studie heeft bijgedragen aan een duidelijker beeld van het beloop en de kosten van gemetastaseerd prostaatkanker. Deze schattingen kunnen in de toekomst worden gebruikt bij de verdere evaluatie van prostaatkankerscreening.

In hoofdstuk 9 wordt een aantal toekomstige uitdagingen ten aanzien van de evaluatie van borst- en prostaatkankerscreening samengevat. Voor borstkanker bestaan deze uitdagingen uit nieuwe ontwikkelingen in diagnostiek, primaire preventie en uit de evaluatie van de screening. Voorbeelden zijn digitale mammografie, dunne naald biopsie en het gebruik van tamoxifen om borstkanker te voorkomen. Uitdagingen voor de screeningevaluatie van het huidige borstkanker screeningprogramma zijn screening van jongere (40-49) en oudere vrouwen (70+), het vaststellen van de sterftereductie ten gevolge van een ingevoerd screeningprogramma en de vraag of het huidige bevolkingsonderzoek verbeterd kan worden.

De belangrijkste vraag waar de evaluatie van de prostaatscreeningstudie antwoord op moet geven is de prostaatkanker sterftereductie. Andere effecten, zoals bijvoorbeeld de kwaliteit van leven, moeten ook worden gekwantificeerd om een evenwichtig oordeel te kunnen geven over de balans tussen gunstige en ongunstige effecten. De uitdagingen van de evaluatie bestaan met name uit het vaststellen van de effectiviteit van prostaatkankerscreening en de ondersteuning van de beslissing over de invoering van een programma.

Screening is altijd een balans tussen gunstige en ongunstige effecten. Deze balans moet in beschouwing worden genomen gedurende de evaluatie van een screeningstudie naar de effectiviteit, gedurende de mogelijke invoering en gedurende de uitvoering van screeningprogramma's. Het is duidelijk dat de nut/risico-verhouding van kankerscreening kan veranderen ten gevolge van bijvoorbeeld verbeteringen in testeigenschappen, diagnose of therapeutische mogelijkheden. Dit maakt continue oplettendheid nodig om het programma aan te passen, opdat de meest optimale resultaten worden bereikt. Het ideaal is zonder twijfel om effectieve vormen van primaire preventie of behandeling te vinden, waardoor screening overbodig wordt.

LIST OF PUBLICATIONS

This thesis

Chapters 2 to 8 are based on the following papers and manuscripts:

Beemsterboer, P.M.M., de Koning, H.J., Warmerdam, P.G., Boer, R., Swart, E., Dierks, M-L., Robra B-P., Prediction of the effects and costs of breast-cancer screening in Germany, *Int J Cancer*, 58, 623-628 (1994).[1]

Beemsterboer, P.M.M., Warmerdam, P.G., Boer, R., Borras, J. M., Moreno, V., Viladiu, P., de Koning, H.J., Screening for breast cancer in Catalonia, which policy is to be preferred? *Eur J Public Health*, 8, 214-246 (1998)[2]

Beemsterboer, P.M.M., Warmerdam, P.G., Boer, R., de Koning, H.J., Radiation risk of mammography related to benefit in screening programmes: a favourable balance? *J Med Screen*, 5, 81-87 (1998).[4]

Beemsterboer, P.M.M., de Koning, H.J., Looman, C.W.N., Borsboom, G.J.J.M., Bartelds, I.A.M., van der Maas, P.J., Mammography requests in general practice during the introduction of nationwide breast-cancer screening, 1988-1995, *Eur J Cancer*, 35, 450-454 (1999).[3]

Beemsterboer, P.M.M., Kranse, R., de Koning, H.J., Habbema, J.D.F., Schröder, F.H., Changing role of 3 screening modalities in the European Study of Screening for Prostate Cancer (Rotterdam), *Int J Cancer* (in press).[1]

Beemsterboer, P.M.M., de Koning, H.J., Kranse, R., Trienekens, P.H., van der Maas, P.J., Schröder, F.H., PSA-testing and use of DRE before and during a randomised trial of screening for prostate cancer (ERSPC, Rotterdam), (submitted).

Beemsterboer, P.M.M., de Koning, H.J., Birnie, E., van der Maas, P.J., Schröder, F.H., Advanced prostate cancer: course, care and cost implications, *the Prostate* (in press).[1]

Other scientific publications

Boer, R., de Koning, H.J., Beemsterboer, P.M., Warmerdam, P.G. and Schröder, F.H., A comparison of disease specific survival of patients who died of and who had newly diagnosed prostate cancer. *J Urol*, 157, 1768-71; discussion 1771-2 (1997).

Borras, J.M., Espinas, J.A., Beemsterboer, P.M.M., Granados, A. and de Koning, H.J., Anticipating the consequences for the primary therapy of breast cancer after introducing screening. A more global picture for health care policy making. *Int J Technol Assess Health Care*, 14, 268-76 (1998).

de Koning, H.J., Boer, R., Warmerdam, P.G., Beemsterboer, P.M.M. and van der Maas, P.J., Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials [see comments]. *J Natl Cancer Inst*, 87, 1217-23 (1995).

Fracheboud, J., de Koning, H.J., Beemsterboer, P.M.M, Boer, R., Hendriks, J.H., Verbeek, A.L., van Ineveld, B.M., de Bruyn, A.E. and van der Maas, P.J., Nation-wide breast cancer screening in The Netherlands: results of initial and subsequent screening 1990-1995. National Evaluation Team for Breast Cancer Screening. *Int J Cancer*, 75, 694-8 (1998).

Fracheboud, J., de Koning, H.J., Beemsterboer, P.M.M., Boer, R., Verbeek, A.L.M., Hendriks, J.H.C.L., van Ineveld, B.M. Broeders, M.J.M., de Bruijn A.E., and van der Maas, P.J., Interval cancers in the nationwide breast cancer screening programme in the Netherlands. *submitted* (1999).

Kranse, R., Beemsterboer, P.M.M., Rietbergen, J.B.W., Habbema, J.D.F., Hugosson, J. and Schröder, F.H., Predictors for biopsy outcome in the European Randomized Study of Screening for Prostate Cancer (Rotterdam region). *the Prostate* (accepted for publication).

Schröder, F.H., van der Maas, P., Beemsterboer, P.M.M., Kruger, A.B., Hoedemaeker, R., Rietbergen, J. and Kranse, R., Evaluation of the digital rectal examination as a screening test for prostate cancer. Rotterdam section of the European Randomized Study of Screening for Prostate Cancer [see comments]. *J Natl Cancer Inst*, 90, 1817-23 (1998).

Warmerdam, P.G., de Koning, H.J., Boer, R., Beemsterboer, P.M.M., Dierks, M.L., Swart, E. and Robra, B.P., Quantitative estimates of the impact of sensitivity and specificity in mammographic screening in Germany [see comments]. *J Epidemiol Community Health*, 51, 180-6 (1997).

DANKWOORD

Na een rustig begin, raakte ik bij het schrijven van dit proefschrift in steeds sneller vaarwater. In het begin deed ik onderzoek om te beoordelen of ik wel onderzoeker wilde zijn. Vervolgens nam de stroom mij mee en kwamen er nieuwe projecten en uitdagende taken bij. Het eindigde als een waterval, door in snel tempo twee artikelen, inleiding en discussie van dit proefschrift af te ronden. Hier het resultaat!

Veel mensen hebben direct of indirect bijgedragen aan het tot stand komen van dit proefschrift en aan hen wil ik graag woorden van dank wijden. Ten eerste aan mijn co-promotor Harry de Koning. Vanaf het begin was jij mijn directe begeleider. De ins en outs van de (klinische) screeningspraktijk heb jij me bijgebracht en alle stukjes die ik geschreven heb zeker 2 keer gelezen (als het niet meer is). Verder bedank ik je voor jouw nooit aflatende optimisme. Vooral in het laatste jaar heeft dat zo nu en dan gezorgd voor een flinke oppepper.

Paul van der Maas wilde meestal op het laatste moment een blik werpen op het door mij geleverde werk. Paul, jouw commentaar heeft iedere keer tot een significante verbetering van mijn stukken geleid en daarom ben ik erg blij dat je ernaar gekeken hebt.

Met Peter Warmerdam mijn kamergenoot en onderzoeksmaatje van de eerste jaren heb ik altijd met veel plezier samengewerkt. De eerste 3 artikelen van dit proefschrift zijn een co-productie van jou en mij geweest. Tegen jou kon ik, in het wilde weg mijn gedachten spuien en we kwamen telkens wel tot een aanvaardbare oplossing voor de obstakels in het onderzoek.

Voor de echt moeilijke dingen was daar Rob Boer. Rob, ik bewonder jouw onderzoekskwaliteiten en waardeer dat je mijn 'problemen' altijd zeer serieus nam en zeer snel een oplossing wist aan te dragen. Ik heb veel van je geleerd!

Met Mark Wildhagen, m'n kamergenoot gedurende de hele MGZ-periode (bijna 6 jaar) heb ik heel wat lief en leed gedeeld. Hoewel we erg verschillende personen zijn,

kon ik altijd mezelf zijn en even m'n ei kwijt over het werk, maar ook privé. Dank daarvoor. De laatste periode hebben we ook samengewerkt aan de prostaatscreening, dat na zo'n lange 'inwerkperiode' prima verliep.

Arry de Bruyn, de duizendpoot voor Harry, maar ook voor de LETB en voor mij. Hartelijk dank voor alle klussen die je altijd goedgehumeurd voor me hebt gedaan en voor je oprechte belangstelling in mij als persoon. Ik heb dat beide altijd erg gewaardeerd. Soms helpt dat om op de been te blijven!

Met Ries Kranse van urologie heb ik heel wat discussies gevoerd over screening en het verloop van het project. Bovendien kon ik voor ingewikkelde koppelingen en andere automatiseringskwesties altijd direct bij jou terecht. Beide zorgden voor nieuwe motivatie.

Professor Schröder bedank ik voor de mogelijkheden om in samenwerking met urologie bepaalde deelaspecten van de screening op prostaatkanker te onderzoeken. Uw snelle en positieve reacties op de manuscripten werkten inspirerend.

Al mijn co-auteurs (18 in totaal) bedank ik voor het commentaar op m'n artikelen. Dat heeft zeker tot verbeteringen geleid!

Ik bedank ook alle andere collega's van MGZ. Aan de periode op MGZ heb ik veel positieve en leuke herinneringen overgehouden. Bij MGZ wordt op een of andere manier geselecteerd op mensen waarmee het prima werken is! Het 25-jarig jubileum van MGZ was voor mij een hoogtepunt. Ik ben nog steeds blij dat ik voor het cabaret in het achtergrondkoor heb mogen optreden.

Door MGZ heb ik zelfs 5 goede vriendinnen overgehouden! Deze 'mutsen' hebben bijgedragen aan heel wat lol tussen de artikelen, analyses en vergaderingen door. De wandelvakantie in Wales was een kroon op het werk! Helaas laat ik het de komende vakantie afweten, maar ik weet zeker dat we elkaar nog heel vaak zullen opzoeken om, onder het genot van een flesje wijn, onze ervaringen uit te wisselen!

Caroline Baan, één van mijn paranimfen (en één van de mutsen), bleef met mij, als laatste van de mutsen over op MGZ. In die periode heb jij mij heel erg geholpen de lijn en voortgang in het werk te blijven zien. Bovendien wisselden we iedere dag ons wel en wee even uit, wat een goed en inspirerend begin van de werkdag was! Hoewel niet meer lijfelijk aanwezig hoop ik voor jou ook tot steun te zijn in de laatste fase; succes met de laatste loodjes!

Hester de Melker, mijn andere paranimf, bedankt voor alle steun en alle gezelligheid! Dat wat jij voor mij hebt betekend is moeilijk in woorden uit te drukken. Als vakgenoot weet je precies wat het is om een proefschrift te schrijven en als vriendin kon ik altijd met alles bij je terecht (en dat was niet gering de afgelopen tijd!). Dat heeft me vaak door moeilijke tijden gesleept. Door je grote relativeringsvermogen (dat het bij

nog wel eens laat afweten) zorgde je er vaak voor dat ik het allemaal weer een stuk positiever zag. Ook veel succes met de afronding van jouw proefschrift!

Ik dacht, als ik de dankwoorden van andere promovendi las over verwaarlozing van vriendschappen, dat dat mij niet zou overkomen. Een proefschrift schrijven is wel belangrijk, maar er zijn ook veel andere belangrijke zaken in het leven. Nu ontkom ik zelf toch niet aan het feit dat ik mijn vrienden en familie de laatste tijd wel een stuk minder aandacht heb gegeven dan ik zou willen. Ik verwacht dat, nu dit werk er bijna opzit, ik weer mijn volle aandacht aan jullie kan besteden. De basis is volgens mij zodanig, dat we weer op de oude voet verder kunnen gaan! Ik kijk ernaar uit!

Mijn ouders bedank ik voor de vrijheid die ze me altijd hebben gegeven om te doen wat ik graag wilde en voor het onvoorwaardelijke vertrouwen in mijn beslissingen en keuzes. Mijn zus, Carolien heeft de voorkant geschilderd en voor dit proefschrift bewerkt. Ik ben heel blij met het resultaat!

Als laatste bedank ik Reijer, al was hij 'alleen' tot steun bij de laatste loodjes. We gaan het avontuur tegemoet; ik heb er zin in!

CURRICULUM VITAE

Petra Beemsterboer werd geboren op 23 juni 1968 in Heemskerk. Na het doorlopen van de lagere school ging zij naar het Pius X College in Beverwijk, waar zij in 1987 haar diploma gymnasium β behaalde. In 1987 begon zij aan de studie Gezondheidswetenschappen (later Biomedische Wetenschappen geheten) aan de Rijks Universiteit Leiden. In de doctoraalfase besloot zij zich te specialiseren in de richting Epidemiologie. Na verschillende onderzoekstages bij de Rijks Universiteit Leiden (vakgroep Fysiologie en Verloskunde, Vrouwenziekten en Voortplanting), TNO Preventie en Gezondheid (afdeling Jeugd) in Leiden en het Nederlands Kanker Instituut (afdeling epidemiologie) in Amsterdam, studeerde zij in 1992 af. In die periode werd zij tevens geregistreerd als epidemioloog A. Aansluitend kwam zij in dienst bij het Instituut Maatschappelijke Gezondheidzorg van de Erasmus Universiteit. De eerste 2 jaar was zij betrokken bij evaluatie van bevolkingsonderzoek naar borstkanker in andere Europese landen. Aansluitend kwamen daar taken bij in het kader van de landelijke evaluatie van bevolkingsonderzoek naar borstkanker, waarvoor zij sinds 1994 lid was van het Landelijk Evaluatie Team bevolkingsonderzoek Borstkanker (LETB). Vanaf 1995 was zij ook betrokken bij de evaluatie van screening op prostaatkanker. In oktober 1998 is zij als wetenschappelijk stafmedewerker in dienst getreden bij de Gezondheidsraad in Den Haag.