Advantages and Disadvantages of Mammography Screening

Sylvia H. Heywang-Köbrunner   Astrid Hacker   Stefan Sedlacek

Referenzzentrum Mammographie München, Germany

Summary

Mammography screening is the only method presently considered appropriate for mass screening of asymptomatic women. Its frequent use, however, warrants diligent analysis of potential side effects. Radiation risk is far below the natural yearly risk of breast cancer and should not be used as an argument against screening. False-positive calls lead to additional imaging or histopathological assessment, mainly percutaneous breast biopsy. These measures are tolerated and accepted fairly well. Their number is limited by strict quality assurance and constant training. Interval cancers represent a limitation of breast screening that should prompt further research for optimization. Evaluation of overdiagnosis is a highly debated topic in the literature. According to the probably most realistic available calculations, overdiagnosis is acceptable as it is compensated by the potential mortality reduction. Nonetheless, this potential side effect warrants optimal adjustment of therapy to the patient's individual risk. The mortality reduction seen in randomized studies was confirmed by results from national screening programs. A recent case referent study indicated that improvements in mortality reduction run parallel to improved mammographic techniques. Use of less aggressive therapies is another valuable effect of screening. Awareness of potential problems, strict quality assurance, and further research should help to further develop screening programs.

Zusammenfassung

Mammographie-Screening ist derzeit die einzige Methode, die als Populations-Screening für asymptomatische Frauen angemessen ist. Wegen des notwendigen wiederkehrenden Einsatzes der Früherkennung in regelmäßigen Abständen sind eventuelle Nebenwirkungen besonders sorgfältig abzuwägen. Das Strahlenrisiko liegt weit unter dem natürlichen Risiko pro Jahr an Brustkrebs zu erkranken und sollte kein Argument gegen das Screening sein. Falsch positive Befunde können zu bildgebenden oder histologischen Klärungen (hauptsächlich Nadelbiopsien) führen. Diese Maßnahmen werden akzeptiert und gut toleriert. Die Anzahl dieser Biopsien wird durch strenge Qualitätssicherung und Training der Befunder so gering wie möglich gehalten. Intervallkarzinome zeigen die Grenzen des Mammographie-Screenings auf und sollten Anlass zu weiterer Forschung und damit zur Optimierung des Screening-Programms geben. Berechnungen zu Überdiagnosen sind ein derzeit in der Literatur kontrovers diskutiertes Thema. Die der Realität wohl am nächsten kommenden Berechnungen ergeben Werte, die unter Abwägung der gleichzeitig zu erwartenden Mortalitätsreduktion durchaus vertretbar sind. Dennoch sollte diese potentielle Nebenwirkung dazu führen, die Therapie dem individuellen Patienten-Risiko anzupassen. Die in randomisierten Studien beobachtete Mortalitätsreduktion wird durch Ergebnisse aus nationalen Screeningprogrammen bestätigt. Eine kürzlich veröffentlichte Referenz-Kontroll-Studie zeigte, dass der Effekt parallel zur Verbesserung der Mammographie-Technik deutlich zunimmt. Schonendere Therapien sind ein weiterer Effekt des Screenings. Das Bewusstsein für eventuelle Probleme, strenge Qualitätssicherung und Forschung werden helfen, das Programm weiter zu verbessern.

Prof. Dr. med. Sylvia H. Heywang-Köbrunner
Referenzzentrum Mammographie München
Einsteinstrasse 3, 81675 München, Germany
Tel. +49 89 9616613-30, Fax -40
heywangkoe@Referenzzentrum-Muenchen.de

Introduction

Cancer registry data clearly show that early detection of breast cancer (lower TNM stage) is associated with better survival [1]. Detection of non-invasive breast cancer (ductal carcinoma in situ, DCIS) exhibits a relative survival of affected women of > 94%. For invasive breast cancers of ≤ 1 cm, relative survival after 15 years ranges from 90 to 92%; for tumors of 1–2 cm, it ranges around 75% and continues to drop with increasing T stage. On average, the death rate increases by 1.3% per millimeter increase in size. With increasing tumor size at detection, a shift towards higher grading becomes visible. All these facts support our knowledge that early detection has a significant impact on prognosis.

Between 1960–1985 several large randomized studies were completed on the subject of mammography screening. In spite of varying designs (differing age groups and screening intervals, sometimes single view mammography), 6 of 8 large studies showed a significant mortality reduction. According to the latest estimates, the mortality reduction calculated from the randomized studies ranges between 15 and 25% [2–5]. Since randomized studies by definition only compare the results of the invited women with those of non-invited women and do not consider the crossover between both groups, the real effect of a screening program for participants is higher than this number [2, 3]. While most results from European screening programs support these data, mammography screening like many other areas of cancer screening continues to be debated with a critical view.

Undoubtedly, screening like any other measure in medicine has side effects. These side effects must be thoroughly weighted against the expected or proven benefit, since screening involves the participation of healthy women. In this article, an overview is given of the effects and side effects of mammography screening, qualitative and quantitative data concerning their occurrence, first experiences in Germany, and ongoing debates.

Definitions and Design of Mammography Screening

Before analyzing effects and side effects of mammography screening, including the effects observed in our program, some definitions need to be clarified. Mammography screening is designed as a program that combines multiple aspects and represents a measure that i) allows to detect breast cancer early in order to permit mortality reduction and improved therapeutic options; ii) is associated with acceptable side effects for the invited population; iii) yields reproducible results; and iv) can be applied at regular intervals to the population at acceptable costs for the society. So far, mammography screening is the only method that proved capable of fulfilling the above demands and for which evidence of mortality reduction exists from several prospective randomized studies.

Considering the available data, mammography screening certainly represents one of the best investigated fields of medicine.

Concerning its design, the German screening program very strictly follows the European guidelines for quality assurance [6, 7]. Mammography screening is designed as a measure to 'screen' healthy women with no symptoms. It does not include a clinical examination, further tests, or discussion of the results with a medical doctor. Clinical examination is not included in the European screening programs due to insufficient evidence concerning its effect on mortality reduction. In the German health system, however, clinical examination is part of the surveillance provided by the gynecologist und thus compliments the screening examination. Immediate communication of the results to the patient directly after the examination is not possible and is thus not part of the screening program. This has the following reasons: Women in whom an abnormality is detected will be recalled. During the recall, the patient will be informed about the finding and the indicated next steps, and the finding will be assessed (usually by additional imaging, sometimes by percutaneous breast biopsy). Thereafter, the final results are discussed and further recommendations are given. For asymptomatic women with no findings, usually no medical need exist for further discussion. Since the probability of a positive finding at the time of the examination is very low (malignancy is expected in about 3.8 to 5.0 per 1,000 women during the bi-annual examination, which adds up to > 100/1,000 in the course of a lifetime), the adequately informed patient should for also be spared increased anxiety during the time span (usually 8 days) in which double reading and possibly consensus reading take place. Double and consensus reading leads to an increase in detection rate by approximately 8% and a reduction in false-positive findings. If a woman was to be prematurely informed by the initial reader or any other doctor before state-of-the-art double reading is completed, up to 8% of cancers would not be communicated correctly. Furthermore, any (unjustified) suspicion of just 1 reader (which may occur in 5–10% of the cases) could lead to anxiety and further unnecessary measures initiated by the anxious patient.

Even though this process is cost-effective and, due to the special training of the readers and double reading, guarantees higher quality of the mammographic interpretation than screening outside of the program, some women desire the direct contact to the screening physician.

If symptoms are present or if a woman is at high risk of or being treated for breast cancer, complimentary diagnostic methods may be needed, and mammography screening may thus not be adequate or sufficient for assessing these women. Therefore, a detailed history is obtained from every woman who presents for screening. Invited women who report a symptom at the time of the screening examination may undergo the screening mammogram but will then usually be recalled for further examinations. Self-referred women who

are symptomatic are not admitted for screening and should undergo multi-modality imaging and clinical evaluation. In order to further assure correct patient selection, breast cancer patients (women who are treated for metastatic or recurrent breast cancer and breast cancer patients during the first 5 years after diagnosis) are not admitted to the program. Women at risk may undergo mammography screening. Identification of the individual risk, counseling, and referral for further measures are in Germany considered to be the task of the gynecologist.

The screening examination only includes mammography. Breast ultrasound or magnetic resonance imaging (MRI) are to date not considered appropriate for mass screening of asymptomatic women in Germany or internationally [8].

Whenever a clinical finding is reported at the time of screening, mammogram or whenever double and consensus reading yield a positive finding, the patient will be recalled for further assessment. Further assessment consists of further state-of-the art imaging (additional mammographic views, ultrasound, clinical examination, rarely MRI – for state-insured patients at no additional costs) and if needed histo-pathological assessment which in 90% of the cases can be performed using percutaneous core needle biopsy or vacuum-assisted breast biopsy. All steps are to be performed in a quality-assured way within the screening program. The goal is to confirm or exclude malignancy in the recalled patients.

Most European screening programs are performed following the European quality assurance guidelines. It is assumed that if the protocols and thresholds of the guidelines are fulfilled, the goals of the screening program concerning early detection and limited side effects are likely to be fulfilled. While quality assurance parameters are monitored with a short delay, some effects or side effects can only be correctly calculated after 10–15 years [9].

Mammography examinations for early detection that are performed in asymptomatic women without increased risk outside of the screening program are considered 'opportunistic screening examinations'. Opportunistic screening is prohibited according to German law since it does not comply with the statutory quality assurance. Opportunistic screening may be offered even at shorter intervals. Also further examinations (clinical examination ultrasound, sometimes even MRI) are often offered. These examinations may in part be paid with public funds (if patients are referred with pseudo-indications) and partly by the examined women themselves. Since there is no systematic documentation and quality assurance of the complete process, neither effects nor side effects or costs of opportunistic screening examinations can be assessed or evaluated.

Mammography screening in the US is not organized in a screening program like in Europe. Screening examinations in the US are usually performed annually after age 40. Training, quality assurance, and requirements for readers are different, and much higher rates of further assessment are common.

Data from screening in the US can thus not be directly applied to European screening programs.

Disadvantages of Mammography Screening

Radiation Risk and Other Risks

Mammography is associated with a small amount of radiation. The average glandular dose is calculated as 4 mGy per breast. The individual dose may differ depending on breast size and compression. However, with modern mammographic techniques and full-field digital mammography in particular, the glandular dose has dropped significantly during the last decades and continues to decrease. Beyond age 40, the risk of radiation is very low and much lower than the natural risk of dying of breast cancer. Today, the natural lifetime risk of a 50-year-old woman to be affected by breast cancer ranges around 8.8%. If a woman undergoes biannual screening mammography from age 50–69, this risk might rise to 8.9% [10–12]. Yaffe et al. [12] assume that this risk will be reached by a screening regimen that includes annual to biannual mammography screening between age 40–75. This includes about twice the number of mammograms scheduled in the German screening program. Hence, considering the fact that mammography can save lives, the radiation risk is irrelevant compared to the advantage of early detection. However, since screening radiation is applied to the healthy population, screening mammography must be subject to strict quality assurance. Furthermore, some women are afraid that compression might cause breast cancer. These women can be assured that no cancer can be caused by compression.

Risk of False Alarm

Like every medical test, screening mammography may detect abnormalities that require further evaluation but will eventually turn out to be benign. Any recall after a screening mammogram is initially treated as a false alarm. Quality assured double and consensus reading assure that the number of recalled women is as low as possible. For subsequent screening rounds, the recall rate presently ranges around 3–4% in Germany, which fully complies with the European guidelines. That is, 30–40 of 1,000 screened women are recalled per subsequent screening round [7, 13]. Quality assured assessment generally starts with imaging evaluation. The latter includes additional mammographic views, ultrasound, and rarely MRI. On the day of the imaging assessment, malignancy can be excluded on average in 20–26 of the 30–40 recalled women based on additional imaging alone [13, 14]. These women are immediately informed about the benign result and will be invited to the regular next screening round 2 years later. Presently, in 4 of the 30–40 women, a recommendation for short-term follow-up is given after the imaging assessment [13, 15]. These women are released with the information of a probably benign change that does not require biopsy. In

approximately 10 of the 30–40 recalled women (that is 1% of the population that undergoes a subsequent screening round), histopathological assessment is recommended which in 9/10 women is performed as an outpatient procedure under local anesthesia [13–15]. In 1/10 women, open surgery is needed to exclude malignancy. This measure is necessary in rare cases for which percutaneous breast biopsy is not possible or may be insufficient (discordant imaging-histopathology correlation or so-called B3 lesions which comprise benign lesions with unknown biological potential). Breast cancer is presently confirmed in 4–5 of 1,000 screened women. Thus, approximately every other biopsy yields a malignant result [13–15].

From the medical standpoint, a screening mammogram (subsequent round) has a false-positive rate of up to 4% (which is very low for a medical test). Histopathological assessment (mostly percutaneous breast biopsy) is necessary in 1% of the screened women in order to detect malignancy in 0.4–0.5% of the screened women per subsequent screening round. Some authors mention the total number of examinations needed in 1,000 women to detect the incident 50–65 breast cancers among 1,000 screened women: during 10 rounds or 20 years of screening 400 recalls and 100 percutaneous breast biopsies are recommended per 1,000 women. Thus, the average chance of having one false-positive call during 20 years is 40% and that of having a biopsy is 10%. Considering the above proportions, the negligible medical risks and generally good tolerance of imaging and histopathological assessment today, medical arguments against mammography screening because of unacceptable medical harm cannot be considered as justified.

Psychologically, a false-positive alarm may cause distress in 3 different ways: i) Since with mammography screening women are recalled for assessment, the time span between the patient being informed and the date (or some times 2 dates) of assessment may cause anxiety. If a benign lesion is confirmed after imaging assessment or after percutaneous breast biopsy, the period of anxiety is, however, limited; ii) A recommendation for 6–12 months follow-up (which in the screening program can only be given after imaging assessment is completed) may be associated with uncertainty for several months. Therefore, such recommendations are kept as low as possible in the screening program. In the individual case, good patient information about the high probability of benignity may be crucial; iii) The need for histopathological assessment and the procedure itself may be frightening for the patient. Most publications do not specify these different situations. It may, however, be important to emphasize that only the first circumstance (i)) is specific for screening. Situations ii) and iii) apply in the same way to opportunistic screening. However, the absolute number of recommended short-term follow-up examinations or histopathological assessments is probably higher for opportunistic screening (due to less experience and frequent use of further imaging modalities). Furthermore, the proportion of surgical biopsies among recommended histopathological work-up is likely to be higher.

Generally, psychologists assume that stress lasting for a limited period is tolerated well by the majority of women [16]. Considering all recalls and subsequent interventions, the existing meta-analyses indicate that women who have experienced a previous recall may be more aware and may exhibit somewhat greater anxiety. However, this effect appears to be small and of little consequence for overall psychological health and compliance with subsequent screening invitations [16–18].

In our own experience, the vast majority of recalled women are capable of dealing with this situation. Most women explicitly confirm that they value the thorough investigation of any suspicion and prefer being recalled to taking any unnecessary risk. There are few exceptions; these can often be explained by pre-existing problems or experiences (for example breast cancer death of a close friend or relative).

Interval Cancers

By definition, interval cancers are cancers which become palpable between 2 screening-rounds. Considering that a screening round may last 22–26 months, in Germany a cancer will be registered as interval cancer, if it occurs within 26 months after the previous screening examination. In other programs, both the definition and the screening interval may vary. These facts have to be considered when comparing results from different countries or programs.

Since breast cancers do not grow periodically, interval cancers are associated with all types of screening unless unrealistically short intervals were used. With shorter intervals (annually, as used in the US) or longer intervals (3 years, as used in the UK), significantly fewer or more interval cancers are to be expected. Costs will also differ considerably.

According to the European guidelines which recommend biannual screening, a rate of up to 10% of the annual breast cancer incidence is considered acceptable for the first year after screening mammography. In the second year, up to 30% is considered acceptable [6, 7]. This means that with bi-annual screening programs approximately 25% of the cancers of a screened population may become clinically apparent during the interval.

Overall, interval cancers represent a limitation of screening and not a side effect. Without screening, these cancers would also have occurred. However, screening does not allow to recognize these cancers at a preclinical stage, which is difficult to accept both for the screeners and the patients. Interval cancers include cancers [19] i) which grow (fast) from zero or from an undetectable size to a clinically apparent lesion during the interval (no signs, fast growing tumor); ii) which exist but are mammographically occult and become clinically detectable during the interval (no signs, limitation of mammography); iii) which can in retrospect be seen as an unspecific change that can usually not be diagnosed prospectively

(minimal signs); iv) which have been missed based on an error of the reader(s).

Blinded double reading, consensus reading, special training and continuous feedback from obligatory conferences, as performed in the German screening program, have proven to be the best measure to optimize the sensitivity of screening mammography. However, due to the many normal variants and overlapping presentation of some benign and malignant changes, mammography is a demanding procedure, and errors (even with 2 readers) may occur. Due to the large spectrum of normal variants, certain changes may be seen in retrospect (recognition of a palpable finding) but cannot be expected to be diagnosed prospectively. Overall, the prospective diagnosis of minimal signs (that may be indistinguishable from other much more frequent benign changes) is comparable to finding the needle in the haystack. According to our own experience and extensive discussions with international experts, a fair judgment concerning detectability may not be possible in an unblinded or retrospective setting without normal variants. About 10–15% of the existing cancers in women aged 50–69 are not visible on mammography. Visibility on mammography mainly depends on the type of breast cancer and the density of the surrounding tissue. With increasing breast density, detectability of breast cancers without microcalcifications drops, and the risk of an interval carcinoma increases [20, 21]. Furthermore, lobular cancers and diffusely growing cancers may completely mimic normal breast tissue [22]. Even though mammography screening allows the detection of a high number of small and early breast cancers, some large cancers may be undetectable if the cancer exhibits the above mentioned histopathological features or if surrounding dense tissue obscures the cancer without microcalcifications. This fact, which may seem like a contradiction in view of the capabilities of modern mammography concerning early detection, may be difficult to accept for some patients as well as doctors who are not involved in regular reading of mammograms.

Cancers which are not detected because they develop during the screening interval constitute a general limitation of any bi-annual screening program. Due to the faster growth of breast cancer in younger women, the Swedish screening program invites women aged 45–54 every 18 months. Being aware of these problems, there is an ongoing discussion how the sensitivity of mammography screening might be further improved, particularly in women with dense breast tissue and those at increased risk.

Overdiagnosis
Overdiagnosis of breast cancer in a screening program describes the fact that in a screened population more breast cancers are detected than in a comparable unscreened population of the same age and composition. Part of the additional cancers that are detected in the screening group would never have become apparent without screening and their detection does not contribute to mortality reduction. The reason for this is that the 'overdiagnosed' cancer (which histopathologically represents a real malignancy) does not represent a threat to this woman's life since she will have died of a different cause before the cancer (without screening) is likely to become apparent. Since the individual life span of each woman is not known in advance, it is impossible to exactly define which breast cancer constitutes an overdiagnosis and which constitutes a potential threat to the patient's life if left undetected and untreated. So overdiagnosis is a statistically calculated number that deals with an extreme form of length time bias.

The most extreme assumption would be that a screening program selectively detects breast cancers that would never progress or would even regress [9]. The latter assumption has been published by Zahl et al. [23] who derived these results by comparing a screened Norwegian population with a control group from a different region and a 5-year-earlier time period. This publication is very critically debated by numerous epidemiologists [24]. Since after the first screening round, previous films are always compared with present films, and since decreasing changes are not called, the above assumption would also imply that an important number of breast cancers first grow and thus become detectable but then regress or remain stable, a phenomenon which (even in patients who refuse therapy) has so far not been observed in a noticeable number of breast cancer cases.

The side effect of an overdiagnosis is that the patient who does not benefit from the diagnosis is informed about her diagnosis and undergoes treatment (so-called overtreatment). However, even though overdiagnosis is classed as a side effect of mammography screening, it must be pointed out that there is no reason to believe that overdiagnosis is less important in opportunistic screening. Rather, it is only impossible to calculate overdiagnosis (like other side effects) outside organized screening programs due to lacking documentation and data.

In the literature, estimates of overdiagnosis exist from randomized studies and from screening programs. Published results, however, diverge strongly ranging from 1 to 50% [3, 9, 23, 25–27]. Correct calculation of overdiagnosis is very difficult and heavily debated, since information from control groups is usually sparse. Other influences (regional differences, time trends of nutrition, hormone intake, decrease of multiparity, and the effect of opportunistic screening) may significantly distort the results and need to be considered with caution. Estimates by Zahl et al. [23] and Joergenssen et al. [28] are mainly based on ecological data. Their extremely high estimates have received extensive criticism from cancer registries whose published data they used and who would not confirm their results or conclusions [28–31]. The criticism concerned both the calculation model and the consideration of other influences. Also, for breast cancer, which is on average detected in a population that is 15 years younger than with prostate cancer, medical experience does not support the assumption of a significant number of stable or regressing

cancers. Most calculations ranged from 5 to 10% of overdiagnosis. That is 1 of 10–100 diagnosed breast cancers might correspond to an overdiagnosis.

A quite realistic and very sophisticated calculation was presented by Duffy et al. [9] in 2010. The authors concluded that in the Swedish randomized two-county trial approximately 4.3 overdiagnoses occurred by screening 1,000 women for 20 years between age 50–69. For the British screening program he calculated on average of 2.3 overdiagnoses during 20 years of screening. This corresponds to a rate of less than 10% of overdiagnosis among screen-detected beast cancers. He concluded that the life-saving effects of mammography screening exceed the potential harm of overdiagnosis by a factor of 2–2.5.

Overall, due to differences in remaining lifespan, the risk of overdiagnosis appears to be low for women below age 60. It increases after age 60 and more again after age 70, since the risk of death from other causes increases. The possibility of overdiagnosis also increases with small and slow growing cancers. Since part of the DCIS (even though being a precursor) may not develop into invasive breast cancer during the remaining lifespan of a woman, DCIS must be considered a potential and real source of overdiagnosis and thus requires special attention. To date, breast cancer always requires adequate treatment. Adequate treatment in fact is adapted to the individual risk, and early breast cancer usually requires less aggressive treatment (see below).

Advantages of Mammography Screening

Mammography screening is performed to achieve the below described advantages. Most screening programs are associated with strict quality assurance to assure that these goals can be guaranteed and that the advantages outweigh potential disadvantages.

Mortality Reduction
Mortality reduction is the major goal of mammography screening. Whereas surrogate parameters allow early checks of function, assessment of mortality reduction requires more effort and takes many years after the program is fully installed. This is due to the fact that death from breast cancer usually occurs much later than its diagnosis. During this long time span, various other influences may interact. Complete registration of all cancer is another demanding task considering the high numbers of screened women. Thus, excellent documentation and matching of cancer registry data are needed to obtain correct results. Also, the full effect of a screening program may only be correctly assessed after as much as 15 years of follow-up.

Initial proof of mortality reduction stems from several randomized studies which were performed about 30 years ago. In spite of differences in design, significant mortality reduction

was demonstrated between the invited groups and the non-invited control groups [2, 3]. The initial criticism by Goetzsche et al. [4] concerned 5/8 excellent studies. Using formal criteria, he excluded certain studies because of the chosen cluster randomization (an accepted statistical method for which correction algorithms exist) and because he suspected incorrect assignment of causes of death. All studies were thus thoroughly re-assessed, but the criticism was refuted by the World Health Organization (WHO), and the International Agency for Research on Cancer (IARC) clearly stated that 'many criticisms were unsubstantiated; remaining deficiencies do not invalidate the results of the trials' [32].

Today, mortality reduction from the randomized trials is calculated to range between 15 and 30% [2–4]. Considering the fact that only about 70% of the invited women participated, while part of the non-invited women did undergo screening mammography (crossover), the expected mortality reduction for a participating woman probably exceeds 35%.

The above mentioned latest evaluation of the randomized Swedish two-county study trial by Duffy et al. [9] takes into account overdiagnosis. It is based on long-term follow-up and on evaluation of participation (not invitation). It showed an absolute benefit of 8.8 saved lives per 1,000 screened women. For the British screening program with a 3-year screening interval, he used an adapted model. The results indicate that 5.7 lives could be saved by 20-year screening of 1,000 women. Other evaluations from European screening programs [33–40] have confirmed these results with a relative reported mortality reduction of 35–45% for participants. In general, numbers on relative mortality reduction could be distorted if a significant number of cancers were overdiagnosed. Absolute numbers, as reported by Duffy et al. [9], or numbers derived from case control or case referent studies are not distorted by potential overdiagnosis. In fact, a recent case referent study from the Netherlands [41] points out an increasing effect of mammography screening after 1992 as compared to before 1992. This change is very probably associated with the significant improvements in mammographic techniques during the last decades.

When contributions to mortality reduction were estimated, the contribution of mammography screening appeared to be at least equal to that of improved therapy [42, 43]. Other authors, however, attribute mortality reduction predominantly to improved therapy. Considering that chemotherapy may be more effective in the early stages, both are likely to contribute cooperatively to the reduction of breast cancer mortality [41].

Improved Treatment of Early Disease
Improved possibilities of treatment are an important advantage of early detection. It is well known that early detection leads to a reduced number of mastectomies, better cosmetic results in cases of breast conservation, reduced adjuvant chemotherapy, and increased replacement of axillary dissec-

tion by sentinel node biopsy. Exactly these trends are presently being observed in Bavaria, predominantly in the age group of 50–69 and parallel to the onset of the mammography screening program [23]. Considering the still moderate participation rate of up to 50%, it is remarkable that population trends already become visible. These trends have also been explicitly reported for the Dutch screening program [44]. Opponents of screening question this effect. Partly outdated data have been used from a time period when most DCIS and even many invasive cancers were treated by mastectomy. Today, this has changed considerably. Even when additional treatments are considered due to overdiagnosis, the rates of chemotherapy, mastectomies, and axillary dissection drop with mammography screening. Also, early and less aggressive cancers that might be suspected to correspond with possible overdiagnosis do not require chemotherapy or axillary dissection. Finally, mastectomies are today only performed for extended DCIS which in general represent a significantly higher risk of developing into invasive (and sometimes extended) invasive breast cancers [45]. The same publication points out that aggressiveness of treatment exclusively is determined by the stage at detection and not by the mode of detection (screen- versus non-screen-detected), as long as the same stages are compared. Thus, considering the much better stage distribution among cancers detected by screening compared to the general population of age 50–69 in Germany [13], the better stage distribution should directly correlate with lower numbers of aggressive treatment. Such improvements in therapeutic options concern another important aspect of screen-detected cancers. It also represents an important benefit in the sense that the main fears of affected women (besides the fear of a fatal disease) concern chemotherapy, mutilating surgery, and suffering after axillary dissection.

Improved Quality Assurance of the Diagnostic Chain
Quality assurance of the complete diagnostic chain is of utmost importance to take full advantage of early detection. Early detection in fact is useless if the cancer is missed during subsequent imaging or histopathological assessment or if it is incompletely excised or treated thereafter. The structure of the German screening program has taken this knowledge and experiences from other screening programs into account. In the Netherlands, for example, a significant number of interval cancers had occurred because detected abnormalities were due to structural insufficiencies incorrectly assessed and thus missed outside the screening program. The existence of certified breast centers is of special advantage for quality assured treatment of the cancers verified during screening. Also, the interface between screening and certified breast centers, which also implies regular pre- and postoperative multidisciplinary conferences, appears to function well by now. Thus, today, a promising infrastructure has been created that allows optimizing the cooperation and success of interdisciplinary diagnosis and treatment. The advantage concerns the consequent treatment of early detected breast cancer and the reduction of side effects. It is achieved by continuous training and feedback of all participants, by monitoring and limiting the number of (false-) positive calls, by consequent introduction and use of minimal invasive procedures in the program, and by regular interdisciplinary conferences and feedback. In contrast, with opportunistic screening, neither detection nor further assessment is quality assured. Due to insufficient documentation, no evaluation is possible and deficits cannot be detected. Furthermore, neither effects nor side effects can be assessed. Depending on the extent of opportunistic screening, both effects and side effects must be suspected. Due to the lower or lacking quality assurance, these may not be balanced. The same applies to the cost-benefit ratio. Thus, overall, mammography screening (at least in Germany) has further advantages associated with quality assurance and systematic documentation.

Discussion

In spite of it being the most frequent malignancy and cause of death, the bi-annual incidence of breast cancer is low. Early detection of breast cancer thus necessitates that a high number of healthy women to be examined at regular intervals in order to detect cancer early. During 20 years of screening (age 50–69), about 65 breast cancers are expected to occur of which up to 50 (5 per round) are detectable by screening whereas 15 will present during the intervals. Being applied at regular intervals in large numbers of healthy women, potential harmful effects of mammography screening need to be diligently considered. Therefore, it is crucial that the effects of screening is well proven and reproducible so that side effects become acceptable. Also, costs must be acceptable to society. So far, no other method is accepted or used as sole or supplementary test for mass screening, probably owing to the very high specificity and good reproducibility of mammography [8].

When considering potential side effects, radiation risk is minor compared to the expected benefit. The risk of a false-positive call is multiplied since mammography needs to be applied repeatedly every 2 years in order to detect breast cancer early. False-positive calls can be reduced by adequate training and quality assurance.

Some scientists insist on relating positive effects to the total number of screened women only and adding up side effects of the complete 20 years of screening. Statements that only 1/1,000 screened women will benefit, while 40% of the screened women will experience a false alarm, do not contribute to correct communication and understanding.

However, a woman does not expect a benefit if she is never affected by breast cancer. She will be more interested in what her odds are if she became affected – even at the time when she decides to have a screening mammogram. These chances are as follows (using an average of the numbers cited in the

literature): 15 out of 65 women with breast cancers (detected during 20 years of screening) would die without mammography screening. With mammography screening at age 50–69, 5 lives can be saved.

It is true that a 4% recall rate and 1% biopsy rate per round add up to 40% and 10% during 20 years of screening. However, 40% of women will be recalled once during 20 years etc. So, much of the presently communicated information may be unclear, and instead of percentages real numbers and proportion might be more sensible to communicate to women. Overall, recalls and even minimal invasive biopsies are tolerated quite well without damage to the overall psychological health and are accepted by most women. This may be due to the fact that a limited time of distress has to be tolerated frequently during life. Most women emphasize that they prefer being recalled to cancer being missed. Nevertheless, individual differences exist and it is sensible to train personnel to recognize women who are very concerned. In these women, appointments for assessment should be organized more quickly and special attention should be paid to their concerns. The fact that psychological support has not been integrated in the screening program may in some cases be a disadvantage.

Overdiagnosis may presently be the most important side effect of breast cancer screening. Using the average figures cited above, up to 5 of the 65 women with breast cancer would not have known about the disease without screening, which results from mammography screening being a method that allows to recognize disease before it is clinically apparent. These women do not benefit because their cancers do not threaten their lives during their individual residual lifespan. However, in other situations, even prognostically favorable cancers continue to grow. Also, a shift of tumor grading is assumed to occur over time. Since other factors (like residual lifespan) are unknown, it is impossible to determine which prognostically favorable cancer is important to detect and treat and which might correspond to an overdiagnosis on an individual level. Calculations as recently presented by Duffy et al. [9] are compatible with the observed biological behavior of breast cancers. They support the value of mammography screening, since more lives will be saved than overdiagnoses will be made [9]. It should be mentioned that overdiagnoses (since they predominantly concern prognostically favorable tumor entities) mostly do not lead to aggressive therapies. Due to the general difficulty of their exact determination, calculations of overdiagnoses are presently associated with a significant potential of error. Since overdiagnosis cannot be avoided, it appears most important that therapy of early detected and of prognostically favorable malignancy is optimally adapted to the patient's individual risk. Also, patients with early breast cancer should be adequately informed about their excellent prognosis. In most cases of invasive and in situ breast cancer, individual risk adaption is already in place. Some critical entities do, however, concern the rare extended cases of DCIS grade 1 and in part DCIS grade 2. Overall, further research will be necessary to further individualize and optimally adapt therapy to the existing risks.

If individual anonymized files systematically remained available for cancer registry matching and longitudinal follow-up, valuable information could help solve such important questions. Unfortunately, to date, uncertain rights of access and data protection issues appear to prevent this source of knowledge from becoming available for research purposes necessary to improve diagnosis and treatment.

One of the most frustrating issues remains the question of interval cancers. Even optimum quality assurance cannot prevent interval cancers from occurring with the present screening regimen. On average, 15 of the 65 breast cancers occurring during 20 years of screening are expected to present as interval cancers. Knowing that interval cancers are more frequent in dense breast tissue, further research should concentrate on evaluating complimentary methods that might help to further improve sensitivity without unacceptable loss of specificity.

The effect of mammography is well proven. The exact effect of the German mammography screening can only be determined several years from now. Considering the high proportion of opportunistic breast screening before and during the screening program, it is unlikely that the same net effect can be achieved as in other countries. This will be due to a smaller difference between opportunistic screening and service screening. Unfortunately, potential side effects and costs of opportunistic screening cannot be calculated either. In order to allow at least some estimates, all available information from these patients should be gathered and made available.

In spite of existing side effects, mammography screening appears to allow both mortality reduction and improvements in treatment options with early disease. Also, detection of DCIS may be useful for better risk management. Both improvement of prognosis and reduction of feared aggressive therapies justify mammography screening. Based on their very different impact on the patient's life, the advantages outweigh the disadvantages. However, we must be aware of this delicate balance and the still existing limitations of mammography screening. These should be an incentive for further research and continued critical re-assessment of our decisions.

References

1 www.tumorregister-muenchen.de/facts/surv/surv_C50f_G.pdf.

2 Becker N, Junkermann H: Nutzen und Risiko des Mammographiescreenings. Dt Ärzteblatt 2008; 105:131–36.

3 Nelson HD, Tyne K, Naik A, et al.: Screening for breast cancer can update for the US Preventive Services Task Force. Ann Int Med 2009;151:727–42.

4 Gøtzsche PC, Nielsen M: Screening for breast cancer with mammography. Cochrane Database Syst Rev 2009;(4):CD001877

5 Gøtzsche PC, Olsen O: Is screening for breast cancer with mammography justifiable? Lancet 2000; 355:129–34.

6 Richtlinien des Bundesausschusses der Ärzte und Krankenkassen über die Früherkennung von Krebserkrankungen (,Krebsfrüherkennungs-Richtlinien'): www.mammo-programm.de/service/infomaterial-fachservice.php.

7 Perry N, Broeders M, De Wolf C, Törrberg S, et al. (eds): European Guidelines for Quality Assurance in Mammography Screening, 4th ed. Office for Official Publications of the European Communities, Luxembourg, 2006.

8 Heywang-Köbrunner SH, Schreer I, Heindel W, Katalinic A: Bildgebung für die Brustkrebsfrüherkennung. Dtsch Ärztebl Int 2008;105:541–7.

9 Duffy SW, Tabar L, Olsen AH, et al.: Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomised trial and from the breast screening programme in England. J Med Screen 2010;17:25–30.

10 National Academy of Sciences: Health risks from exposure to low levels of ionizing radiation: BEIR VII – phase 2. Washington DC, National Academies Press, 2006.

11 Strahlenschutzkommission: Evaluierung von Nutzen und Risiken im qualitätsgesicherten Mammographie-Screening in Deutschland. Empfehlung der Strahlenschutzkommission 2010. www.ssk.de.

12 Yaffe MJ, Mainprize JG: Risk of radiation-induced breast cancer from mammographic screening. Radiology 2011;258:98–105.

13 Heywang-Köbrunner, Hacker A, Sedlacek S, Engel J: Mammographie-Screening – eine Zwischenbilanz. Gyn und Gebhilfe 2010;7:31–8.

14 www.mammo-programm.de.

15 Qualitätsbericht des Mammographie-Screening-Programms von 2005–2007. www.mammo-programm.de/service/infomaterial-fachservice.php.

16 Brewer NT, Salz T, Lillie SE: Systematic review: the long-term effects of false-positive mammograms. Ann Intern Med 2007;146:502–10.

17 Armstrong K, Moye E, Williams S, Berlin JA, Reynolds EE: Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. Ann Intern Med 2007;146:516–26.

18 Salz T, Richman AR, Brewer NT: Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. Psychooncology 2010;19:1026–34.

19 Hofvind S, Geller B, Skaane P: Mammographic features and histopathological findings of interval breast cancers. Acta Radiol 2008;49:975–81.

20 Boyd NF, Guo H, Martin LJ, et al.: Mammographic breast density and the risk and detection of breast cancer. N Eng J Med 2007;356:297–300.

21 Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH, White E: Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 2000;92:1081–7.

22 Braisch U, Geiss K, Meyer M: Bevölkerungsbezogene Betrachtung der Tumorstadienverschiebungen durch Mammographie-Screening in verschiedenen histologischen Untergruppen. www.ekr.med.uni-erlangen.de/GEKID/Vortraege/GMDS2010_VortragBraisch.pdf.

23 Zahl PH, Maehlen J. Welch HG: The natural history of invasive breast cancers detected by screening mammography. Arch Int Med 2008; 168:2311–6.

24 Paci et al, Wedon-Fekjoer et al, Kalager M, et al.: Rapid responses to Zahl PH, Maehlen J, Welch HG (Arch Int Med 2008;168:2311–6). Arch Int Med 2009;169:997–9.

25 Puliti D, Zappa M, Miccinesi G, et al.: An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. Eur J Cancer 2009;45:3166–71.

26 De Koning HJ, Draisma G, Fracheboud J, de Bruijm A: Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast Cancer Res 2006;8:202.

27 Zackrisson S, Andersson I, Janzon L, et al.: Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. BMJ 2006;332:689–92.

28 Jørgensen KJ, Gøtzsche PC: Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 2009;339:b2587. Review.

29 Pisani P, Foreman D, et al.: (Rapid responses to Jørgensen KJ, Gøtzsche PC BMJ 2009). BMJ 2009. www.bmj.com/content/339/bmj.b2587/reply.

30 Euler-Chelpin M, Njor SH, et al.: (Rapid responses to Jørgensen KJ, Gøtzsche PC BMJ 2009). BMJ 2009. www.bmj.com/cgi/eletters/339/jul09_b2587.

31 Kopans DB: A Major error. BMJ 2009. www.bmj.com/cgi/eletters/339/jul09_1/b2587#217871.

32 International Agency for research on Cancer (WHO), press release no.139 19–03–02: Mammography screening can reduce deaths from breast cancer. www.iarc.fr/en/media-centre/pr/2002/pr139.html.

33 The Swedish Organized Screening Evaluation Group: Reduction in breast cancer mortality from organized service screening with mammography: 1. further confirmation with extended data. And 2. validation with alternative analytic methods. Cancer Epidemiol Biomarkers Prev 2006;15:45–51.

34 Jonsson H, Bordás P, Wallin H, Nyström L, Lenner P: Service screening with mammography in Northern Sweden: effects on breast cancer mortality – an update. J Med Screen 2007;14:87–93.

35 Allgood PC, Warwick J, Warren RM, Day NE, Duffy SW: A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. Br J Cancer 2008;98:206–9.

36 Parvinen I, Helenius H, Pylkkänen L, et al.: Service screening mammography reduces breast cancer mortality among elderly women in Turku. J Med Screen 2006;13:34–40.

37 Schopper D, deWolf C: How effective are breast cancer screening programmes by mammography? Review of the current evidence. Eur J Cancer 2009;45:1916–23.

38 Gabe R, Tryggvadottir L, Sigfusson BF, et al.: A case-control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. Acta Radiol 2007;48:948–55.

39 Roder D, Houssami N, Farshid G, et al.: Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia. Breast Cancer Res Treat 2008;108:409–16.

40 Coldman A, Phillips N, Warren L, Kan L: Breast cancer mortality after screening mammography in British Columbia women. Int J Cancer 2007; 120:1076–80.

41 Van Schoor G, Moss SM, Otten JD, et al.: Increasingly strong reduction in breast cancer mortality due to screening. Br J Cancer 2011;104:910–4.

42 Berry DA, Cronin KA, Plevritis SK, et al.: Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005;353:1784–92.

43 Vervoort MM, Draisma G, Fracheboud J, et al.: Trends in the usage of adjuvant systemic therapy for breast cancer in the Netherlands and its effect on mortality. Br J Cancer 2004;91:242–7.

44 Health Council of the Netherlands: The benefit of population screening for breast cancer with mammography. The Hague: Health Council of the Netherlands, 2002; publication no. 2002/03E. ISBN: 90–5549–420–8.

45 Holland R, Rijken H, Hendriks J: The Dutch population-based mammography screening: 30-year-experience. Breast Care 2007;2:12–8.