Based on the results of the two pilot mammographic screening projects in Nijmegen and Utrecht, a population-based programme was set up in the Netherlands in 1989. The main characteristics of the programme are the centralised organisation including centralised technical and medical quality control and audit, the 2-year interval between examinations, and the eligible age of 50–74. The National Expert and Training Centre in Nijmegen is responsible for the technical and medical quality control and the National Evaluation Team in Rotterdam for the data collection evaluation and annual reporting of the performance results. After 15 years, the Dutch breast cancer screening programme can be regarded as an effective health care intervention at reasonable costs.

A new challenge for the Dutch screening programme is the planned transformation of analogue to digital screening within the next 2 years. It will be a major logistic operation with many, at this moment still unforeseen obstacles.
intervention at reasonable costs. In this article, we report on the programme’s history and design as well as its features and achievements.

History and Design

In 1975, observational studies on mammographic screening were started in Nijmegen and Utrecht. These studies were initiated by the encouraging early results of the American so-called HIP study [1] involving women covered by the Health Insurance Plan (HIP) of Greater New York. Started in 1963, the HIP trial was the first randomised controlled trial (RCT) designed to examine the efficacy of screening for breast cancer. The non-randomised so-called case control studies in Nijmegen and Utrecht showed that the relative risk of dying from breast cancer among women ever screened compared with women never screened was 0.48 and 0.30, respectively [2, 3]. Based on these results, it was recommended by the Dutch Health Council and the National Council for Public Health in 1987 to establish a national breast cancer screening programme. The National Expert and Training Centre for Breast Cancer Screening (Dutch abbreviation LRCB) was set up in Nijmegen. Its responsibilities include training radiographers, radiologists and pathologists who are participating in the screening programme and monitoring the technical and medical quality of the programme. The professional associations agreed that targeted training and continuing education are a necessary precondition for qualitatively acceptable screening. Only LRCB-certified professionals are allowed to participate in the screening.

Before the political decision was taken, a study was conducted to assess the likely benefits and risks, the consequences for the health care system, and the financial and staffing implications of a national screening programme. The conclusion was that, if women between the ages of 50 and 70 were screened for breast cancer every other year in the context of a national programme, the effects would on balance be positive, and that the benefits of such a programme would justify the costs [4]. On the basis of RCTs conducted in Sweden (Kopparberg, Östergötland, Malmö), it was estimated that screening would reduce breast-cancer mortality among women between the ages of 50 and 70 by 27% after 15 years of screening in comparison to the mortality data in the pre-screening era. This result can be translated into 700 fewer women dying of breast cancer each year than would otherwise be the case.

The programme for women aged 50–69 years has been gradually implemented between 1989 and 1997. Between 1998 and 2000, the screening was extended to the age group 70–75. Screening is organised regionally. There are 9 screening regions which coincide with the regions covered by the Comprehensive Cancer Centres. The 9 regional organisations operate a total of 63 screening units (SU) of which 54 are mobile to improve attendance. For technical quality assurance, each SU exposes and develops a test film of a phantom every day, just before starting to screen. These data will then be analysed by the physics group of LRCB, and immediate actions will follow if correction is deemed necessary. The personal data of the eligible women are provided by the municipal population register which is fully computerised. Every 2 years, women get a personal invitation letter with a fixed appointment for a screen examination in one of the SUs. Non-responding women are issued a reminder after 2–3 months. At the initial screen, 2-view mammography is performed; at subsequent screens, medio-lateral-oblique (MLO) views are standard. Additional cranio-caudal (CC) views are performed only on indication which include dense glandular tissue, postoperative changes, implants and whenever abnormalities are suspected by the radiographer. At present, an additional CC view will be performed in about 40% of cases at subsequent examination. Films are processed on site in the mobile units so that additional views and/or repeats for technical reasons can be made immediately. X-rays of previous examinations are also available for the radiographer for comparison. There is no radiologist present at the SU. Physical examination is not part of the screening.

X-rays are transported daily to the 27 central reading units (CU) where they are read independently by 2 radiologists. A decision on recall for further assessment is made usually on the basis of consensus. In some units, the second reader decides whether the woman will be recalled. All examined women receive the results in writing; in the event of a positive result (recall for further examination), the general practitioner is informed in advance.

The above described centralised organisation of the Dutch screening is often referred to as one of the most efficiently set up national mammographic screening project in Europe. However, little is said about the organisation of the further assessment of the recalled women based on a suspicious breast lesion. This part of the screening has been improperly organised in the Dutch programme. Assessment of recalled women should be an integral part of a screening programme and should be carried out preferably by the screening radiologists. Assessment units should be incorporated in the regional screening organisations, thereby facilitating a high level of diagnostic performance and a direct feedback of the results to the screening radiologist. In the Dutch system, the radiology departments of the hospitals act as assessment units. The general practitioner and the woman decide jointly which hospital the woman will be referred to for assessment. The screening radiologists are often not involved in the assessment. The communication between the screening and hospital radiologist may occasionally be suboptimal which may lead to diagnostic errors. Such errors are usually categorised as ‘doctor’s delay’ when the detection and treatment of the malignant lesion is delayed due to inadequate assessment procedure. Presently, plans are unfolding to repair this organisational shortcoming. These plans include the set up of assessment
centres within the screening organisations and the set up of specialised breast units with specially trained multidisciplinary teams [5]. Evaluations of the process and its effects is undertaken by the National Evaluation Team for Breast Cancer Screening (Dutch abbreviation LETB) at the Erasmus University in Rotterdam. The LETB publishes an annual report on the results of the screening programme based on tabulated data on invitations, attendance, screen examinations, referrals, assessment and screen-detected breast cancers including tumour stage. Data on interval cancers, breast cancer incidence and therapy are obtained after linking the regional files of screened women to the files of the regional cancer registry. Due to an inevitable delay in the cancer registry and because of the screening interval of 2 years, records of women screened in a certain calendar year cannot be linked to cancer registry records earlier than in the 3rd year after the screening round. An alternative method is to collect the data of interval cancers within a shorter period through the pathology reports of the centralised pathology data system.

**Screening Results 1990–1999**

Between 1990 and 1999, 5.7 million invitations were issued and 4.5 million examinations performed (a participation rate of 79%). 45,600 women have been referred for further assessment (10.1 per 1,000) and 21,500 breast cancers detected (4.7 per 1,000) [6]. Table 1 shows the tumour stage distribution of screen-detected cancers for the same period. The programme functioned broadly as anticipated [7]. In some respects, things have actually worked out better than expected; the 79% participation rate was well above the forecast 70%, for example. However, there have also been less favourable results. For the period of 1990–1995, the detection rate for subsequent screening examinations was lower than expected (observed rate 3.4%; expected rate 4.3%), the interval cancer rate was higher than expected (observed rate 0.99%; expected rate 0.96%), and the stage distribution of screen-detected cancers was less favourable than expected [8, 9]. These results were suggested to be attributable to the very low recall rate (<1%) in the subsequent screening examinations of the Dutch programme in comparison with that of screening programmes in other countries, which have recall rates of 2–5% [10]. It is anticipated that an increase in the recall rate should improve the breast cancer detection rate by reducing the number of both interval and late screen-detected cancers (i.e., cancers that could be detected earlier, collectively referred to as ‘missed cancers’). However, an increase in the recall rate would also increase the false-positive rate, resulting in undesirable consequences, such as excess further assessment, unnecessary surgery, patient anxiety and additional financial costs. Recent studies of different population-based screening programmes have shown that the relationship(s) among recall rate, detection rate and false-positive rate is not straightforward [11]. That is, a high recall rate will not necessarily result in a high detection rate. Moreover, it is not possible to define an optimal recall rate because desirable screening performance is influenced by several factors, including health policy issues, that differ from country to country. Smith-Bindman et al. [11] reported recently that although recall rates are twice as high in the United States as they are in the United Kingdom, cancer detection rates are similar.

**The Dutch Review Study of Late Detected Breast Cancers**

We have set up a retrospective review study [12] to see whether the low recall rate of the Dutch programme could have indeed been the reason of the higher number of late detected interval and screen-detected cancers. The study was based on a review of a series of interval and screen-detected breast cancers from the Dutch screening programme. The low recall rate of the Dutch screening programme (<1%) provided a unique opportunity for such a study. It allowed for the collection of a random selection of originally screen-negative, pre-diagnostic mammograms from non-recalled cancer case patients whose mammograms may have contained subtle signs of possible cancer. In other screening programs with lower thresholds for calling a mammogram suspicious, these women could have been recalled for additional testing. By performing a blinded review of screening mammograms from the Dutch programme, we aimed to determine the effect of changes in recall rate on earlier detection of cancers that were not recalled, in relation to false-positive rates. Our panel of 15 screening radiologists included 10 radiologists from the

### Table 1. Tumour size distribution of all screen-detected breast cancers. Lymph node involvement and distant metastases of invasive cancers, 1990–1999

<table>
<thead>
<tr>
<th>Screen</th>
<th>Breast cancers, n</th>
<th>Tumour size distribution, %</th>
<th>Node involvement, %</th>
<th>Distant metastases (M1), %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Tis (DCIS)</td>
<td>T1a</td>
<td>T1b</td>
</tr>
<tr>
<td>Initial</td>
<td>10,465</td>
<td>15</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Subsequent</td>
<td>9,737</td>
<td>14</td>
<td>4</td>
<td>22</td>
</tr>
</tbody>
</table>

DCIS = Ductal carcinoma in situ.
Netherlands and 5 radiologists who were from countries whose screening mammography programmes had much higher average recall rates than the Dutch screening programme. We anticipated that the 5 non-Dutch radiologists might pick up more suspicious lesions with subtle radiological signs than the 10 Dutch radiologists. Hence, we used individual and panel-based review results to calculate detection sensitivity rates at various levels of false-positive recall rates. These data further allowed us to estimate the effect of different recall rates on earlier detection of both interval and screen-detected cancers. Mean detection sensitivity was estimated based on the sensitivity data at different false-positive threshold values for all 15 radiologists and for the 10 Dutch and 5 non-Dutch radiologists separately. Based on these data, localisation-response receiver operating characteristics (LROC) curves were constructed for each individual radiologist and for all radiologists together as average (fig. 1). The mean detection sensitivity for all radiologists at a false-positive value of 0.5% (i.e., a recall of approximately 1 in 250 of the control subjects in this study set) was 12.5% (95% confidence interval (CI) = 10.8–14.4%). At this detection sensitivity, 31 of the 245 women with originally screen-negative breast cancer would have been recalled based on their pre-diagnostic mammograms. With false-positive rates of 4 and 8%, the mean detection sensitivities increased to 21.5% (95% CI = 19.2–23.8%) and 27.3% (95% CI = 24.5–30.1%), respectively.

The reviewed mammograms of the cancer cases in this study were from prior screening examinations, which were originally categorised as screen-negative. Thus, the observer performance we measured is related to non-recalled cancer cases, which may explain the low detection sensitivity of 21–27% in our study in comparison with the detection sensitivity in routine screening practices (70–80%). Similar detection sensitivity (27%) results were reported recently by Burhenne et al. [13] in a multicenter U.S. review study. The total increase in the number of earlier detected cancers consists of the number of interval cancers and the number of late screen-detected cancers recognised on the pre-diagnostic index mammogram. The major effect of earlier detection is the potential to detect these cancers at a less advanced and more curable stage. The results of a screening review study are highly dependent on the mammography reading skills of the review panel. On average, the 15 radiologists had a median of 10 years experience in mammography screening and read more than 6,000 mammograms per year. Nevertheless, the radiologists operated with different recall policies in their respective screening programs in their own countries. From the beginning of mammography screening in the Netherlands (in 1975), the potential for a large number of false-positive findings was a major concern to the officials of the screening programme because of the possible consequences of unnecessary surgery, patient anxiety, negative influence on participation, and costs. Therefore, in their daily practice, the 10 Dutch radiologists were operating at a relatively low recall rate (approximately 1%), whereas the 5 non-Dutch radiologists maintained a recall level 2–5 times higher. In addition, there could have been a performance difference between the 2 groups of radiologists due to the longer mean screening experience of the non-Dutch radiologists. Although the average detection sensitivity of the non-Dutch radiologists was somewhat higher than that of the Dutch radiologists, the difference was not statistically significant (p = 0.29; fig. 1).

In conclusion, the study showed that a delicate balance exists between recall, detection and false-positive rates. At a low recall rate (i.e., < 1%), when only women with more or less obvious mammographic abnormalities are recalled, a number of cancer cases will surface as interval cancers or late screen-detected cancers at a more advanced tumour stage. By lowering the threshold for recall by focusing on more subtle mammographic abnormalities, a substantial number of cancers could be detected earlier. The effect of increasing the recall rate is most obvious between 1 and 4%. At still higher recall rates, cancer detection rates level off, with very low positive predictive values of recall as an expression of gradually increasing numbers of false-positives per one single earlier detected cancer. To benefit most from a higher recall rate, radiologists should be able to identify the cancers with subtle mammographic features which are often based on temporal changes.
Site Visits of the National Expert and Training Centre for Breast Cancer Screening

One of the instruments of quality control of the medical performance of the Dutch mammographic screening is the regular audit visit of the professional staff of the National Reference Centre to the 27 CUs of the regional screening organisations [14]. The frequency of the visits is once every 3 years. At these site visits, the performance of the screening radiologists and radiographers is monitored and the technical and positioning quality of the mammograms are checked. Performance data of 2 screening rounds of a 4-year period are presented and compared to desirable target values and to the mean national performance values. At the site, 80 consecutive interval cancer cases and 40 stage-II screen-detected cancer cases are reviewed and subdivided into 3 categories: 1, no suspicious signs on the screening mammograms prior to the diagnostic mammogram, 2, so-called minimal signs although not prompting a recall, 3, significant signs for which the women would have been referred; commonly called ‘missed’ cancers. At review, 2 previous screening mammography sets are available for comparison, called the prior and the current mammograms, prior to the diagnostic mammograms. Review results of the different CUs are compared and published annually. These retrospective review studies of various series of interval cancers showed that in about 50% of cases no abnormality is seen on the last screening mammogram before diagnosis. In about 25–30%, the abnormality can be classified as minimal sign and in 20–25% as significant sign.

Much emphasis has been put on the quality of the positioning technique. Poor positioning is one of the reasons of missed interval cancers. During the site visits, all mammograms made on one of the previous days are reviewed for positioning. Percentages for optimal, acceptable and unacceptable positioning are recorded for each SU. Target value for optimal positioning according to national guidelines is 90%. The reviews disclosed that areas where interval cancers are most likely to be overlooked due to poor positioning are the infra-mammary angle, the high axillary tail and the sub-areolar area.

Main Results of the Population-Based Breast Cancer Screening Programme in the Netherlands

After 15 years of nation-wide breast cancer screening, breast cancer mortality in women aged 55–74 was 25.5% lower in 2004 than before the start of the screening programme [15]. As cancer mortality can only be analysed after a certain period of time after the diagnosis (lag time), the above data on mortality decrease concerns women aged 50–69 when screened. Since 2002, breast cancer mortality also declines significantly in women aged 75–84 relating to women aged 70–74 at screening.

The total cost of the programme was 44.4 million Euro in 2004, including training, audits, evaluation and all management costs, equal to 49.91 Euro per screen examination. When adjusted for the increasing number of screen examinations, the creation of financial reserves for the forthcoming digitalising and for deflation, total costs remain rather stable since 2002, and costs per screen even show a slight decrease since 1997 (fig. 4).

The above mortality decrease was to be achieved only with a high attendance rate. In 2003, the attendance was 80.8% (fig. 5).
All ages contributed to this highest attendance rate ever. Even for women aged 70 and over, the attendance increased from 65% in 1998 to 74.5% in 2003. An important result is, in addition, that the participation for the first screens aged 49–50 increased as well to 80% in 2003. 94% of women attending one examination will also show up for the second one.

In 2003, 866,000 screen examinations were performed, 4% more than in 2002. The referral rate remains stable at 13–14 per 1,000 screens since 2002, after a gradual increase as of 1996 partly due to the recommendations of the LRCB site visits and the review study of late detected interval cancers (fig. 6). In 2003, 4.4 breast cancers were detected per 1,000 subsequent screens, a 30% higher detection rate than in 1994.

Besides a temporary strong increase of early stage breast cancers, screening leads also to a significant decrease of advanced diseases. The number of stage II+ cancers of women 50–69 years of age decreased by 12% in the period of 1989–1997. This significant reduction of advanced cancers preceded by two years the mortality decrease by the same amount indicating that the mortality decrease can largely be attributed to the screening [17]. There has been a continuous debate on the proportion of shared contribution of the systemic adjuvant therapy and mammographic screening to mortality decrease of breast cancer. In 1999, one third of all breast cancers were detected by screening. 64% of screen-detected tumours were stage 0–I cancers, in comparison to 35% of the non-screen-detected cancers. As a consequence, the proportion of breast conserving treatment increased from 26% in 1990 to 41% in 1999. In 1999, almost half of the targeted women with breast cancer had breast conserving surgery.

Despite an increasing screen detection rate, the interval cancer rate was somewhat higher than expected (1.1 per 1,000 woman-years of follow-up in 1999; expected 0.96). This may be explained by an improved registration of interval cancers, or a lower detection rate of some screening regions. One performance measure is the relationship between the number of screen-detected cancers and interval cancers (S:I ratio). The regional differences of the S:I ratio for first screenings was 4:1 to 3:1, and for subsequent screenings 2:1 to 1:1.5. Target values are respectively > 3:1 and > 2:1. Low regional detection rates tend to be followed by higher interval cancer rates.

A new challenge for the Dutch screening project is the planned transformation of analogue to digital screening within the next 2 years. This will be a major logistic operation with many,
at this moment still unforeseen obstacles. The National Expert and Training Centre with its physics group is participating in the coordination of this national project. In 2004, 3 pilot digital screening projects were started. The first preliminary results indicate a 50–100% increase in referral rate and a significantly higher detection rate of ductal carcinoma in situ. The results are presently under analysis.

Conclusion

Based on the results of the 2 pilot mammographic screening projects in Nijmegen and Utrecht, a population-based programme was set up in the Netherlands in 1989. The main characteristics of the programme are the centralised organisation including centralised technical and medical quality control and audit, the 2-year interval between examinations and the eligible age of 50–74 years. The National Expert and Training Centre in Nijmegen is responsible for the technical and medical quality control and the National Evaluation Team in Rotterdam for the data collection evaluation and annual reporting of the performance results.

The 10-year results of the programme were encouraging. Between 1990 and 1999, 5.7 million invitations were issued and 4.5 million examinations performed (a participation rate of 79%), 45,600 women have been referred for further assessment (10.1 per 1,000) and 21,500 breast cancers detected (4.7 per 1,000). However, some results fell short of the expectations based on the microsimulation model of the Rotterdam epidemiology group. The number of interval cancers was somewhat lower than expected and the number of screen-detected cancers somewhat lower. A review study of interval and screen-detected cancers disclosed that the low recall rate (around 1%) was the most likely reason for these shortcomings. The study showed that a delicate balance exists between recall, detection and false-positive rates. At a low recall rate (i.e., around 1%), when only women with more or less obvious mammographic abnormalities are recalled, a number of cancer cases will surface as interval cancers or late screen-detected cancers at a more advanced tumour stage. By lowering the threshold for recall by focusing on more subtle mammographic abnormalities, a substantial number of cancers could be detected earlier. The effect of increasing the recall rate is most obvious between 1 and 4%. At still higher recall rates, cancer detection rates level off, with very low positive predictive values of recall as an expression of gradually increasing numbers of false-positives per one single earlier detected cancer. Based on these study results, the present recommendation of the National Expert Centre is to increase recall rate to 2% or above, by recalling more women with subtle mammographic abnormalities.

After 15 years, nation-wide breast cancer screening, breast cancer mortality in women aged 55–74 was 25.5% lower in 2004 than before the start of the screening programme. As cancer mortality can only be analysed after a certain period of time after the diagnosis (lag time), the above data on mortality decrease concerns women aged 50–69 when screened. Since 2002, breast cancer mortality also declined significantly in women aged 75–84 relating to women aged 70–74 at screening. The total cost of the programme was 44.4 million Euro in 2004, including training, audits, evaluation and all management costs, equal to 49.91 Euro per screen examination. A new challenge for the Dutch screening programme is the planned transformation of analogue to digital screening within the next 2 years. It will be a major logistic operation with many, at this moment still unforeseen obstacles. After 15 years, the Dutch breast cancer screening programme can be regarded as an effective health care intervention at reasonable costs.

References

5 Holland R: Met vroege ontdekking zijn wij er nog niet. [Early detection is just the first step.] Inaugural speech. Catholic University Nijmegen, 2000.