Quantitative estimates of the impact of sensitivity and specificity in mammographic screening in Germany

Peter G Warmerdam, Harry J de Koning, Rob Boer, Petra M M Beemsterboer, Marie-Luise Dierks, Enno Swart, Bernt-Peter Robra

Abstract

Study objective – To estimate quantitatively the impact of the quality of mammographic screening (in terms of sensitivity and specificity) on the effects and costs of nationwide breast cancer screening.

Design – Three plausible "quality" scenarios for a biennial breast cancer screening programme for women aged 50–69 in Germany were analysed in terms of costs and effects using the *Microsimulation Screening Analysis* model on breast cancer screening and the natural history of breast cancer. Firstly, sensitivity and specificity in the expected situation (or "baseline" scenario) were estimated from a model based analysis of empirical data from 35 000 screening examinations in two German pilot projects. In the second "high quality" scenario, these properties were based on the more favourable diagnostic results from breast cancer screening projects and the nationwide programme in The Netherlands. Thirdly, a worst case, "low quality" hypothetical scenario with a 25% lower sensitivity than that experienced in The Netherlands was analysed.

Setting – The epidemiological and social situation in Germany in relation to mass screening for breast cancer.

Results – In the "baseline" scenario, an 11% reduction in breast cancer mortality was expected in the total German female population, ie 2100 breast cancer deaths would be prevented per year. It was estimated that the "high quality" scenario, based on Dutch experience, would lead to the prevention of an additional 200 deaths per year and would also cut the number of false positive biopsy results by half. The cost per life year gained varied from Deutsche mark (DM) 15 000 in the "high quality" scenario to DM 21 000 in the "low quality" setting.

Conclusions – Up to 20% of the total costs of a screening programme can be spent on quality improvement in order to achieve a substantially higher reduction in mortality and reduce undesirable side effects while retaining the same cost effectiveness ratio as that estimated from the German data.

(*J Epidemiol Community Health* 1997;51:180–186)

Department of Public Health, Erasmus University Rotterdam, PO Box 1738, 3000 DR, Rotterdam, The Netherlands
P G Warmerdam
H J de Koning
R Boer
P M M Beemsterboer

Abteilung Epidemiologie und Sozialmedizin, Medizinische Hochschule Hannover, Germany
M-L Dierks

Institut für Sozialmedizin, Otto-von-Guericke Universität, Magdeburg, Germany
E Swart
B-P Robra

Correspondence to:
Mr P G Warmerdam.

Breast cancer screening for women aged 50 and over is being implemented in a large number of European countries. These programmes aim to achieve a relatively high level of quality by means of strict organisation, training, and evaluation.[1][2] The definition of screening quality is often narrowed down to two properties – *sensitivity* (ie the capability of accurately detecting cancer at a screening examination and *specificity* (the ability to pinpoint accurately those women without breast cancer). High sensitivity can be assumed to be one of the key elements in achieving a reduction in breast cancer mortality, while high specificity will reduce the negative side effects of screening.[3]

The nationwide programmes in the United Kingdom and The Netherlands have made a good start in providing facilities which ensure high standards. It is not certain, however, that these high standards can be achieved in all countries or programmes, especially where the organisation of screening is not centralised. Where screening is carried out in the private practices of gynaecologists or radiologists it may be of poorer quality than that undertaken in specialist screening units. The reasons for this are that in a decentralised screening setting quality control is less easily achieved, radiologists who are not specialists in this area may be less skilled in screening mammography, and the mammography equipment may be of an inferior quality.

The decentralised health care setting in Germany has occasioned discussion about the introduction of a nationwide breast cancer screening programme.[4][5] In 1990, a pilot study was set up in the Aurich and Braunschweig regions to investigate quality and to test the quality assurance necessary for the implementation of a nationwide programme (the Deutsche Mammographie Studie, DMS).[6] Data from this DMS study support the hypothesis of less than optimal quality: relatively old equipment, poorer image quality, and a relatively low number of mammograms per radiologist. Although it has been possible to estimate the effects and costs of breast cancer screening in other countries with different backgrounds on the basis of incidence and mortality levels,[7] the consequences of lower quality on the expected results of a screening programme have not yet been quantified in detail. It is possible that a decentralised screening setting, such as the one in Germany, might affect the quality of screening so adversely that a breast

cancer screening programme is not worth implementing.

In this study we have estimated the expected sensitivity and specificity for a nationwide German programme, based on the results from the Aurich and Braunschweig pilot projects. Screening performance results predicted by the *Microsimulation Screening Analysis (MISCAN)* model[8] were compared with the observed data in these projects to estimate sensitivity. Since the true extent of the future quality in a decentralised system is of course difficult to assess, we have defined two other sets of plausible quality assumptions as follows:

1 A "high quality" scenario based on the situation in the Dutch nationwide programme; and
2 A hypothetical "low quality" scenario with a 25% lower sensitivity than that experienced in The Netherlands.

The impact of these various quality assumptions (sensitivity and specificity) on the outcomes of a screening programme was evaluated and quantified in terms of a reduction in breast cancer mortality and costs and in negative side effects. The results from this study give indications for the cost effectiveness of quality assurance and evaluation.

Methods

THE *MISCAN* APPROACH

In this study the *MISCAN* model for screening analysis and evaluation was used.[8 9] The natural history of breast cancer was reflected in this model in four states – one ductal carcinoma in situ DCIS) and three invasive states with increasing tumour size. By incorporating demographic aspects, individual life histories can be generated to simulate a "real life" dynamic population with breast cancer incidence and mortality, and death from other causes. A screening programme can be applied to this population to determine what changes would occur. Various characteristics of screening policies or programmes can be defined, such as screening ages, interval, and attendance. Important screening parameters embedded in the model were the sensitivity and specificity of the screening test and the improvement in prognosis after screen detection.

The sensitivity of a screening examination is an age and state dependent model parameter, defined as the probability of a positive screen result when screening a woman with preclinical breast cancer. Preclinical cancer is defined here as a lesion that has not yet been clinically diagnosed. Specificity is defined as the probability of a negative test result in a woman without preclinical cancer. Given the preclinical prevalences and detection rates as model outcomes, specificity of screening can be modified in the model by adjusting the positive predictive values (PPVs) of the subsequent additional diagnostic steps. In this way, the numbers of diagnostic assessments without breast cancer being diagnosed (false positives) can be determined.

Improvement in prognosis due to early detection was defined in the model as the reduction in risk of dying from breast cancer for screen detected cases, and was dependent on age and cancer stage. The estimate of this improvement in prognosis after screening detection was based on the breast cancer mortality reduction reported in the overview of the Swedish randomised trials.[9 10]

The design of the analysis on effects and costs of breast cancer screening has been described before.[11 12] In short, the effects and costs of different policies are compared to the situation in which mass screening is not applied, based on national data on assessment and treatment. Adjustments have been made for the German situation if data were available. This method has been shown to be useful for predicting the effects and costs of screening in Australia[13] and Italy.[14]

THE GERMAN MODEL: DEMOGRAPHY AND EPIDEMIOLOGY

The demography was based on the 1989 age specific distribution of the total German female population; death rates from causes other than breast cancer were based on 1990 data from the former West Germany, combined with less detailed data from the former German Democratic Republic (GDR). Breast cancer mortality rates were calculated on the basis of 1990 data from the former West Germany, incorporating 25% lower rates for the former GDR.[15 16] The clinical stage distribution of breast cancer was based on published series[17 18] and assumed not to be different from The Netherlands.[19] Data from the Saarland cancer registry were used to determine the breast cancer incidence at national level.[20] Regional differences in incidence were assumed to be negligible on the basis of the regional mortality data for breast cancer.[21] Relative survival rates in the model 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 observed data.

THE GERMAN MODEL: PRECLINICAL DURATION OF BREAST CANCER AND SENSITIVITY

Age specific assumptions on the mean durations of preclinical states in The Netherlands were based on data from the Dutch screening projects in Utrecht and Nijmegen and have been published before.[8 11] The mean duration of preclinical disease was estimated to range from 2.7 (age 50) to 6.2 (age 70) years. Durations for the preclinical stages of breast cancer in Germany were assumed not to be different from those estimated for The Netherlands.

The expected sensitivity of a screening programme in Germany has been assessed from the results of two German pilot projects, Aurich and Braunschweig, in the Deutsche Mammographie Studie.[6] In these regions, breast cancer screening has been carried out since 1990. Data from 35 000 screening examinations (27 000 of which were first screens) in the first two years (1990.III-1992.III) were

Table 1 Sensitivity in relation to tumour stage, and referral rates (for additional examinations and biopsy) and positive predictive values (PPV) (of a referral for biopsy), for first and subsequent screens as proxy indicators for specificity in three quality scenarios – baseline, high quality and low quality

Sensitivity* per tumour stage (size)	Baseline		High		Low	
Ductal carcinoma in situ	0.350		0.400		0.300	
Invasive <10 mm	0.620		0.700		0.525	
Invasive 10–19 mm	0.840		0.950		0.713	
Invasive ≥20 mm	0.840		0.950		0.713	
Referral rates (%)† and PPV for 1st and subsequent (Subs) screens‡	Baseline		High		Low	
	1st	Subs	1st	Subs	1st	Subs
Referred for additional examination	3.9	2.1	0.9	0.5	3.3	1.9
Referred for biopsy after additional examination	24	24	76	79	24	24
PPV (biopsy)	34	53	52	73	34	53

* In the *MISCAN* screening model sensitivity is defined as the stage specific probability of a positive screen result when screening a woman with preclinical screen detectable breast cancer.
† Women with a positive screen result are referred for additional examination; women with a positive result for additional examination are referred for biopsy.
‡ Mean values for a screening programme in Germany for women aged 50–69 (2 year interval) starting in 1994 and running for 27 years.

analysed for specific age distribution and screening interval. To simulate these pilot projects a *MISCAN* analysis model was constructed using this information together with the assumptions on German breast cancer incidence, clinical stage distribution, and mean durations of preclinical disease. Observed detection rates and stage distribution of screen detected cancers were compared with the values predicted by the model for different assumptions on sensitivity. Calibration in this model based analysis lead to an approximately 12% lower sensitivity for the DMS pilot projects than had been experienced in The Netherlands (see Results section).

THREE SCENARIOS

For the "baseline" scenario (the expected situation) of a screening programme in Germany, the results from the DMS pilot projects were used, ie sensitivity 12% lower than estimated for the Netherlands. Specificity in this scenario was based on PPVs of referrals for additional examination and biopsy as observed in these pilot projects.

The "high quality" scenario was based on the more favourable diagnostic results in the Dutch screening projects and the nationwide programme. The specificity of the Dutch programme was estimated with the use of the age specific referral and detection rates (all rounds) in the Utrecht and Nijmegen projects.[11] The new results in the nationwide programme were consistent with these earlier assumptions.[2] In the high quality scenario Dutch sensitivity and PPVs were used.

Thirdly, a "low quality" variant was analysed, incorporating a hypothetical 25% lower sensitivity than in the Dutch model. Positive predictive values, however, were assumed to be the same as in the baseline scenario.

Values for sensitivity as well as referral rates and predictive values of the three scenarios are shown in table 1.

For all scenarios the German population and epidemiology are considered. All results/computations were made for a screening programme starting in 1994, with a build up period of five years, and running for 27 years. Effects

and costs which occurred during, as well as after, this period were included. From a public health perspective we have chosen to consider the effects in the total German female population (all ages). Attendance rates, steeply decreasing from 65% to under 30% in the age group 50–69, were based on the experience in the general screening programme in Germany,[12 22] and on rates obtained from a recent telephone survey, since data on attendance rates in the DMS study were not available.

All costs due to screening, diagnosis, and treatment of breast cancer were considered,[11 12] and are presented in Deutsche mark (DM). 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.[23] This PPP was 1.19, which means that health care in Germany is 19% more expensive.

Results

MODEL BASED ANALYSIS OF EMPIRICAL SCREENING DATA: ASSESSMENT OF EXPECTED SENSITIVITY FOR A SCREENING PROGRAMME IN GERMANY

This first part of the Results section describes the model based analysis of the pilot projects in Aurich and Braunschweig, with which the expected sensitivity for a screening programme in Germany (baseline scenario) has been estimated.

In table 2, the observed numbers of detected breast cancers and detection rates of the DMS

Table 2 Number of observed and model expected breast cancers detected in the DMS study (and detection rates per 1000 screens)*

	No of screens	No of detected tumours		
		Observed	Expected†	Expected‡
1st screen	27 099	91 (3.4)	104 (3.8)	93 (3.4)
2nd screen	8 188	14 (1.7)	14 (1.7)	14 (1.7)
Total	35 287	105 (3.0)	118 (3.3)	107 (3.0)

* Women screened aged 35–84, average interval 1.1 year.
† Assuming sensitivity as experienced in The Netherlands (see table 1).
‡ Sensitivity for all stages lowered by approximately 12% (see table 1).

pilot projects are displayed. In the age group 35–84, a total of 105 tumours was detected (of which 11 were after a six months early recall). The *MISCAN* model, adapted for German underlying breast cancer incidence and assuming sensitivity as experienced in The Netherlands, predicted 118 tumours (given the specific ages of the women screened and an observed average screening interval of 1.1 year in these projects). Hence, the observed detection rates in the DMS study were lower than expected, although not statistically significant.

The same analysis was carried out assuming an approximately 12% lower sensitivity (all stages) than experienced in The Netherlands, leading to a better fit (107 tumours predicted) with the observed data. These results were consistent in the major age categories (under 50, 50–69, and over 70). In the age group 50–69, a total of 70 screen detected tumours was observed. The model with Dutch sensitivity resulted in 78 predicted tumours, whereas the 12% lower sensitivity variant predicted 71 tumours, which came closer to the observed number. It seems that the detection rates in the second round have hardly been influenced by assuming a lower sensitivity in the model, presumably since some of the extra number of missed cancers at the prevalence screens will be detected at the subsequent screens now, and probably also due to the short re-screen interval.

Based on these results, together with the present discussion on quality of screening in Germany and DMS data,[6] the expected sensitivity in the baseline scenario was assumed to be 12% less than that experienced in The Netherlands.

In the following parts of the Results section, this scenario will be evaluated and compared with the other two scenarios.

IMPACT OF SENSITIVITY ON THE NUMBER OF BREAST CANCERS DIAGNOSED

Introducing a screening programme will result in an increase in the total number of newly diagnosed breast cancer cases, especially in the beginning of the programme when the prevalence of screen detectable cases is relatively high.[24] A five year build up period in the programme will cause this increase in annual incidence to be spread more evenly over time.

In the baseline scenario, biennial screening of women aged 50 to 69 in Germany is expected to lead to increases in the total annual incidences (all ages) of 6%, 10%, 6%, and 3% in the 2nd, 4th, 6th, and 8th years of the programme respectively. The maximum increase is to be expected in the 4th year, when 3900 more women will be diagnosed with breast cancer than in a situation without screening. Different levels of sensitivity are expected to result in similar sequences of increases in annual incidence, with a maximum of 9% in the low quality scenario compared with 11% in the high quality scenario. The present screening programme in The Netherlands is expected to lead to a maximum increase of 15%, mainly because of a higher screening attendance.

In all three scenarios the stage distribution of screen detected breast cancers is far more favourable than that of clinically diagnosed breast cancers in the same age group. The impact on the overall stage distribution (all ages), however, is relatively small, since only 17% of all breast cancers are expected to be detected by screening. In the present (no screening) situation, 62% of the breast cancers are invasive and larger than 20 mm in size. With a screening programme (baseline scenario) only about 23% of the screen detected cancers will be larger than 20 mm, while among all diagnosed cancers this stage is still expected to comprise 55%.

IMPACT OF SENSITIVITY ON BREAST CANCER MORTALITY REDUCTION

The level of breast cancer mortality reduction due to screening is directly dependent on the benefit assumed after screen detection. This estimated benefit, or improvement in prognosis, is based on the 29% mortality reduction reported in the five Swedish trials.[10] However, if we look at breast cancer mortality at population level (all ages), the mortality reduction due to a screening programme will be less than that. In the baseline scenario, breast cancer mortality in the total German female population (all ages) is ultimately expected to be reduced by 11%. A total of almost 17 000 women is expected to die of breast cancer each year, which is 2100 women fewer than in the situation without mass screening.

A higher sensitivity will lead to more (early) detected breast cancers with a better prognosis, hence more breast cancer deaths will be prevented. Increasing sensitivity in the German programme from the intermediate to a high level would eventually result in an additional 200 breast cancer deaths prevented per year (a 12% reduction compared with a no screening situation). In figure 1 the effect on total breast cancer mortality is shown for different levels of sensitivity achieved in a screening programme.
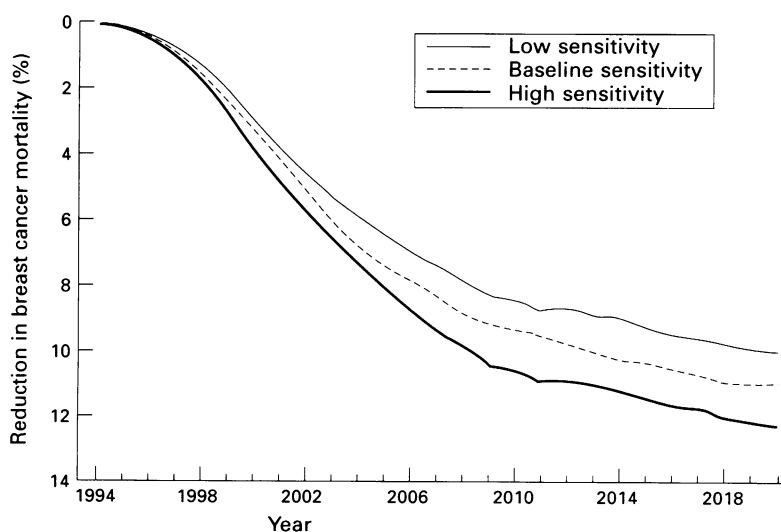
Figure 1   Breast cancer mortality reduction in the total German female population (all ages) due to a screening programme for women aged 50–69 (two year screening interval) associated with different assumptions on sensitivity.

Table 3 Referrals and positive predictive values (PPV) in a screening programme in Germany for women aged 50–69 (2 year interval) in relation to two different assumptions on quality* – baseline and high quality scenarios

	Baseline		High	
	1998†	2008‡	1998†	2008‡
No of screens (× 1000)	2 212	2 421	2 212	2 428
No of positive screen results	73 974	58 931	17 063	13 702
No of women referred for biopsy	17 531	13 977	13 090	10 676
Biopsy diagnosis:				
Malignant	6 966	6 481	7 714	7 075
Benign	10 565	7 496	5 376	3 601
PPV of positive screen result (%)	9.4	11.0	45.2	51.6
PPV of referral for biopsy (%)	39.7	46.4	58.9	66.3

* As defined in table 1; referral rates and predictive values in a screening programme are results of a mixture of 1st and subsequent screens.
† End of build up phase.
‡ Steady state.

Table 4 Costs* (× 10⁶ DM) and effects of a screening programme in Germany for women aged 50–69 (2 year interval) in relation to three different assumptions on quality – baseline, high, and low quality scenarions† (5% discount rate)

	Baseline	High	Low
Breast cancer deaths prevented	19 627	21 843	17 510
Life years gained	206 525	230 079	182 919
Costs of screening	4 005	4 002	4 008
Costs of false positive biopsies	576	232	500
Costs of advanced disease	−889	−989	−793
Other costs	199	215	175
Total extra costs	3 891	3 460	3 890
Costs (DM) per life year gained	18 842	15 037	21 264

* Cost amounts are expected differences between situation with and without screening.
† Values as in table 1.

IMPACT OF SENSITIVITY AND SPECIFICITY ON CONSEQUENCES FOR THE WOMEN

A screening programme will lead to a decrease in the number of advanced breast cancers. Consequently, women who, in a no screening situation, would have undergone mastectomy, have a chance of being treated less radically because of earlier detection. Screening will therefore cause a shift from mastectomies to breast conserving therapies. In the steady state situation of a screening programme, almost 2000 (14%) more breast conserving therapies are expected to be performed each year than in the situation without mass screening. Part of this favourable increase, however, is due to the extra incidence caused by screening, ie cases that are detected earlier as well as latent cases that otherwise would not have been diagnosed.[24] This extra annual incidence in Germany is estimated to be 1.8%, or 750 cases each year (of which approximately 40% otherwise would not have been diagnosed). Increasing sensitivity in the programme is expected to result in 40 additional detected breast cancers; the number of breast conserving therapies is expected to increase by 160 each year.

A higher specificity will have favourable consequences for the women, since fewer women will be exposed to the anxiety of undergoing additional examination or even being referred for biopsy without breast cancer being diagnosed (table 3). In the baseline scenario almost 74 000 women will be subjected to additional examinations after a positive screen (3.3% of all screened women) in the year 1998, at the end of the build up phase. With a high specificity only 0.8% would have to be referred that year, which is 57 000 fewer women. Ten years later these figures will be more favourable since, by then, only a small portion of the screening examinations will be first screens, which have relatively high referral rates compared with subsequent screens (table 1). The increase of specificity is expected to cause the yearly number of women referred for biopsy without breast cancer being diagnosed to be halved: 10 600 compared with 5400 in 1998; 7500 compared with 3600 in a steady state situation.

EXTRA COSTS AND COST EFFECTIVENESS

Costs and effects of the three scenarios for an age 50–69 years screening programme with a two year interval are displayed in table 4 (at a 5% discount rate to establish a time preference). The bulk of the costs (about DM 4.0 billion) is directly attributable to the screening programme as such (overhead costs, mammography). The main costs incurred by a programme are due to biopsies performed on women who are apparently without breast cancer (DM 576 million in the baseline scenario). The main cost savings are due to treating fewer women with advanced disease (DM 889 million).[25] The entry "Other costs" (DM 199 million) includes costs of diagnostic assessment in the programme, costs due to the increase in primary treatments and follow up, savings on diagnostic assessment outside the programme and savings on adjuvant therapies.

Hence, in the baseline scenario almost DM 3.9 billion (discounted) needs to be spent to gain 207 000 life years (discounted): a cost effectiveness ratio of DM 19 000 per life year gained. A screening programme with high sensitivity and specificity is expected to be less expensive (DM 3.5 billion). The direct screening costs are about the same, but many fewer biopsies are performed on women without breast cancer, and more money is saved on treatment of advanced disease. A high quality programme is also more effective (230 000 life years gained), resulting in a cost effectiveness ratio of DM 15 000 per life year gained. A low sensitivity screening programme leads to a ratio of more than DM 21 000 per life year gained.

In figure 2 it can be seen that varying the sensitivity has a large impact on the effects, but almost no impact on the total difference in costs. A higher sensitivity, while maintaining the level of positive predictive values, leads to more savings in terms of treatment of advanced disease, but also to extra costs incurred by a larger number of women referred, treated, and followed up. Consequently, a higher specificity can have a strong impact on costs, but will not increase the number of life years gained.

Discussion

It is of course difficult to make assumptions on the quality of a future population based screening programme in Germany that are based only on the available screening data from the German pilot projects. Using the *MISCAN*
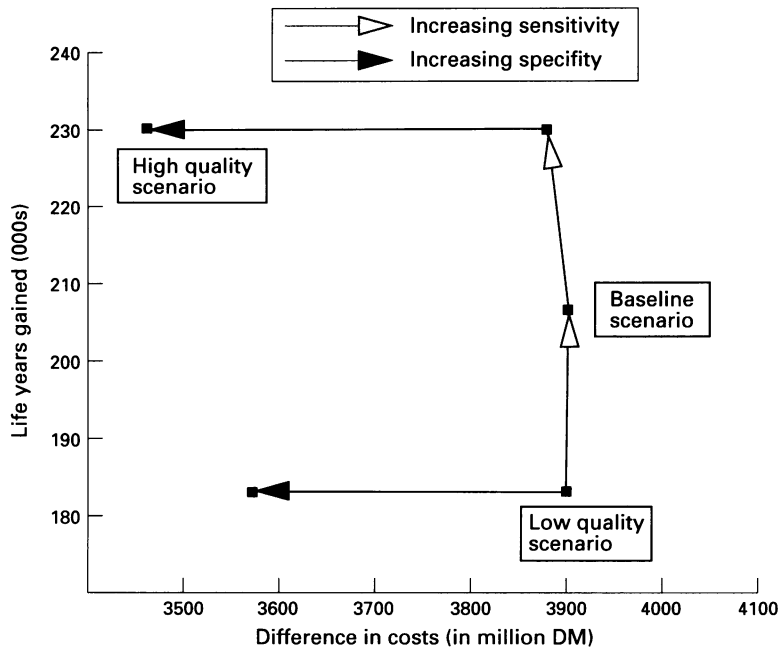
Figure 2 Impact of changes in sensitivity and specificity of a screening programme in Germany for women aged 50–69 (two year interval) on the total number of life years gained and on the total extra costs; (5% discount rate).

model we can analyse screening data by making estimates for age specific detection rates, while taking into account the underlying incidence and clinical stage distribution of breast cancer and by comparing these with the observed rates. Because of the decentralised health care system in Germany, there is reason to believe that the screening quality in Germany could be improved. The results of the model, an expected 12% sensitivity difference compared with The Netherlands, support this opinion, although there were no data on interval cancers. To account for the uncertainty in assessing the extent of this loss in quality, we have defined two other sets of quality assumptions. The range of differences in sensitivity assumed for the three German scenarios in this study was somewhat arbitrary, based on expert opinion, but seems plausible, considering the outcomes of the model. However, larger differences in quality between European countries might exist.

The various assumptions made about several factors are, of course, open to discussion. For instance, one could question how representative the Saarland registry data on breast cancer incidence rates are. National data on these incidence rates are lacking, but national breast cancer mortality rates closely resemble those of the Saarland region. Furthermore, regional differences in incidence were assumed to be negligible on the basis of the regional mortality data for breast cancer.[21] Therefore, it is likely that breast cancer incidences are comparable, unless major differences exist in survival between Saarland and all of Germany. Because of an absence of data, this question cannot be addressed. The extrapolation is not likely to have significantly influenced the results.

The attendance rates used in the analyses are low compared to attendance in other European countries. It is not unlikely that attendance in a German screening programme with a personal invitation system would be higher than estimated here. However, it is impossible to predict attendance in this situation since the health care system is different. Since the attendance rate has a large impact on reducing breast cancer mortality, more attention to this aspect is needed.

The assumption that the clinical stage distribution in Germany is equal to that in The Netherlands is rather fragile. Diversity of data sources makes it hard to compare accurately, but the scarce German data give no reason to assume a different stage distribution. There are limited data on the baseline level of screening, and the extent to which this has caused a shift in the clinical stage distribution prior to screening can be assumed to be negligible.[26] A screening programme may have an eliciting effect on spontaneous screening in women not (yet) eligible for screening. We have not considered this effect, since it is difficult to estimate the extent of this spontaneous screening, as well as what effect this has on mortality.

It is clear that the importance of sensitivity and specificity in a mass screening programme for breast cancer should not be underestimated. The favourable and unfavourable effects for the women involved are highly dependent on the levels of sensitivity and specificity achieved. Increasing sensitivity will lead to more detected cases in an early stage, and thus to a more effective programme with a higher number of prevented breast cancer deaths. Also, with a higher sensitivity more mastectomies can be prevented in favour of breast conserving therapies. However, if specificity in the screening programme is low, these favourable effects are only achieved by referring a considerable number of women for additional examinations and biopsies, thereby making the programme more expensive.

In Germany, introducing a programme with the baseline sensitivity and specificity assumptions is expected to result in a total difference in costs of DM 3.9 billion (5% discount rate). A high quality programme is expected to lead to only DM 3.5 billion (discounted). Therefore, 11% of the total costs of a programme can be invested in quality assurance (in technique and interpretation), provided that the higher levels of sensitivity and specificity can be achieved. In this way, an equally expensive programme will gain an additional 24 000 life years (5% discounted) and will prevent an additional 200 cases of breast cancer death per year.

If DM 19 000 per life year gained is considered to be cost effective, then in order to gain as many life years as in a high quality programme an amount of DM 4.3 billion (DM 18 842 × 230 079 life years, table 4) is considered good value for money. Since the costs of a high quality programme is estimated to be only DM 3.5 billion, an amount of almost DM 0.9 billion (20% of the total costs of the programme) can be invested in quality

assurance while maintaining a cost effectiveness of DM 19 000 per life year gained. Of course, the effects and costs (including quality assurance measures) have to be submitted to continuous evaluation during the programme. In The Netherlands, at present, approximately 10% of the cost of screening is allocated to quality assurance and evaluation of the national programme.

Under the assumption of low quality, the cost effectiveness ratio of a programme is expected to be more than DM 21 000 per life year gained. This amount might be underestimated since the lower sensitivity has not been accompanied by worse specificity. It is very likely that in practice extra costs will be induced by an even larger number of referred women. Furthermore, in this analysis the decrease in sensitivity was assumed to be equally distributed over the tumour stages, while in practice it is more likely that the smaller tumours will be missed relatively more frequently. Thus the effect of a screening programme might be reduced even more than estimated in our calculations.

The cost effectiveness of a low quality screening programme may not be unacceptably high when compared with, for instance, the ratio for cervical cancer screening.[27] However, this does not in any way imply that it is an acceptable goal for quality in a national programme. Nevertheless, this analysis shows that for all countries considering breast cancer screening, much effort should be put in improving sensitivity and specificity. It is very likely that allocating up to 20% of the total costs of a screening programme to strict quality assurance measures and evaluation will be cost effective.

1 Chamberlain J, Moss SM, Kirkpatrick AE, Michell M, Johns L. National Health Service breast screening programme results for 1991–2. *BMJ* 1993;**307**:353–6.
2 de Koning HJ, Fracheboud MD, Boer R, *et al*. Nation-wide breast cancer screening in the Netherlands; support for breast-cancer mortality reduction. *Int J Cancer* 1995;**60**:777–80.
3 Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer. *J Natl Cancer Inst* 1993;**85**:1644–56.
4 Werneke U. Threat to German breast cancer screening plan [letter]. *Lancet* 1995;**345**:1167.

5 Robra BP. Threat to German cancer-screening plan. *Lancet* 1995;**345**:1574.
6 Frischbier HJ, Hoeffken W, Robra BP. *Mammographie in der Krebsfruherkennung; Qualitätssicherung und Akzeptanz*. Stuttgart; Ferdinand Enke Verlag, 1994.
7 van Ineveld BM, van Oortmarssen GJ, de Koning HJ, Boer R, van der Maas PJ. How cost-effective is breast cancer screening in different EC countries? *Eur J Cancer* 1993; **12**:1663–8.
8 van Oortmarssen GJ, Habbema JD, van der Maas PJ, *et al*. A model for breast cancer screening. *Cancer* 1990;**66**: 1601–12.
9 de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PMM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer screening trials. *J Nat Cancer Inst* 1995;**87**:1217–1223.
10 Nyström L, Rutqvist LE, Wall S, *et al*. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;**341**:973–8.
11 de Koning HJ, van Ineveld BM, van Oortmarssen GJ, *et al*. Breast cancer screening and cost-effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer* 1991;**49**:531–37.
12 Beemsterboer PM, de Koning HJ, Warmerdam PG, *et al*. Prediction of the effects and costs of breast-cancer screening in Germany. *Int J Cancer* 1994;**58**:623–28.
13 Carter R, Glasziou P, van Oortmarssen G, *et al*. Cost-effectiveness of mammographic screening in Australia. *Aust J Public Health* 1993;**17**:42–50.
14 Paci E, Boer R, Zappa M, *et al*. A model-based prediction of the impact on reduction in mortality by a breast cancer screening programme in the city of Florence, Italy. *Eur J Cancer* 1995;**3**:348–53.
15 Statistisches Bundesamt. *Todesursachen 1990, Gesundheitswesen Fachserie 12 (Reihe 4)*. Stuttgart; Metzler Poeschel, 1991.
16 Statistisches Bundesamt. *Statistisches Jahrbuch 1991 für das vereinte Deutschland*. Stuttgart; Metzler Poeschel, 1992.
17 Leonhardt A. (Significance of mammography study in the detection of 2,511 breast cancers 1977 to 1987). (In German). *Geburtshilfe Frauenheilkd* 1988;**48**:72–77.
18 Paterok EM, Rosenthal H, Richter S, Säbel M. (Breast cancer: trends from 1964 to 1990. Results of a long-term study). (In German). *Röntgenpraxis* 1992;**45**:325–29.
19 Netherlands Cancer Registry. *Incidence of cancer in the Netherlands 1989*. Utrecht: SIG (Health Care Information), 1992.
20 Saarland in Zahlen. *Morbidität und Mortalität an Bösartigen Neubildungen im Saarland 1986, 1987, 1988,1989, Jahresbericht des Saarlandischen Krebsregisters*. Saarbrücken: Sonderhefte Statistisches Amt des Saarlandes, 1987, 1988, 1989, 1990.
21 Smans M, Muir CS, Boyle P. *Atlas of cancer mortality in the European Economic Community (IARC Scientific Publication 107)*. Oxford: Oxford University Press, 1992.
22 Berghof B, Robra B-P. *Krankheitsfrüherkennung Krebs Frauen und Männer. Aufbereitung und Interpretation der Untersuchungsergebnisse aus den gesetzlichen Früherkennungsmassnahmen 1985–1986*. Cologne: Deutscher Arzte-Verlag, 1988:42–49.
23 Organization for Economic Co-operation and Development. *Health data file CREDES*. Paris: OECD, 1991.
24 Boer R, Warmerdam P, de Koning H, van Oortmarssen G. Extra incidence caused by mammographic screening. Letter. *Lancet* 1994;**343**:979.
25 de Koning HJ, van Ineveld BM, de Haes JC, van Oortmarssen GJ, Klijn JG, van der Maas PJ. Advanced breast cancer and its prevention by screening. *Br J Cancer* 1992; **65**:950–55.
26 Robra B-P. *Evaluation des deutschen Krebsfrüherkennungsprogramms; Struktur, Prozess, Resultate und Entwicklungsperspektiven*. Bremerhaven: Wirtschaftverlag NW, Verlag für neue Wissenschaft, 1993.
27 Koopmanschap MA, Lubbe KT, van Oortmarssen GJ, van Agt HM, van Ballegooijen M, Habbema JK. Economic aspects of cervical cancer screening. *Soc Sci Med* 1990; **30**:1081–87.