Overdiagnosis of Breast Cancer at Screening is Clinically Insignificant

Stephen A. Feig, MD

Long-term follow-up of randomized trials provide the most accurate estimates of overdiagnosis. Estimates from follow-up of service screening studies are almost as accurate if there is sufficient adjustment for lead time and risk status. When properly analyzed data from both of these types of trials indicate that the rate of overdiagnosis at screening mammography is clinically negligible: 0–5%. Population trend studies are a potentially highly inaccurate means to estimate overdiagnosis. Most cases of DCIS detected at screening are medium and high grade with substantial potential to become an invasive disease. To avoid overtreatment, clinicians need to tailor their treatment of DCIS to the histologic and molecular characteristics of each case.

Key Words: Screening mammography; breast cancer screening; overdiagnosis.

As the evidence that screening can substantially reduce breast cancer mortality was being confirmed in numerous studies during the past 30 years, the focus of screening controversies shifted from the existence and measurement of benefit to potential "harms" and costs of screening, as well as to proposals to reduce the frequency of screening and to limit the age of women offered screening to those aged 50–70 years or to deny screening to those with no known risk factors. It is sadly ironic that these issues have gained the forefront of media attention, whereas the lifesaving results of screening have been marginalized. Indeed, women should now be aware that breast cancer mortality among screened women aged 40–69 years in the Swedish Two-County Randomized Trial was reduced by 31% among those invited to screening (1). Randomized trials underestimate the actual benefit from screening due to noncompliance of some study group women and contamination of some control group women. Service screening studies provide higher, more accurate estimates.

Among seven European service screening studies analyzed with incidence-based mortality methods, breast cancer mortality was 25% lower for invited versus not invited women, and 38% lower for screened versus not screened women (2). Among seven other European service screening studies analyzed using case-control methods, corresponding breast cancer mortality reductions were 31% for invited versus not invited women and 52% for screened versus not screened women (2).

The purpose of this review article on overdiagnosis, which has recently gained the spotlight as a purported major harm from screening, was to demonstrate that among screen-detected cancers the possibility of overdiagnosis is extremely low, less than 5%. Furthermore, this review article demonstrates that overdiagnosis has less clinical significance than the vastly larger clinical benefits of early detection established in screening studies. It should also be appreciated that more recent improvements in imaging technology such as 2D digital mammography and 3D digital tomosynthesis should allow even greater benefits than shown in the randomized trials and service screening studies (3).

The concept of overdiagnosis postulates that some breast cancers detected at screening would never be known to the patient or her physician in the absence of screening. It has been alleged that such overdiagnosed breast cancers never produce any clinical signs or symptoms and never represent a cause of death. There is no way to determine by pathologic examination whether an individual cancer has been overdiagnosed. Thus, the existence and frequency of overdiagnosis has only been inferred by mathematical calculation based on trends of breast cancer incidence or on data from screening trials. Yet, such calculations may be grossly misleading if based on basic misassumptions or improper flawed techniques such as insufficient follow-up or failure to correct for risk factors in the populations (4,5). If overdiagnosis does actually occur in the real world, women with overdiagnosed cancers would receive "unnecessary" treatments such as lumpectomy, mastectomy, chemotherapy, and radiation therapy. Additionally, these women would experience the unnecessary anxiety of knowing that they have breast cancer. These women and their families, employers, and medical insurance providers would incur "needless" costs for the consequent diagnostic and therapeutic procedures. Thus, if existent, overdiagnosis would represent harm from screening. The frequency of overdiagnosis would determine whether this harm is trivial or substantial when weighed

Acad Radiol 2015; 22:961–966

From the Department of Radiological Sciences, University of California Irvine Medical Center, 101 City Drive South, Orange CA 92869-3298 (S.A.F.). Received January 12, 2015; accepted January 22, 2015. **Address correspondence to:** S.A.F. e-mail: sfeig@uci.edu

http://dx.doi.org/10.1016/j.acra.2015.01.020

against the benefits from early detection of breast cancer. Overdiagnosis is completely different from a "false-positive biopsy," which is an abnormality that is biopsied on the basis of suspicious imaging findings and subsequently found to be benign on examination of the biopsy specimen.

The frequency of overdiagnosis has been estimated on the basis of data from randomized screening trials, service screening studies, changes and trends in incidence, and stages of cancer in populations. Analytic methods applied to these databases differ substantially in their accuracy and their conclusions. Accurate estimates indicate that the frequency of overdiagnosis is extremely low. Between 0% and 5% of screen-detected cancers are overdiagnosed (6). Inaccurate estimates have led to the erroneous conclusion that as many as 30% of breast cancers are being overdiagnosed. It is understandable that many women may feel confused and frightened and consequently deterred from being screened, and some physicians may be dissuaded from advising screening for their patients. To clarify the controversy, this review article demonstrates why the risk of overdiagnosis has been greatly exaggerated and that the risk is negligible or nonexistent compared to the substantial benefits from screening.

WHY ESTIMATION OF OVERDIAGNOSIS BY MEANS OF TREND STUDIES IS UNRELIABLE AND INACCURATE

In a widely publicized article in the New England Journal of Medicine in late 2012, Bleyer and Welch (7) made the audacious claim that as a result of screening 31% of all breast cancers in the US are being overdiagnosed. To reach this implausible conclusion, the author compared the breast cancer incidence from 1976 to 1978, when very few women were being screened, with the breast cancer incidence from 2006 to 2008, when 60% of women in the US aged 40 and older were being screened at least once every 2 years. The key misassumption in their analysis was that in the absence of screening, breast cancer incidence during that 30-year period would have increased by only 0.25% per year. They ignored the fact that for a longer 40-year period (1940–1980) the increase in breast cancer incidence had been 1% per year, four times greater than used in Bleyer's model (8–10). In fact, breast cancer incidence during 2006–2008 (128 invasive cancers per 100,000 women per year) was actually lower than expected (132 cancers per 100,000 women per year) from a 1% per year increase (10). Thus, temporal comparison of breast cancer incidence rates in the US population is an unreliable method to gauge the frequency of overdiagnosis.

From 1950 to 2014, the breast cancer incidence rate has varied each year and each decade, but the overall trend of increase has been continuous. Changes in diet, lifestyle, and environmental factors are the likely reasons. Screening mammography appears to be a relatively minor reason. In the US, the increased incidence began many years before the screening mammography era. As evident from 30 years of follow-up in the Swedish Two-County Trial, screening did not lead to any cumulative increased incidence in the study group compared to the control group that was not offered screening (11). Rather, screening leads to a temporary increase in incidence because of earlier detection of already existent disease. The claim by Bleyer and Welch that 31% of breast cancers in the US are overdiagnosed is implausible and should have received a more critical journal manuscript review before publication. Equally disturbing is the widespread publicity for the claim provided by television broadcasts, newspaper headlines, and magazine articles. Such irresponsible coverage by the media may discourage women from being screened and persuade medical insurance providers to curtail their support for screening.

Relative changes in incidence rates for early-stage and late-stage breast cancers in the US were the basis for the opinion of Esserman et al. (12) that screening detects an "excessive" number of slow-growing or biologically inert cancers. The authors observed that after the introduction of screening in our country, breast cancer incidence never returned to prescreening levels and that although the incidence of regional cancer has decreased during the screening era, this decrease was less than the increased detection of early-stage disease. The authors acknowledged that breast cancer mortality decreased during the screening era, but were uncertain about the relative contributions of screening versus treatment. The authors fail to appreciate that increased detection of early disease may precede the decrease in late-stage disease by as long as 5–30 years because breast cancer is a chronic disease comprising a whole spectrum from slower growing to rapidly growing tumors. They do not appreciate that the incidence of all regional disease is influenced by many factors, including the overall increasing breast cancer incidence, inadequate screening compliance, and excessively long screening intervals. The breast cancer mortality reduction of 30%–50% seen among women screened in randomized trials and service screening studies cannot be negated by distracting observations requiring the persistence of regional disease. Fully 34% of American women aged 40 years and older have not been screened in the past 2 years and 50% have not been screened annually as recommended by the American Cancer Society (13). In my opinion, the disproportionate number of regional cancers cited by Esserman et al. can be attributed to late-stage cancers among women not being screened often enough or not being screened at all.

ACCURATE ESTIMATION OF OVERDIAGNOSIS USING SERVICE SCREENING REQUIRES SUFFICIENT ADJUSTMENT FOR DETECTION LEAD TIME AND RISK STATUS

Service screening studies compare breast cancer death rates among screened versus nonscreened women but unlike randomized trials, do not randomize individual women (14). Comparison may be made between women offered screening and those in the same geographic area or an adjacent

geographic area not offered screening or with historic death rates (after adjustment for incidence trends) from the same area. Service screening studies also require retrospective adjustments for differences in risk factors to ensure accurate calculations of mortality reduction.

Similarly, accurate estimation of overdiagnosis rates from service screening studies also requires adjustments for differences in risk status between study and control group women, changes in breast cancer incidence over the time periods compared for historic comparison, and correction for lead time bias from earlier detection. Lead time is the length of time by which screening advances the time of detection. Sojourn time is the maximum attainable lead time if detection occurred at the earliest possible time at the moment that the tumor reached the threshold size. Lead time is always less than sojourn time, but approaches sojourn time as screening frequencies become shorter (15).

Additionally, accurate estimation of overdiagnosis from service screening requires at least 25 years of follow-up, including at least 10 years after the last screen, unless there is adequate correction for lead time (5). Failure to adequately correct the lead time and differences in risk factors between study and control groups as well as for changes in incidence rates when historic controls are used will inflate estimates of overdiagnosis.

Estimates of overdiagnosis made by investigators at European service screening programs have varied from 0% to 54% (6). These widely different estimates have been explained by Puliti et al. (6) in a landmark article which found that the 10 studies lacking adequate adjustment for lead time and risk status had mistakenly calculated overdiagnosis rates of 30%–54%. Among the six studies making adequate adjustment, rates of overdiagnosis varied from 1% to 10% (mean = 4.7%). Rates in these six studies included 2.8% in the Netherlands, 1% and 4.6% in two Italian studies, 7% in Denmark, and 10% and 3.3% in England and Wales (16–21).

Similar to trends in the US, European countries also experienced increasing breast cancer incidence in the late 20th century. Among screened women, these trends were accelerated by lead time. For example, if screening confers an average lead time of 2 years, the 2000 incidence would have been observed in 1998 and the 2001 incidence in 1999, and so forth. Screening also accelerates age effects so that the age 52 incidence would have been observed at age 50, and so on. A major surge of additional cancers (beyond those expected from the annual incidence rate) occurs on the initial (prevalence) screen of women screened for the first time. A similar surge of prevalence screen tumors continues each year as new patients enter the screening program. Overdiagnosis is a relatively minor reason for the "excess" incidence observed at screening (4).

LONG-TERM FOLLOW-UP OF RANDOMIZED TRIALS PROVIDES THE MOST ACCURATE ESTIMATES OF OVERDIAGNOSIS

Randomized trials measure long-term follow-up differences in breast cancer death rates between a study group of women offered screening and a control group of women not offered screening represent that gold standard of proof that screening can reduce breast cancer mortality rates. Unlike survival rate studies, randomized trial results are not influenced by three types of study bias: 1) lead time bias, the possibility that screening advances the time of diagnosis but does not alter the time of death. Thus, the measured survival rates are increased but not the mortality rates, 2) Length-biased sampling, the possibility that screening preferentially detects slower growing cancers but misses the faster growing killer cancers that surface clinically between screens as interval cancers, and 3) Self-selection bias, the possibility that women who volunteer for screening have better survival rates primarily because of factors unrelated to the screening process itself such as generally better access to healthcare. Randomized trials circumvent the possible effect of these biases for two reasons (14). First, they measure mortality rates among women offered screening versus not offered screening, rather than between the subsets of women actually screened versus not screened. Secondly, the study and control groups do not differ in any demographic characteristic because they are randomized. For these same reasons, randomized trials also provide a much more accurate measurement of the frequency of overdiagnosis among screened women than obtainable through trend studies or even than through case-control studies.

Direct observation of the cumulative breast cancer detection rates among study and control groups in the Swedish Two-County Screening Trial on 30 years of follow-up was published by Yen et al. (11) and provides the most reliable estimate.

When the Swedish Two-County Screening Trial began, detection rates were initially higher in the study group than in the control group of unscreened women because screening lowered the detection threshold from a larger size when a cancer is palpable to a much smaller size of mammographic visibility. For at least 10 years after screening began, the cumulative incidence of cancers in the study group offered screening exceeds that of the largely unscreened control group. Eventually, the cumulative incidence curves for study and control groups merge. On longer-term follow-up, screening did not increase the cumulative number of cancers but rather hastened their date of appearance to an earlier more favorable stage. The observation by Yen et al. that there was virtually no difference in the cumulative number of invasive and in situ cancers in study and control groups in the Swedish Two-County Trial on 30 years of follow-up indicates that overdiagnosis was nonexistent or negligible among the 38,589 women offered screening. The same observation was made when in situ and invasive diseases were analyzed separately. There was of course a shift from advanced invasive disease (node positive or >20 mm of size or both) to nonadvanced invasive disease (node negative and <20 mm of size). This is the desired outcome from screening.

When the cumulative incidence of cancers in the study group versus the control group on 30 years of follow-up was evaluated according to age at entry into the screening trial, no difference was found for those in any of these three decade groups: 40–49, 50–59, and 60–69. Cumulative cancer

incidence among study group women aged 70–74 years at entry was higher than that for those in the control group, but the difference was not statistically significant because of the small number of women in that age group (11).

OTHER EVIDENCE THAT MOST CASES OF SCREEN-DETECTED DUCTAL CARCINOMA IN SITU DO NOT REPRESENT OVERDIAGNOSIS

Ductal carcinoma in situ (DCIS) is the type of breast cancer, which most frequently comes to mind when the topic of overdiagnosis is discussed and debated. In the premammography era, DCIS was extremely uncommon, representing less than 5% of the annual incidence of breast cancer. Yet today, it accounts for 30% of all breast cancers detected at screening and 20% of all newly diagnosed breast cancers (both screen detected and non–screen detected) in the US (22). The major controversies relate to how frequently screen-detected DCIS would become invasive in the absence of screening and how long this transition would take. Most invasive ductal breast cancers are believed to represent the final step in a sequence of steps from normal disease, hyperplasia, atypical ductal hyperplasia (ADH), DCIS, and invasive ductal carcinoma (IDC) (23,24). In this process, DCIS is considered a nonobligate precursor of potentially fatal breast cancer.

Most IDCs develop from DCIS, but not all cases of DCIS may proceed to invasive cancer during the patient's lifetime. In fact, ductal hyperplasia and ADH, but not DCIS, are potentially reversible to the earlier step.

Because the final goal of screening is to detect breast cancer at the earliest, most curable stage, detection of DCIS would seem to be one of the primary aims of the screening process. Indeed, several studies have shown that with appropriate treatment, screen-detected DCIS has a 20-year survival rate of 99% (25–27). Yet some observers would decry screening detection of DCIS, describing it as a "pseudocancer," which results in unnecessary biopsies, excessive surgery, extra costs, patient anxiety, and other harms.

Several clinical studies indicate that most DCIS detected at screening would otherwise progress to invasive cancer within a short or intermediate timeframe. Tumor histology grade and the presence/absence of necrosis in 445 cases of screen-detected DCIS in the USA were reported by Silverstein et al. (28). Among these, 30% were high grade, 37% were nonhigh grade with no necrosis. In the UK National Health Service Screening Programme, 60% of cases of DCIS were high grade, 20% were intermediate grade, and 20% were low grade (29). It was estimated that if undetected, 84% of high-grade DCIS would progress to invasive disease in 5 years, most intermediate grade DCIS would progress to invasive disease in 10 years, and low-grade DCIS could become invasive in 15 years or longer.

Screen-detected cases of DCIS at multiple European screening programs were evaluated both histologically and by mathematical modeling by Yen et al. (30). These investigators found a substantial difference in the proportion of aggressive versus nonaggressive DCIS on initial (prevalence) versus subsequent (incidence) screens. Among DCIS cases detected on a first screen, 63% were progressive and 37% were nonprogressive. However, on all incidence screens combined, 96% were assessed as likely to become invasive and only 9% appeared to be nonprogressive. This observation has great clinical significance because among women undergoing continuous screening over their lifetime beginning at age 40 years, virtually all screen-detected DCIS would be potentially progressive to invasive and less than 5% of DCIS would be nonaggressive.

Reduction in the subsequent incidence of invasive disease because of detection of DCIS at screening can also be used to indicate the effect of DCIS detection on the natural history of breast cancer. Among women participating in a screening program of limited duration in the UK, 75% of detected cancers were invasive and 25% were DCIS (31). After the start of the program, the detection rates for both invasive disease and DCIS were higher than projected from prescreening incidence trends. After the screening program ended, there was a drop in the incidence of invasive disease for several years to levels lower than projected from prescreening incidence trends. McCann et al. (31) estimated that only 75% of the decreased incidence of invasive disease and mortality reduction was because of earlier detection or invasive cancer and 25% was the result of detection of DCIS. The authors concluded that "cancer for cancer there is as much benefit from detection and treatment of DCIS as from early detection and treatment of invasive cancer."

Screening studies can also be used to estimate the growth rate of DCIS. Sojourn time (mean duration of preclinical disease) can be calculated as a function of prevalence divided by the expected incidence in the absence of screening. Slower growing tumors have longer sojourn times, whereas faster growing tumors are associated with shorter sojourn times. Applying Markov models to data from the Swedish Two-County Screening Trial results, Chen et al. and Tabar et al. have estimated that the mean sojourn time for DCIS is 4.8 years, which is shorter than that for IDC grade 1, but longer than that for IDC grade 2 (2.9 years) and grade 3 (2.2 years) (27,32).

Inferences regarding the growth rates for DCIS can also be made on the basis of the relative size and grade of DCIS among women screened at various intervals. Carlson et al. observed that compared to women screened less frequently, those undergoing annual mammography were more likely to have DCIS tumors which were smaller, had less comedo histology, and were of lower nuclear grade. The findings were consistent with a transition time from smaller, lower grade DCIS to larger, higher grade DCIS of 1–3 years (33). In another study by Feig et al., the progression from in situ to invasive disease was charted according to the tumor size. The ratio of invasive/in situ cancer was graded as pure DCIS (0%), minimal (0%–10%), moderate (11%–74%), and marked (75%–100%). Results suggest that DCIS progresses rapidly to invasive disease in 50% of cases by 1 cm size and to virtually all invasive disease by 2.5 cm (34). These two studies suggesting fairly rapid progression in size and grade

of DCIS and final transition to invasive disease may at first seem inconsistent with its relatively long sojourn time. This apparent paradox may be resolved if one remembers that sojourn time is calculated as a function of prevalence of DCIS at screening, divided by the incidence of DCIS in the absence of screening. DCIS incidence rates are kept low by the relatively rapid transition to invasive disease described by Feig et al., making the sojourn time window of opportunity for mammographic detection of DCIS seemingly longer than in actuality. This transition occurs before most lesions reach a palpable size and explains why DCIS incidence has increased with screening mammography. The paradox is thus resolved by the explanation that DCIS largely represents a preclinical stage of IDC rather than a distinct type of breast cancer or a high-risk marker.

There are several origins for the mistaken notion that DCIS is not a "real" cancer. First, several autopsy studies of women dying from causes other than breast cancer in the 1980s found a 0.2%–15% prevalence of DCIS, higher than the detection rate at screening (35,36). These studies have been used to suggest that many cases of screen-detected DCIS are not clinically significant. There are several flaws in that argument. One is that 50% of these could not be visualized on specimen radiography, so that even fewer would be apparent on mammography of an intact breast. Another weakness in that reasoning is that most of these cases labeled "DCIS" would be reclassified as ADH using current criteria. Finally, among the few remaining cases of true DCIS, virtually all would be borderline ADH or extremely low grade, and noncalcified (another sign of lowest grade) and microscopic rather than mammographic size. Thus, autopsy-detected DCIS is an obsolete diagnosis by current pathology standards, or at the opposite end of the DCIS spectrum, from the type of DCIS detected at screening. As previously discussed in this article, clinical studies indicate that screen-detected DCIS is clinically significant.

Another element in the argument that DCIS is not "real cancer" is cases of DCIS mistakenly diagnosed as ADH on initial pathologic interpretation and subsequently managed by excision along rather than lumpectomy or mastectomy because most of these cases did not have recurrence on follow-up. There are two reasons why these studies should lead to just the opposite conclusion. First, some of these lesions had been completely removed because the initial biopsy margins were sufficiently wide. Second, more important reason is that the case series consisted of only very low-grade DCIS. Even so, Page et al. reported the development of invasive carcinoma in 28% (7 of 10) at 10 years and 36% (9 of 25) at 24 years of follow-up (37,38). Using another case series, Betsill and Rosen found that invasive carcinoma developed in 60% (6 of 10) at 9.7 years, and 60% (9 of 15) at 22 years of follow-up (39,40).

DISCUSSION

The issue of overdiagnosis among screen-detected breast cancers has been widely publicized and sensationalized by the public media, largely on the basis of one flawed study suggesting that 31% of all newly diagnosed breast cancers in the US have been overdiagnosed and that women having these cancers subsequently underwent unnecessary surgery, chemotherapy, and radiation therapy with consequent anxiety for them and excessive costs for our medical care system (7).

So-called "trend studies" such as that article is generally regarded as an unreliable method to measure the effect of screening. On the contrary, evaluation of data from randomized clinical trials (acknowledged as the gold standard method to evaluate screening data), such as the Swedish Two-County Trial, found no evidence of overdiagnosis on 30 years of follow-up of study and control groups (11). Additionally, results from service screening studies may also be used to estimate the frequency of overdiagnosis. Accurate estimation from service screening studies requires sufficiently long follow-up, as well as adjustment for detection lead time and risk status. Using such analysis, the frequency of overdiagnosis at screening has been estimated at 0%–5% (6).

If overdiagnosis does occur even this infrequently, it would be among very old women who die from a cause other than their breast cancer or among women with very low-grade DCIS. Current screening methods cannot reliably distinguish low-grade breast cancers from medium- and high-grade cancers. Thus, the major challenge to reduce the frequency of overdiagnosed cancers would seem to rest with pathologists and clinicians rather than with breast imagers. Clinicians need to implement better compliance with American Cancer Society guidelines, which advise against screening women with fewer than 5 years of life expectancy and/or those with significant comorbid conditions (41). Pathologists can further develop genetic/biomarker tests to assess tumor specimen growth potential (42–44). Imaging is a highly sensitive screening modality but treatment specificity is the responsibility of pathologists, surgeons, medical, and radiation oncologists.

Of paramount importance is the fact that mammography has reduced breast cancer mortality by as much as 50% among screened women (2) and the expectation that newer modalities such as screening with digital tomosynthesis, ultrasound, and magnetic resonance imaging will reduce breast cancer death rates to even further. It would be most unfortunate if women were dissuaded from obtaining screening because of unnecessary concerns regarding overdiagnosis.

REFERENCES

1. Tabar L, Vitak B, Chen THH, et al. Swedish Two-County Trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011; 260:658–663.
2. Broeders M, Moss S, Nystrom L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012; 19(suppl 1):14–25.
3. Feig SA. Digital breast tomosynthesis—the road ahead (invited commentary). Radiographics 2014; 34:E103.
4. Feig SA. Pitfalls in accurate estimation of overdiagnosis: implication for screening policy and compliance. Breast Cancer Res 2013; 15:105.

5. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. Breast Cancer Res 2013; 15:R41.
6. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen 2012; 19:42–56.
7. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast cancer incidence. N Engl J Med 2012; 367:1998–2005.
8. Seidman H, Gelb SK, Silverberg E, et al. Survival experience in The Breast Cancer Detection Demonstration Project. CA Cancer J Clin 1987; 37: 258–290.
9. Garfinkel L, Boring CC, Heath CW, Jr. Changing trends: an overview of breast cancer incidence and mortality. Cancer 1994; 79(suppl):222–227.
10. Kopans DB. The New England Journal of Medicine article suggesting overdiagnosis from mammography screening is scientifically incorrect and should be withdrawn. J Am Coll Radiol 2013; 10:317–319.
11. Yen AMF, Duffy SW, Chen THH, et al. Long-term incidence of breast cancer by trial arm in one county of the Swedish Two County Trial of mammography screening. Cancer 2012; 118:5728–5732.
12. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA 2009; 302:1685–1692.
13. American Cancer Society. Cancer prevention and early detection facts and figures 2009. Atlanta GA: American Cancer, 2009.
14. Feig SA. Methods to identify benefit from mammographic screening of women aged 40-49 years. Radiology 1996; 201:309–316.
15. Morrison AS. Screening in chronic disease. 2nd ed. New York: Oxford University Press, 1992.
16. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, et al. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev 2011; 33:111–121.
17. 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; 45:3166–3171.
18. Paci E, Miccinesi G, Puliti D, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A Service Screening Study in Italy. Breast Cancer Res 2006; 8:R68.
19. Olsen AH, Agbaje OF, Myles JP, et al. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J 2006; 12:338–342.
20. Waller M, Moss S, Watson J, et al. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. Cancer Epidemiol Biomarkers Prev 2007; 16: 2257–2261.
21. Duffy SW, Tabar L, Olsen AH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening from a randomized trial and from the Breast Cancer Screening Programme in England. J Med Screen 2010; 17:25–30.
22. SEER cancer statistics review. Bethesda, MD: National Cancer Institute, 2013.
23. Gallager HS, Martin JE. Early phases in the development of breast cancer. Cancer 1969; 24:1170–1178.
24. Gallager HSW, Martin JE. The study of mammary carcinoma by mammography and whole organ sectioning. Cancer 1969; 23:855–873.
25. Silverstein M. Non-invasive breast cancer: the dilemma of the 1990's. Obstet Gynecol Clin North Am 1994; 21:639–658.
26. Smart CR, Byrne C, Smith RA, et al. Twenty year follow-up of the breast cancers diagnosed during the Breast Cancer Detection Demonstration Project. CA Cancer J Clin 1997; 47:134–149.
27. Tabar L, Duffy S, Vitak B, et al. The natural history of breast cancer. What have we learned from screening? Cancer 1999; 86:449–462.
28. Silverstein MJ, Poller DN, Waisman JR, et al. Prognostic classification of breast ductal carcinoma-in-situ. Lancet 1995; 345:1154–1157.
29. United Kingdom National Health Service. NHS Breast Screening Programme & Association of Breast Surgery at BASO: an audit of screen detected breast cancers for the year of screening April 2007 to March 2008. UK West Midlands Cancer Intelligence Unit; 2009.
30. Yen MF, Tabar L, Smith RA, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 2003; 39:1746–1754.
31. McCann J, Treasure P, Duffy SW. Modeling the impact of detecting and treating ductal carcinoma in situ in a breast screening programme. J Med Screen 2004; 11:117–125.
32. Chen HH, Duffy SW, Tabar L, et al. Markov chain models for progression of breast cancer. Part I: tumor attributes and the preclinical screen-detectable phase. J Epidemiol Biostat 1997; 2:9–23.
33. Carlson KL, Helvie MA, Roubidoux MA, et al. Relationship between mammography screening intervals and size and histology of ductal carcinoma in situ. AJR Am J Roentgenol 1999; 172:313–317.
34. Feig SA, Schwartz GF, Nerlinger R, et al. Prognostic factors of breast neoplasms detected on screening by mammography and physical examination. Radiology 1979; 133:577–582.
35. Nielsen M, Jensen J, Andersen J. Precancerous and cancerous breast lesions during lifetime and at autopsy: a study of 83 women. Cancer 1984; 54:612–615.
36. Nielsen M, Thomsen JL, Primdahl S, et al. Breast cancer and atypia among young and middle-aged women: a study of 110 medicolegal autopsies. Br J Cancer 1987; 56:814–819.
37. Page DL, Dupont WD, Rogers LW, et al. Intraductal carcinoma of the breast: follow-up after biopsy only. Cancer 1982; 49:751–758.
38. Page DL, Dupont WD, Rogers LW, et al. Continued local recurrence of carcinoma 15-25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. Cancer 1995; 76:1197–1200.
39. Rosen PP, Braun DW, Jr, Kinne DE. The clinical significance of preinvasive breast carcinomas. Cancer 1980; 46:919–925.
40. Betsill W, Rosen PP, Lieberman PM, et al. Intraductal carcinoma: long-term follow-up after biopsy along. JAMA 1978; 239:1863–1867.
41. Smith RA, Saslow D, Sawyer KA, et al. American Cancer Society guidelines for breast cancer screening: update 2003. CA Cancer J Clin 2003; 53:141–169.
42. Masood S. Focusing on breast cancer overdiagnosis and overtreatment: the promise of molecular medicine. Breast J 2013; 19:127–129.
43. Colin C, Devonassoux-Shisheboran M, Sardanelli F. Is breast cancer overdiagnosis also nested in pathologic misclassification? Radiology 2014; 273:652–655.
44. Katz SJ, Morrow M. Addressing overtreatment in breast cancer. Cancer 2013; 15:3584–3588.