Chapter 6

Cost–effectiveness of population-based breast cancer screening

Cost–effectiveness analysis: What and why ?

As resources are limited, more and more decisions about health care interventions are based on cost–effectiveness analyses, so that health care is spread as equitably and efficiently as possible. In many countries, it has become routine policy to assess the costs of new (promising) health care interventions in relation to their expected benefits before actually implementing them. Interventions have a price, and most do not save total expenditure, but a minor change to an intervention strategy can lower the cost without a substantial loss of benefit, or, on the contrary, more benefit can be expected for similar cost (van den Akker-van Marle *et al.*, 2002). The most accurate instrument for comparing different strategies is a cost–effectiveness analysis, to calculate outcome measures of effectiveness, such as a decrease in mortality and/or morbidity, as economic costs. Usually, a cost–effectiveness analysis is used to compare alternative health care interventions, including current or proposed policy, with no intervention, taking future costs and benefits into account and estimating the cost per life–year gained with the different policies (Brown & Fintor, 1993). Preferably, the costs per life–year gained are adjusted for quality of life, but quality of life is not always measured in practice.

Published analyses

International studies of the cost–effectiveness of breast cancer screening show substantial differences in cost per life–years gained (Brown & Fintor, 1993; de Koning, 2000b). The cost–effectiveness ratio appears to be more favourable for most well-organized screening programmes, often European ones, than for spontaneous screening. The probable explanation is that having a special organization only for screening helps keep costs low, promotes more efficient use of resources, with high attendance of invited women and good quality screening leading to a health benefit. Moreover, as the direct cost for the screening examination is probably the most important single factor in total costs (Brown, 1992), organized large-scale screening may reduce the average cost per screen.

Comparisons of cost–effectiveness ratios between programmes in different countries is complex. Even with similar quality of mammographic screening (e.g. sensitivity), differences are found in almost all the factors that affect both effectiveness and cost. Thus, not only the epidemiology of breast cancer but also the organization and the costs of health care in general may differ. It is therefore surprising that one of the lower (and therefore favourable) estimated cost–effectiveness ratios (2650 euros per year of life gained; 5% discount rate) is seen in Navarra, Spain, where the breast cancer incidence is substantially lower than in northern countries (de Koning, 2000b; Table 58). The Navarra programme had a very high participation rate of invited women (90%), a high breast cancer detection rate, indicating a high-quality programme, and a relatively unfavourable clinical stage distribution of breast tumours before the introduction of screening (van den Akker-van Marle *et al.,* 1997). Conversely, in Germany, the estimated cost–effectiveness ratio was high—9600 euros per life–year gained—which must be attributed to the decentralized health care system, the lack of centralized screening settings and of personal invitations to screening and lower breast cancer incidence and mortality rates than in, for example, the Netherlands (Beemsterboer *et al.*, 1994).

The estimated cost–effectiveness ratios for the Netherlands and the United Kingdom were similar and relatively low (de Koning *et al.*, 1991). Both countries have nationally organized health care systems, high rates of breast cancer incidence and mortality and strictly nationally coordinated screening programmes with clear quality assurance and evaluation criteria. During the 1990s, a reduction in breast cancer mortality was observed among women aged 55–74 in both countries. The reduction is likely to be due partly to the screening activities, but other components of breast cancer control may also have played a role (Quinn & Allen, 1995; van den Akker-van Marle *et al.*, 1999; Blanks *et al.*, 2000b), particularly in the United Kingdom, where breast cancer mortality had already decreased in the early 1990s.

Table 58. Estimated effects, costs and cost–effectiveness of breast cancer screening every 2 years (unless stated otherwise) for women aged 50–69 in various countries

Country (age range)	Breast cancer deaths prevented (if 27 years of screening)	Life-years gained	Difference in life-years gained, 5% discounting	Difference in costs (euros)[a], 5% discounting	Cost-effectiveness ratio (euros/life-year gained)[a], 5% discounting
Spain, Navarra (45–65)	1 100	22 000	Not reported	60	2650
Germany	54 300	860 000	206 500	2000	9600
Spain, Catalonia	195 per year	Not reported	19 450	90	4475
United Kingdom, north-west (50–64)[b]	4 880	81 000	15 000	60	3950
Australia	Not reported	250 000	53 500	450	8300
Spain	22 000	316 000	79 000	560	7125
France	42 000	649 000	155 000	765	4950
United Kingdom (50–69)	72 000	1 046 000	252 000	730	2900
Netherlands	17 000	260 000	61 000	210	3400

From de Koning (2000b)
[a] www.exact.nl
[b] 6% discount rate

Improvements in clinical care may be favoured by implementation of a screening programme, because of improved diagnostic assessment and treatment, and this can be regarded as a positive side-effect of screening programmes. For this reason, it is important that cost–effectiveness analyses also take into account possible changes in treatment patterns.

Application of strict rules

A major problem in comparing the cost–effectiveness ratios of different screening programmes is differences in the analyses. Brown and Fintor (1993) presented a good example of how differences in screening modality, in the assumptions made with respect to the expected effects and in the assessment result in very different cost–effectiveness ratios (see box below). They used a report from the Office of Technology Assessment (US Congress, Office of Technology Assessment, 1987) in the USA and the Dutch study (de Koning *et al.*, 1991). After adjusting the data for the differences, the outcome of the Office of Technology Assessment study was very similar to the alternative described in the study of de Koning *et al.* The cost–effectiveness ratios for different studies cannot be compared unless such adjustments are made. Therefore, an overview of cost–effectiveness ratios based on the same method of analysis provides a better insight into how a screening programme can be ranked internationally. The cost–effectiveness ratios for Navarra, Spain, and for Germany were derived from studies in which the 'Dutch model' was applied (de Koning, 2000b).

Elements of cost–effectiveness

The outcomes of a cost–effectiveness analysis are standard, in the following hierarchical order:

- number of prevented breast cancer deaths and life–years gained in absolute terms;
- discounted effects (see below); and
- discounted cost and cost–effectiveness ratio, adjusted for quality of life.

Effectiveness

The most important benefit of an effective breast cancer screening programme is a reduction in breast cancer mortality, together with life–years of relatively good quality gained. In a cost–effectiveness analysis, this is the most important element. Screening conducted in the 1970s and 1980s was shown to be

Example of revision and recalculation of cost–effectiveness ratios: reconciling calculations from the USA and from de Koning *et al.* (1991)

Report from the Office of Technology Assessment, USA, 1987

Cost–effectiveness = US$ 34 600 / life–year saved

Adjustment for lag effects

↘ Cost–effectiveness = US$ 26 183 / life–year saved

Adjustment for screening price, US$ 50 → US$ 20

↘ Cost–effectiveness = US$ 11 267 / life–year saved

Adjustment for biopsy, costs → saving

↘ Cost–effectiveness = US$ 8931 / life–year saved

Adjustment for effectiveness, 13% → 16%

↘ Cost–effectiveness = US$ 7256 / life– year saved

de Koning *et al.* (1991)

Cost–effectiveness = US$ 7250 / life–year saved

From Brown and Fintor (1993)

effective when compared with no screening (see Chapter 4). On the basis of the early outcomes of three Swedish screening trials, a 16% reduction in breast cancer mortality, i.e. 600 fewer women dying from breast cancer annually, was estimated to be realistic for a nationwide programme of breast cancer screening every 2 years for women aged 50–69 in The Netherlands (de Koning *et al.*, 1991). Integration of more data from five Swedish screening trials published in 1993 (Nyström *et al.,* 1993) indicated a probable 17% reduction in total breast cancer mortality in The Netherlands, that is to say 800 fewer breast cancer deaths per year and 15 life–years gained per individual (de Koning *et al.,* 1995a). In the United Kingdom, it was estimated before implementation of the nationwide breast cancer screening programme that screening of women aged 50–64 every 3 years should reduce breast cancer mortality by 25%, assuming 70% participation.

If screening is effective, it also leads to a reduction in advanced stage disease. This is important not only from the point of view of reduced costs due to less radical treatment but especially from the perspective of improved quality of life, less morbidity and fewer out-patient clinic visits (de Haes *et al.,* 1991; de Koning *et al.*, 1992).

Unfavourable effects

The impact of national programmes on quality of life has been the subject of much discussion. The potential negative effects of the screening examination itself (Ellman *et al.*, 1989), the referral of a significant number of women with benign lesions (Gram *et al.*, 1990) and the consequences of earlier and often more intensive treatment cannot be ignored.

Many factors determine the favourable and unfavourable effects of screening and, possibly, its cost–effectiveness. Important variables are improved prognosis of cases detected at screening, the predictive value of the screening test and the detection of DCIS that would have progressed to invasive carcinoma. Although mortality reduction is the fundamental effect, other desirable and undesirable consequences of screening may influence a woman's quality of life.

Figure 44 summarizes the most important favourable and unfavourable effects of a screening programme (per million screens), other than mortality reduction or gain in crude number of life–years. The scale represents the relative weights given to various types of morbidity at different phases, 100 representing perfect quality of life and 0 representing the worst possible state. The value 82 for adjuvant hormonal treatment implies an estimated 18% loss in effect during this phase as compared with the situation of perfect health (de Haes *et al.,* 1991; de Koning *et al.*, 1991). Screening 1 million women is expected to make adjuvant hormonal treatment unnecessary for 525 women, owing to the smaller number of women with lymph-node metastases. Therefore, this effect would lead to an increase of (525 × 0.18 × 2 years) = 189 quality-adjusted life–years.

The screening examination itself is estimated to have only a slight, short-term (1 week) negative impact, resulting in a decrease of (1 million × 0.006 × 1/52 year) = 115 quality-adjusted life–years. Even though it is estimated that 15.8 million women will have been screened during the period 1990–2017, only 7% of the total negative quality adjustment is incurred by these examinations. More importantly, breast cancer will be diagnosed in approximately 4500 women an average of 4 years earlier

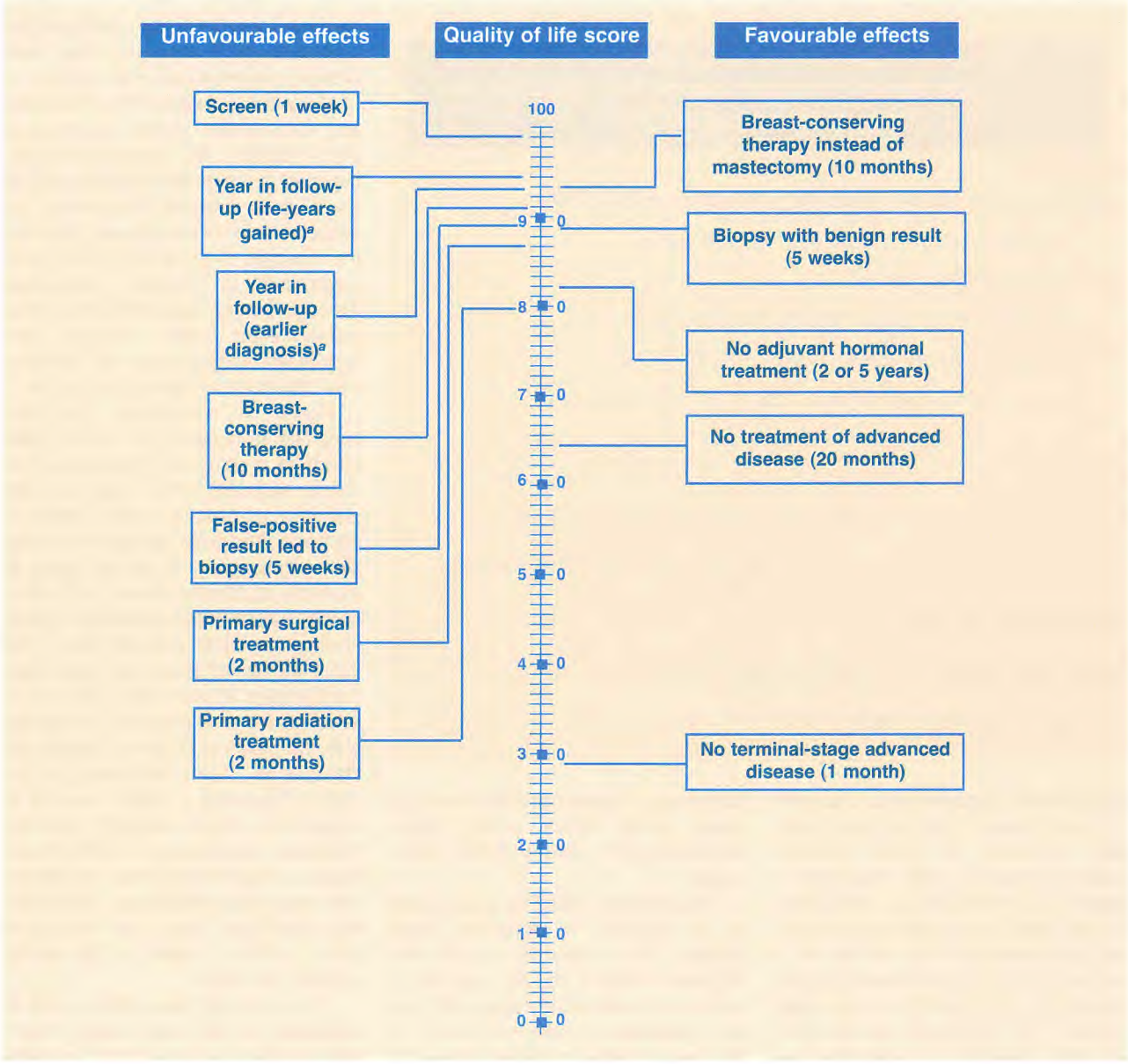

Figure 44 Unfavourable and favourable effects, other than reduction in mortality, of 2-yearly screening of women aged 50–70
Differences are given per million screens (not discounted). Scale represents relative weights given to various types of morbidity, with decreases in quality of life due to unfavourable changes on the left and increases due to favourable changes on the right. Effects on morbidity are classified as short-term (treatment phase; 2 months), intermediate (first year after treatment; 10 months) and long-term (life–years). Durations (in parentheses) are assumed durations of impact on health-related quality of life for each episode.

From de Koning *et al.* (1991)

[a] Appear on negative side of the balance to correct the total gain in life-years for quality of life

(lead time), and more than 1000 patients will experience a longer disease-free interval (16 500 life–years gained). In Figure 43, both types of additional years appear on the negative side of the balance, in order to correct the total gain in life–years for quality of life. The small loss in cost–effectiveness during follow-up is explained by the introduction of several medical follow-up procedures and by the negative impact on a woman's quality of life resulting from breast surgery and the knowledge that she has had breast cancer. The large increase in woman–years in follow-up is almost entirely responsible for the negative quality-of-life adjustment, whereas the decrease in the number of patients with advanced breast cancer as a result of screening is responsible for 70% of the positive quality-of-life adjustment. In the 27-year programme, a total of 252 000 quality-adjusted life–years are gained, which is a very small decrease from the 260 000 unadjusted life–years gained. As the more favourable effects are preceded by unfavourable effects, this difference becomes larger when effects and costs are discounted.

Differences in quality-adjusted life–years is a preferable measure to crude life–years gained. When associated morbidity in all possible phases is taken into account, favourable and unfavourable effects other than mortality reduction have only a limited impact and appear to cancel each other out. More extreme assumptions about expected cost–effectiveness in the various phases result in an adjustment of between – 20% (most unfavourable) and + 3% (most favourable) on life–years gained for 2-yearly screening of women aged 50–69 (de Haes *et al.,* 1991).

It can be estimated that only one-quarter of the women in whom breast cancer is detected by screening benefit, in the sense that they do not die from breast cancer (de Koning, 1995). Even then, they have to live for longer with the knowledge of having cancer, which can negatively affect their quality of life. Some women in whom breast cancer is diagnosed at a screening will die from it. Other women with screen-detected breast cancer might not have died from the disease, depending on the general diagnostic and therapeutic quality and survival rate. Screening generates a number of false-positive results, leading to temporary anxiety and additional assessment. False-negative, and also true-negative, results may falsely reassure women, so that they are less aware of symptoms and wait too long to see their general practitioner.

Costs

The costs of systematic breast cancer screening can be divided roughly into those for organizing and implementing the programme, those for assessment, including that for false-positive results, those for additional primary treatment and savings in treatment costs due to a decreased number of cases of advanced disease. There has been debate about whether costs for health care not related to the treatment of breast cancer should be considered as well (Johnston, 2001).

The direct costs for screening include all those for inviting and screening women, e.g. for employing and training staff and for housing and material. These depend not only on the characteristics of a country but also on the organization of the screening programme. Separate screening units require high investment but guarantee high performance and better use of capacity. A centralized organization will keep overhead costs for coordination, quality assurance and monitoring low. Other determinants of the costs of screening are the type of invitation (personal letter or only general announcement), the number of views per screening examination, single or double reading of films and, especially, the total number and age range of the women invited and the frequency of screening examinations.

The costs of assessment depend initially on the number of recalled or referred screen-positive cases and secondly on the setting in which the further diagnostic assessment is carried out. In general, the higher the recall or referral rate, the higher will be the proportion of false-positive screen results. As diagnostic assessment of women with a true-positive result will almost always result in some kind of biopsy, the costs can be estimated precisely; however, this is not the case for false-positive results. In some cases, assessment will be limited to clinical investigation and a review of screening mammograms; in others, the assessment will be extended by magnification views and/or ultrasound examination, and a proportion of women with false-positive results will undergo a diagnostic biopsy. It is therefore difficult to estimate the costs related to false-positive screening results, and reliable data on the distribution of diagnostic procedures are often available only in an ongoing programme.

In the first few years after implementation of a screening programme, the treatment costs will rise owing to the increase in breast cancer detection. Thereafter, when most women have been invited for incidence screening rounds, the number of breast cancers detected in an advanced stage can be expected to decrease, and the costs of extensive breast cancer treatment can be saved (de Koning *et al.*, 1992; Richards *et al.*, 1993).

Implementation of a breast cancer screening programme can lead to a broad tendency to earlier detection of symptomatic breast cancers, as a consequence of publicity, increased awareness and improved early detection methods in clinical care. Although this generates additional costs, it will ultimately lead to a shift towards prognostically more favourable tumour stages and the possible saving of treatment costs for palliative care. A screening programme may also lead to less diagnostic

assessment of breast symptoms among screened women. This assumption in the Dutch cost–effectiveness analysis was confirmed later by other studies in The Netherlands, showing that the demand for mammography outside the screening programme decreased among targeted women and remained stable in groups that were not targeted (Beemsterboer *et al.*, 1999; van Leiden *et al.*, 1999). Interestingly, during the first 2 years after the start of implementation of the screening programme, use of mammography outside the programme increased in all age groups but was significant only in the targeted age band. After 2 years, the frequency of spontaneous mammography returned to the previous level in the age groups that were not targeted but was significantly lower than before the start of the screening programme in the targeted population (Beemsterboer *et al.*, 1999).

Discounting effects and costs

The costs and savings of a screening programme are not all seen at the same time. For example, the costs of the screening test(s) itself (seen at the start of the screening) represent the largest share, while much of the savings is due to avoidance of future treatment of disease. Furthermore, various target populations can be screened for the same disease, leading to different time profiles. It is generally accepted that earning an amount of money today is preferable to earning the same amount next year, because it can be put into a bank account where it will 'grow' by earning interest. This concept is called 'time preference' in economics. For example, if the 'real' interest rate (the interest without inflation) is 3%, a sum of 1000 will grow to 1000 + (1000 × 3%) = 1030 in 1 year. Conversely, the amount of 1030 of next year can be regarded as equivalent to an amount of 1000 in this year, provided the interest rate is 3%. The interest rate in this reverse reasoning is called the 'discount rate', and the amounts obtained by applying discount rates to future costs and savings are called 'present values'.

Table 59 gives an example of how cost–effectiveness indices should be presented, for the Dutch situation. First, the table presents the number of breast cancer deaths and the number of life–years lost as a result of breast cancer in the absence of mass screening and in the presence of mass screening, respectively. All effects are evaluated without discounting, but discount rates are applied to the costs. The effect of discounting is that the later certain costs arise, the less heavily they weigh in the cost–effectiveness analysis. The higher the discount rate, the more strongly this mechanism works. Various discount rates have been proposed (3%, 5%, 6%, 10%). In this table, a discount rate of 3% was used. The costs for screening, diagnosis and breast cancer therapy are

Table 59. Effectiveness, cost and cost–effectiveness of mammographic breast cancer screening in The Netherlands

	No screening	Screening (difference from no screening)
No discounting		
Effectiveness		
No. of deaths from breast cancer	351 364	– 31 195
Life–years lost from breast cancer (× 1000)	6 374	– 514
Quality-adjusted life–years lost (× 1000)	7 168	– 468
3% discounted		
Effectiveness		
No. of deaths from breast cancer	140 520	– 16 180
Life–years lost from breast cancer (× 1000)	2 395	– 203
Quality-adjusted life–years lost (× 1000)	2 715	– 179
Costs (x 10^6 euros)		
Screening	0	+ 630
Diagnosis	2 921	– 58
Primary treatment	4 159	+ 119
Follow-up	1 456	+ 43
Palliative care	5 481	– 287
Total	14 017	+ 448
Cost–effectiveness (euros)		
Cost per life–year gained		2 209
Cost per quality-of-life year gained		2 496

Adapted from de Koning *et al.* (1991)
Assuming women aged 50–69 screened every 2 years

distinguished to provide insight into the costs and savings at various stages of breast cancer. Finally, in the computation of cost–effectiveness ratios, a discount rate equal to that used for the cost should be applied to the effects. Although this principle has been debated, it is based on theoretical grounds. One is that not discounting effects will always lead to a situation in which postponing a programme (discounted less cost) is more cost–effective than starting the programme today. The cost–effectiveness is expressed as the cost per breast cancer death prevented or as the cost per life–year gained. If the effects have been adjusted for quality of life, the outcome is cost per quality-adjusted life–year gained.

Modelling for policy decisions

The Netherlands was one of the first European countries to begin organized breast cancer screening. In 1974 and 1975, population-based, experimental, mammographic screening programmes were started in the cities of Utrecht and Nijmegen. The two programmes differed with respect to the targeted age groups, the screening interval and the re-invitation policy. In a case–control study, the two programmes were estimated to have resulted in a reduction in breast cancer mortality among screened women of 50–70% (Collette *et al.*, 1984; Verbeek *et al.,* 1984).

In the 1980s, the Department of Public Health at Erasmus University, Rotterdam, developed a computer-simulation package for analysing the effects of screening (MIcrosimulation SCreening ANalysis, MISCAN). The natural history and epidemiology of the disease, the design of the screening programme and the performance of screening are incorporated in this application. Roughly summarized, it simulates life histories in the absence of a screening programme and evaluates how these would be changed by various screening strategies (van Oortmarssen *et al.*, 1990; de Koning *et al.,* 1995a). At the request of the Dutch Ministry of Health in 1986, a national research group was set up to determine the expected effects of a nationwide breast cancer screening programme based on model calculations with data from three randomized screening trials and from the two Dutch experimental programmes. With the inclusion of estimates of various cost aspects, this evaluation became an extensive cost–effectiveness analysis (de Koning *et al.,* 1991). It takes into account various screening strategies with respect to the total number of screen examinations per woman, the length of the interval between successive screening rounds and referral modalities.

In general, the age at which a programme is started, the interval at which the test is applied and the age at which the programme is stopped are considered the major organizational aspects (Commission of the European Communities, 2000). Unfortunately, these aspects cannot be simply copied from experience elsewhere, because each country and trial is unique in terms of the underlying incidence and stage distribution of breast cancer and the screening design, which must be taken into account in interpreting 'efficacy' (de Koning *et al.*, 1995a; de Koning, 2000b). Small differences in general circumstances or in design can have heavy consequences on both effects and costs. The same applies to modelling and its assumptions.

Policy decisions on age categories to be screened

Whereas, in general, screening of postmenopausal women by mammography is considered to result in a reduction in breast cancer mortality, there remains uncertainty about its effect in women under 50.

Younger ages

It appears to be more cost–effective to increase the frequency of screening examinations in a programme for women aged 50–69 than to screen women under 50 (de Koning *et al.,* 1991). The same conclusion was drawn from a study on the screening programme for breast cancer in Catalonia, Spain, for women aged 50–64: on the basis of proven benefits and costs, extension of the programme to older women would be more effective than including younger women (Beemsterboer *et al.*, 1998a). This conclusion was drawn in spite of the fact that in the Catalonian study extension to older and to younger ages appeared to be almost equally cost–effective and that, theoretically, younger women could gain more life–years. Extension to older women, however, would prevent more breast cancer deaths. Furthermore, screening has proved to be effective for women aged 50–69 years, whereas the effectiveness in younger women remains uncertain.

The lower cost–effectiveness of screening younger women is due to the lower breast cancer incidence and the poorer performance of the screening test due to denser breast tissue, resulting in a lower positive predictive value of an abnormal mammogram and higher rates of false-positive and false-negative results. These disadvantages could be partly outweighed by a higher frequency of screening examinations; that, however, would increase not only the costs but also the risk for radiation-induced breast cancers (Beemsterboer *et al.*, 1998b).

Figure 45 shows the marginal costs per additional life–year gained and the corresponding changes in total costs with different screening policies, on the basis of inequal effectiveness by age group (de Koning *et al.*, 1991). In comparison with increasing the invitation frequency of a 2-yearly screening programme for women aged 50–69, extension to women aged 40–49 would lead to a relatively high marginal cost–effectiveness ratio (additional cost

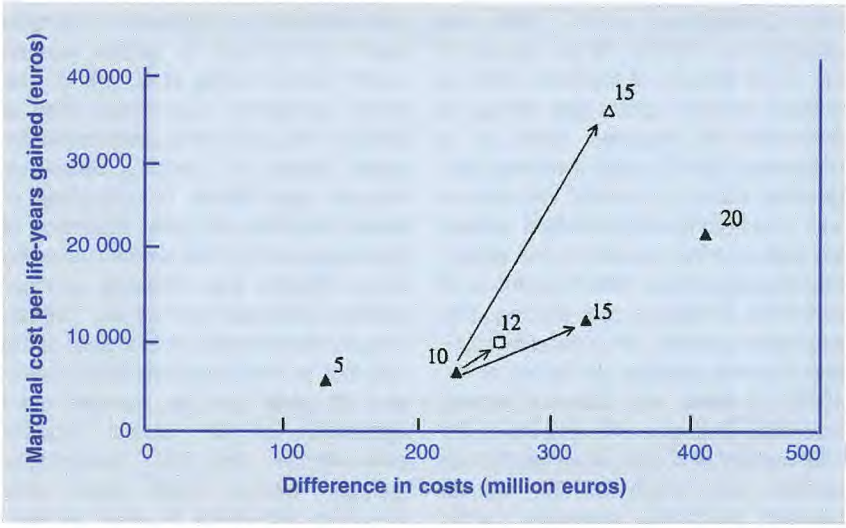

Figure 45 Marginal cost–effectiveness (additional euros per additional life–year gained) of six breast cancer screening policies: 5, 10, 15 or 20 invitations to women aged 50–69 (filled triangles), 12 invitations to women aged 50–75 (square) and 5 invitations to women aged 40–49 followed by 10 invitations when they are 50–69 (open triangle).The corresponding differences in cost for each screening policy are shown on the horizontal axis; 5% discount rate.
From de Koning *et al.* (1991)

per additional life–year gained) of 35 500 euros. This finding is in line with the analyses of Salzmann *et al.* (1997), who showed that the cost–effectiveness was almost five times that in older women.

Upper age limit

Another point that has not been completely resolved is the upper age limit for mass screening. The increasing breast cancer incidence with age favours better performance of a screening test among older women in contrast to younger women. Nevertheless, participation rates among older women may be lower, relatively more of the breast cancers detected may be of lesser clinical importance, and competing causes of death will play a greater role and limit the number of life–years gained. All these factors will affect the cost–effectiveness of screening older populations unfavourably (Commission of the European Communities, 2001). However, few empirical data are available from screening trials and pilot programmes on women over 70; thus, the choice of an upper age limit of 69 or 70 would seem arbitrary. Some studies have suggested that mortality from breast cancer is also reduced among women aged ≥ 70 (Tabár *et al.*, 1989; van Dijck *et al.*, 1996), but no large-scale randomized controlled trials have been performed to settle this question.

Model simulations show that breast cancer screening is cost–effective for women aged > 69 years, on the assumption that the efficacy is the same as in women aged 50–69 (Boer *et al.*, 1995; Kerlikowske *et al.*, 1999). It is likely that, in organized programmes, reasonable attendence rates can be achieved for this age group. However, it is conceivable that, after a certain age, the balance between the benefits and harms of screening will become unfavourable. This depends theoretically on the 'behaviour' of the preclinical sojourn time of breast tumours, i.e. whether it increases continuously with age or whether it remains stable in women over a certain age, for instance 65 years (Figure 46). In the first case—a continuous increase with age—unfavourable effects of screening, such as detection of clinically less important cancers and concomitant loss of quality of life, will outweigh the benefit of screening from a certain age (Boer *et al.*, 1995).

Policy decisions on screening interval

The choice of frequency of screening depends directly on the epidemiology and natural history of the disease and especially on the sojourn time. If this increases with age, as is generally assumed for breast cancer, a longer screening interval would be justified for older women. A study in which an increase in the sojourn time for preclinical breast cancer was observed showed a more favourable cost–effectiveness ratio in women aged ≥ 65 when they were screened less frequently than younger women (Boer *et al.*, 1999a). However, the logistics of a population-based breast cancer screening programme with different invitation schedules for different subgroups, especially for a programme in which mobile units are used, may become complex and expensive. In the Swedish trials, the screening intervals varied from 18 to 33 months. Currently, most organized breast cancer screening programmes invite eligible women every 2 years (Shapiro *et al.*, 1998).

Of the large-scale nationwide programmes, that in the United Kingdom represents the most important exception, as it provided mammographic screening only every 3 years. Despite this important difference in programme design from the Dutch programme, the cost–effectiveness ratios are similar, but the effectiveness is expected to be lower (de Koning, 2000b). A study in 1998 showed the cost–effectiveness of shortening the screening interval in the programme in the United Kingdom from 3 to 2 years, and estimated that 2-yearly

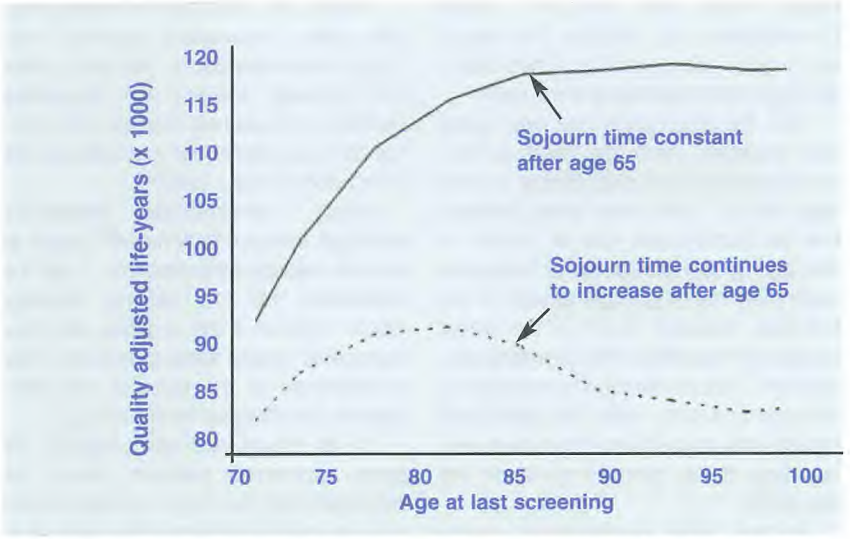

Figure 46 Numbers of 5% discounted quality-adjusted life–years gained as a function of the upper age limit of invitation to screening in a programme with a 2-year screening interval, starting at age 51
From Boer *et al.* (1995)

screening would cost UK£ 2709 per life–year gained versus UK£ 2522 with the current policy (Boer *et al*., 1998). In the same study, it was calculated that extending the upper age limit for invitation from 64 to 69 years would increase the cost per life–year gained from UK£ 2522 to UK£ 2990. The study, which concluded that either of the two alternatives, shortening the screening interval or extending the age range, would be effective, demonstrates that the policy choice of screening interval is related to the choice of age group to invite and vice versa (see also Figure 44).

Policy decisions on high-risk groups

Instead of screening large populations at relatively low average risk of developing the disease in question, one alternative could be to limit screening to persons at high(er) risk, in order to reduce costs and unfavourable effects. This alternative approach would require accurate identification of 'high-risk' individuals. In the case of breast cancer, this is unrealistic, as there are no clear markers of risk. To date, high-risk breast cancer screening is an option for women at high lifetime risk for familial breast cancer (see Chapter 4). This concerns predominantly young women, however, for whom mammography is probably not the most effective screening method. It is therefore still uncertain whether, from the public health point of view, screening high-risk groups would be cost–effective.

Policy decisions on recall or referral

The method used for reading screening mammograms and the recall or referral policy affect the cost of screening and its effectiveness. A low recall threshold results in a high proportion of screened women with a false-positive result and thus lower specificity; a (too) high threshold leads to an inadequate rate of breast cancer detection and a high frequency of interval cancers. These aspects depend on the number of views used, the quality of the films and the training and experience of the readers. At the start of a screening programme, a decision must be taken whether one, two or more readers, generally radiologists, are to read the screening mammograms. If there is to be double (or multiple) reading, various strategies are possible, such as independent double-reading of all films or pre-screening of all films by a first reader followed by double-reading of the initially selected mammograms. In the near future, computer-aided detection may play a role in supporting the reading process. All double or multiple readings will lead to discrepant findings, interpretation and referral recommendations in a proportion of cases, and clear guidelines should be drawn up to deal with the discrepancies and to decide which of them requires a recommendation for recall or referral, e.g. always referral, decision by one reader, decision by a third reader or after consensus of the two readers.

Several studies of the effect of single versus double reading on sensitivity and specificity generally agree that double reading leads to a higher detection rate and to lower specificity (Anderson *et al.,* 1994; Denton & Field, 1997; Blanks *et al.,* 1998). However, it is not clear whether the increased effect of double reading compared with single reading is also cost–effective (Haiart & Henderson, 1991; Brown *et al*., 1996; Leivo *et al.,* 1999).

Participation

Inadequate data are available to allow estimation of the cost–effectiveness of recruitment strategies or the effects of different attendance rates.

Cost-effectiveness in practice

The continuing cost-effectiveness of a breast cancer screening programme in actual practice should be monitored carefully. Breast cancer screening should be offered only if there are

concomitant efforts to maximize the ratio of benefits to harm. This requires effective routine monitoring of performance (Commission of the European Communities, 2001). Any unfavourable development with regard to effects or cost might adversely affect the cost–effectiveness and will require interventions.

In The Netherlands, a national team is responsible for evaluating and monitoring the effects of the Dutch breast cancer screening programme (Fracheboud *et al.*, 2001a). In general, the results of subsequent screens are considered the most important indicator of the effectiveness of the programme. Subsequent examinations account for over 85% of all screening examinations now that the Dutch programme has been fully implemented.

As the decision to implement the Dutch nationwide breast cancer screening programme was based on the favourable outcomes of computer simulations with the validated MISCAN model, the observed effects have regularly been measured against those expectations. Now, more than 10 years after the start of screening, it is still not possible to answer the question of whether the programme has resulted in a reduction in breast cancer mortality. Although the mortality rate has been decreasing throughout the past decade in both The Netherlands and the United Kingdom, it is not clear to what extent the screening programmes contributed to this reduction. In anticipation of the answer to this question, the monitoring and evaluation of the programme concentrate on short-term results, such as participation rate, breast cancer detection rate, tumour stage distribution, false-positive and false-negative screening rates and sensitivity and specificity.

Although favourable short-term results do not guarantee a reduction in mortality, they are essential prerequisites. In Table 60, the short-term results of the Dutch screening programme for 1990–97 are compared with the expected results from the MISCAN model (Fracheboud *et al.*, 2001b). The results are largely in line with the expectations, although some deviations were seen.

First, the attendance rate was higher than expected. When the Dutch screening programme was extended to women aged 69–75, there was some concern that the participation rate of women in this age group would be considerably lower than that of younger groups. In the first year, however, 64.6% of the invited women participated in the screening programme. This percentage is expected to increase in future, when this group will include only women who have been participating in the programme since the age of 50.

Second, fewer women were recommended for further diagnostic testing than expected. In initial screens, the high predictive value of referral resulted in a breast cancer detection rate that conformed to expectation (taking into account the falling average age of the women attending initial screens over the course of the years). The breast cancer detection rate at subsequent screens, however, was lower than expected. The question is whether the relatively low referral rate—which is four to six times lower than that in the United Kingdom—is the cause. Efforts must be made to increase the referral rate, in order to increase the detection rate without a disproportionate increase in the number of false-positive screening results.

Third, the stage distribution of cancers detected at subsequent screens was similar to that at initial screens, although it had been expected that the distribution at subsequent screens would be more favourable. Possible explanations are that the natural history of breast cancer is different from that currently assumed, shortcomings in the quality and assessment of the mammograms or a difference in the interpretation or policy in national screening programmes from that in some of the trials (Boer *et al.*, 1999b).

Fourth, the incidence of interval cancers after subsequent screens was higher than expected in the early years but gradually changed with increasing numbers of screened women and interval cancers towards the expected values (Fracheboud *et al.*, 1999).

Finally, considerable differences emerged between regions with regard to several important parameters. It can be concluded that the national average would improve if the regions with less favourable results were to improve their programmes to the level of the other regions (Fracheboud *et al.*, 2001b).

In its annual evaluation reports, the team expressed concern about the detection rate, the stage distribution and interval cancer incidence observed at or after subsequent screens. A study was initiated in 1999 to find ways of optimizing the effectiveness of the Dutch screening programme and to reduce the variation in regional results.

Quality of life

In the early 1990s, there was little empirical evidence for the effects of screening on the short-term quality of life of women who participated or for the long-term quality of life resulting from the expected shift in the number of women experiencing early and advanced disease. The adjustment for quality of life in the MISCAN model was based on the results of a literature review and on the assignment of values to various disease and treatment phases by experts in breast cancer and public health (de Haes *et al.*, 1991). The early computer simulations predicted that 2-yearly mammographic screening for women aged 50–70 would be 8% 'less effective' after adjustment for quality of life. The conclusion was that quality of life was not a major issue in the decision to undertake a large-scale breast cancer screening programme.

Since 1990, there has been increased interest in health-related quality of life, and this aspect has been assessed in

Table 60. Observed and expected results of initial and subsequent screening examinations in The Netherlands, 1990–97

Result	Initial screen		Subsequent screens (< 2.5 years after previous screen)	
	Observed	Expected	Observed	Expected
Attendance (%)	78.5	70.0	78.5	70.0
Referrals/1000	13.1	16.0	6.9	7.5
Biopsies/1000	9.2	12.0	4.5	6.0
PPV of referral (%)	47	41	51	57
PPV of biopsy (%)	66	54	78	72
Screen-detected cancers/1000	6.1	6.5	3.5	4.3
Tumour size distribution of screen-detected cancers (all carcinomas)				
Ductal carcinoma *in situ* (%)	14	13	14	14
T1a–T1c (%)	61	65	64	73
≥ T2 (%)	20	18	17	9
Unknown	5	5	5	5
Axillary lymph node status (% of invasive carcinomas)				
Positive	27	26	23	23
Negative	67	68	71	71
Unknown	6	6	6	6
Interval cancers/1000 woman–years of follow-up	0.96	1.00	0.93	0.96

PPV, positive predictive value
From Fracheboud *et al.* (2001b)

the various phases of breast cancer diagnosis and treatment. The Interdisciplinary Group for Cancer Care Evaluation (GIVIO, 1994) has undertaken programmes to assess the health-related quality of life of patients with early breast cancer since 1980 and in 1994 compared the quality of life in groups with intensive and conservative follow-up. Cockburn *et al.* (1994) and many others have reported on the psychological consequences of mammographic screening (see also Chapter 3).

The measurements of quality of life made about a decade ago are still largely valid; however, their influence on cost–effectiveness may have changed. For instance, with less invasive treatment such as breast-conserving therapy, quality of life will be better than with invasive treatment such as mastectomy. Any shift towards less invasive treatment, resulting not only from mass screening but also from improved medical techniques or new insights, will result in a more favourable cost–effectiveness ratio.

Cost

In the late 1980s, much effort was made to obtain an overview of all the relevant costs of breast cancer screening, diagnosis and treatment. In order to adjust for time preference, the cost–effectiveness ratio was calculated with 3% and 5% discount rates. The development of costs that has been observed subsequently in The Netherlands is largely in line with the estimates. However, some changes warrant renewed cost–effectiveness analysis. For instance, the duration of hospital stay after surgical procedures has been reduced, which will lower the cost.

Developments in the costs of diagnosis and treatment were not monitored as continuously as the effects. Treatment for breast cancer in particular has become more expensive during the past decade, but that will be partly outweighed by the shortening of in-patient stays. Furthermore, the increased costs will be accompanied by increased savings.

Cost–effectiveness

There is no sign that the cost–effectiveness of the Dutch breast cancer screening programme is less favourable than was expected. Many surrogate measures have been monitored and evaluated, and the results do not refute most of the assumptions made in earlier cost–effectiveness analyses. The participation rates, detection rates, size distribution, interval cancer rates and referral rates have been fairly stable or improved over time, suggesting possible effectiveness at the nationwide level. In The Netherlands, assumptions on effectiveness were based initially on three, and later five Swedish randomized controlled trials (de Koning *et al.*, 1991, 1995a). Although recent reviews have given rise to discussion and the results can be used in sensitivity analyses for cost–effectiveness, they do not change the general picture that breast cancer screening can be very cost–effective. Assumptions of 50% lower effectiveness lead to ratios in the order of those of the Dutch cervical cancer screening programme.

Limitations of cost-effectiveness analysis

Evaluation of the cost–effectiveness of any screening programme is highly recommended. This requires adequate quantification of all the relevant effects and costs, which, in turn, requires continuous collection and registration of relevant data. The effort required to accomplish this is often enormous. In practice, a number of obstacles may limit a comprehensive cost–effectiveness analysis. These include practical limitations, gaps in knowledge, current and future developments in the diagnosis and treatment of breast cancer and alternative screening policies. Model-based cost–effectiveness analysis can provide relevant estimates of these long-term effects, but comparison with published analyses is often hampered by differences or lack of clarity in the assumptions made. Many so-called model-based cost–effectiveness analyses have been published which do not have the scientific rigour, clarity about assumptions and sensitivity analyses that should be provided. The box below provides an overview of the shortcomings of cost–effectiveness. Some of these shortcomings are discussed below in the context of the Dutch breast cancer screening programme.

Practical limitations

All the information considered important for an optimal cost–effectiveness evaluation of the Dutch breast cancer

Shortcomings in cost–effectiveness analysis

Practical problems

- Limitations of obtaining data (privacy regulations, informed consent)
- Limitations of registries (incompleteness, not national)

Gaps in knowledge

- Natural history of breast cancer (duration of preclinical detectable phase with increasing age, biology of ductal carcinoma *in situ*)
- False reassurance

New developments

- Diagnosis and treatment of cancer (such as large-core needle biopsy)
- Screening methods (digital mammography)
- Alternative screening strategies (< 50 years)

Quality of life evaluation

- Empirical data on episodes induced and prevented

screening programme is simply not available.

- Monitoring the effects of screening, particularly the effect on breast cancer mortality, is hindered by privacy regulations. All women participating in the Dutch screening programme are asked to give informed consent for use of their data in evaluating the screening programme. As such permission is not obtained from unscreened women, evaluation of breast cancer mortality at an individual level is impossible.
- The increasing amount of data registered by the screening organizations and regional cancer registries leads to a longer delay until delivery of the data.
- Some relevant data are not available nationally. For instance, in The Netherlands, there are no national data on the grade of malignancy of breast cancers.

Gaps in knowledge

The model-based cost–effectiveness analysis of 1990 was based partly on a number of assumptions, because of lack of empirical data. Some of the assumptions and uncertainties are discussed below.

- The duration of the preclinical detectable phase with increasing age (≥ 65) is unknown. In order to estimate this duration, it is essential to have information about the detection rate in prevalence rounds and in subsequent rounds at an interval longer than 2.5 years and about the interval cancer rate. In The Netherlands, empirical evidence is expected to become available around 2003, 5 years after introduction of mass screening for women aged 70–75. At that time, the balance between favourable and unfavourable effects of breast cancer screening for women aged ≥ 75 will be estimated more carefully;
- The 1990 cost–effectiveness analysis was based on the assumption that the number of requests by the target population for mammography outside the programme (opportunistic screening) would decrease. Opportunistic screening in a target population can negatively influence the effectiveness of screening and the costs of health care, and opportunistic screening of women in adjacent age groups would be another negative effect of screening from the public health perspective. In order to quantify these aspects, the effect of the start of the Dutch screening programme on the number of mammographies requested in general practice was examined (Beemsterboer *et al.*, 1999). The study showed an increased number of requests by general practitioners after the start of the screening programme, for women in all age groups. More than 2 years after the start of screening, however, the number of requests for mammography in all age groups had decreased to that before the start. In the age group 50–69, the number of mammographies was significantly lower than before screening started, probably due to the introduction of the national screening programme. Opportunistic screening was not clearly demonstrated in adjacent age groups.
- Negative results in a screening examination may falsely reassure screened women and lead to delayed diagnosis of symptomatic breast cancer, either during the interval or at the next screening examination. However, it is not known whether false reassurance indeed plays an important role and, if so, to what extent.
- At the time the decision was made to introduce breast cancer screening in The Netherlands, the radiation dose used in modern medicine was assumed to be negligible (Health Council, 1987). As a result of new techniques and continuous improvement of image quality, the radiation dose has probably increased from 0.5 mGy to about 2 mGy per examination. In order to study the risk of mammographic radiation and the implications for screening programmes for different age groups and intervals, model-based estimates were made of the number of breast cancer deaths induced by low-dose radiation (2 mGy per view) in breast cancer screening programmes and the numbers prevented (Beemsterboer *et al.*, 1998b). This study showed that the balance between the number of deaths induced and those prevented was favourable in the age group 50–69, assuming a screening programme with a 2-year interval. If screening is extended to the age group 40–49, the results are less favourable: one induced breast cancer death versus 66 prevented with a 1-year interval and one versus 97 with a 2-year interval.
- Little is known about how the effects of adjuvant systemic therapy interact with the proposed benefits of screening in trials performed in periods when such treatment was not available.

New developments

New developments in breast cancer diagnosis and treatment may also necessitate a review of cost–effectiveness analysis, as they may have consequences for both the effectiveness and the costs of breast cancer screening and for quality of life. Important developments are:

- *New diagnostic procedures for impalpable breast lesions.* Less invasive diagnostic procedures, such as stereotactic core biopsy, have increasingly replaced needle-localized biopsy for the evaluation and treatment of impalpable breast lesions. In general, these procedures are cheaper and are assumed to have fewer negative effects on quality of life than open breast biopsy. Introduction of these procedures will lead to fewer unnecessary open surgical biopsies, thus lowering health care costs and, as a consequence, influencing cost–effectiveness.
- *Digital mammography.* Replacing conventional screen–film mammography by digital mammography will require heavy initial financial investment (purchase of expensive equipment) but in the long run may save costs (transport, files, archives). Another favourable aspect of digital mammography may be a reduction in risk from radiation and better performance of screening programmes due to high-quality imaging. Computer-assisted pre-selection and enlargement of mammographic images are related techniques. The possible consequences are not yet clear.

Cost–effectiveness analysis is essential if an alternative screening policy is being considered, such as extension of the age limit of the target population, an increase or decrease in the screening interval or choice of another screening instrument. For instance, the results of current screening trials involving women under the age of 50, such as the trial in the United Kingdom for women aged ≥ 40 (Moss, 1999), may warrant reconsideration of the lower age limit in the Dutch breast cancer screening programme and an update of the cost–effectiveness analysis.

Because of the relatively unfavourable detection rate after subsequent screens, in combination with the relatively low referral rate, in the Dutch programme, an alternative referral strategy might be considered. Currently, all screening mammographies are read by two readers, and women are referred only if there is consensus between the two. An alternative strategy would be to refer women if at least one of the two readers recommends further assessment of the mammographic lesion. This would result in a higher referral rate and a larger number of false-positive results, but with an expected increase in the detection rate. A revision of the cost–effectiveness analyses would then be required.

The differences found between observed and expected outcomes in The Netherlands may have several implications. There may be some dysfunction in actual screening practice, which should be corrected. The regional differences in particular imply that some improvement is possible. Another explanation is that the expectations were too optimistic. It is not unlikely that the results of 10 years' screening in The Netherlands are more representative of reality than the model-based expectations. In that case, the model used to estimate cost–effectiveness should be reconsidered and revised. At least in The Netherlands, both explanations may be true.