Mammographic Screening and "Overdiagnosis"[1]

Daniel B. Kopans, MD
Robert A. Smith, PhD
Stephen W. Duffy, MSc

Between 1990 and 2007, the mortality rate from breast cancer in the United States has decreased by 31% (1,2), principally due to the contribution of mammographic screening and improved therapies. This means that, each year, approximately 31% fewer women die of breast cancer than would have died were it not for the progress in the control of the disease. Furthermore, the decline in breast cancer mortality between 1991 and 2006 represents 36.7% of the total decrease in cancer deaths in women, which means approximately 75 500 fewer women died of breast cancer than would have if mortality rates had not declined during the past 20 years (3).

There has been debate over the relative contributions of screening and therapy to this decrease in the death rate. Using computer modeling, a consortium of investigators estimated that mammography was responsible for 28%–65% of the reduction in the mortality rate (4), a wide range of effect estimates attributable to varying model assumptions. Conversely, several large observational studies from Sweden (5–8) and the Netherlands (9) have directly measured the effects of mammographic screening in large general populations, finding that most of the decrease in deaths from breast cancer was due to screening, consistent with the upper end of the modeling estimates (4).

Despite this evidence, some not only deny the important role of screening in the decline of the breast cancer death rate (10) but also question whether the human costs of screening justify what they judge to be a small benefit (11). The issue that has achieved the highest public profile in recent years is "overdiagnosis," with extravagant claims of very high rates of overdiagnosed tumors as a result of screening and rather tenuous theories about overtreated screening-detected breast cancers that would have spontaneously regressed if not detected with mammography receiving the most attention (12–14).

To estimate the efficacy of screening in preventing deaths from breast cancer without distortion by lead time or length bias, randomized controlled trials (RCTs) are required (15). Meta-analyses of the RCTs have demonstrated a significant decrease in breast cancer deaths among women invited to screening from age 40 years onward on the order of 20%–25% (16,17), and, because these were all trials in which women were invited to be screened, the effect of modern mammography among women who actually undergo screening is considerably greater (5–7). It is in the context of a substantial benefit from screening in reducing breast cancer mortality that the overdiagnosis debate should be considered.

Overdiagnosis

The latest challenge to screening mammography has been the argument that screening leads to the diagnosis of a large number of breast cancers that, if left undiscovered, would never become clinically evident and, thus, would never become potentially lethal (13,14). Those who regard overdiagnosis as the greatest harm of screening argue against population-based screening, insisting either that the small number of lives saved is not worth the human cost or that informed and shared decision making should emphasize that, for most women, the harms of screening outweigh the benefits (11). There are two lines of reasoning underpinning this argument. In the first, breast cancers detected during autopsy of women who had not been diagnosed with breast cancer (and who died for other reasons) are often cited as the basis for a prevalence of disease of which a large fraction must be indolent. However, if we screen 1000 women,

we will detect one to two cases of ductal carcinoma in situ (DCIS) (0.1%–0.2%); results of autopsy studies suggest that 9%–15% of women have undiagnosed DCIS at death (18). This indicates that microscopically detected DCIS in cadavers is not relevant to the issue of mammographically detectable DCIS in living women. The second and more common approach to estimating overdiagnosis derives from a comparison of two populations. In these comparisons, one population has been exposed to mammographic screening and has experienced a substantial increase in incidence, whereas the other population, unexposed to mammography, has had a smaller increase in incidence. It has been asserted that the excess incidence represents overdiagnosis and that the treatment of these excess cases is, de facto, overtreatment.

The most commonly accepted definition of overdiagnosis is not a pathologic but an epidemiologic one: the detection with screening of cancer that would not have been diagnosed in the host's lifetime if the screening had not taken place. This is the definition used by the most energetic critics of screening. Because there are no pathologic features that can help differentiate a progressive cancer from a nonprogressive cancer, overdiagnosis is estimated as the difference between observed versus expected incidence. It is this last quantity that is the source of most disagreement among estimates of overdiagnosis.

To prevent deaths from breast cancer, screening must advance the time of diagnosis within the clinically occult phase of the disease's natural history, when treatment has the greatest potential to be successful. Thus, for screening to be effective there must be lead time. This, in turn, implies the possibility of overdiagnosis because there must be some subset of tumors whose lead time exceeds the lifetime of the patient. On the basis of the average lead times and their variability, the number of these tumors must be very small (19). There is disagreement about the size of this population and no way to identify it at this time, but it is intuitively thought that it must largely be represented by DCIS and small invasive cases.

In the broader context of overtreatment, it is accepted that large numbers of breast cancers, including those that are clinically evident, also are "overtreated" insofar as the majority of patients receiving cytotoxic chemotherapy would survive without it and a minority of patients would unfortunately die of breast cancer despite chemotherapy. Between these two extremes is a group of patients with breast cancer whose lives are saved as a result of chemotherapy. However, because we also cannot reliably identify this group, large numbers of women in the other two groups undergo chemotherapy "unnecessarily." Thus, overtreatment is a problem for all breast cancers, and, indeed, overdiagnosis, according to the definition of "indolence," should apply to clinically evident tumors as well.

Spontaneous Regression of Breast Cancers

By definition, an overdiagnosed case is detectable with either screening or palpation. In one model, a breast cancer progresses at an infinitesimally slow rate or progression is arrested; in either case, however, it would not have been detected without screening. In the other model, a breast cancer progresses to a detectable size but then is expected to regress altogether or to a size that is not detectable. We have been unable to find any convincing evidence that a significant fraction of breast cancers regress spontaneously, as implied in some recent publications. Documentation of this phenomenon is mainly limited to a small number of individual case reports (20–22). This suggests that, in one of the most common cancers worldwide, spontaneous regression is a rare phenomenon indeed. In 1976, Lewison (23) reported on regression in 12 patients from his practice; however, despite the apparent "regression" of cancer in the breast, most of these women died of breast cancer. It should also be noted that in those settings—in the developing world, where the natural history of breast cancer is allowed to run its course—the consequences are usually tragic (24).

The most recent controversies about overdiagnosis stem from two articles. In the first article, Zahl et al (14) claim to have found evidence of spontaneous regression in a study of incidence rates in relation to the inception of a screening program in Norway. In the other article, Jørgensen and Gøtzsche (13), from their study of incidence rates in four countries, conclude that there are extremely high rates of overdiagnosis in screening programs. Both studies have serious methodologic errors.

In the first study, Zahl et al (14) compared a cohort of women in the prescreening era (age range in 1992, 50–64 years) with one undergoing screening in the same area (age range in 1996, 50–64 years) and, because the observed incidence appeared to be considerably higher in the latter cohort, concluded that this was evidence of overdiagnosis. However, the incidence of breast cancer was increasing during this time period in Norway, independently of screening, and the cohort of women aged 50–64 years in 1996 has a major component that is the same population as those aged 50–64 years in 1992, only 4 years older. Because the incidence of breast cancer increases with age—and was increasing with each calendar year in Norway at the time—the incidence in the postscreening cohort would be expected to be greater than that in the prescreening cohort *even without screening*. In addition, the statistical modeling did not take into account the complex correlation structure resulting from the same population and, in many cases, the same cancers being included twice in the same data analysis. Thus, incidence is expected to be higher among screened women without overdiagnosis. The authors' sweeping biologic conclusions regarding spontaneous tumor regression are speculative and cannot be taken seriously owing to their fundamentally flawed analysis.

The estimates of overdiagnosis in the second article (13) are equally unreliable due to a preponderance of elementary epidemiologic errors (25). To illustrate, let us take the example of the analysis and conclusions related to the United Kingdom screening program.

Jørgensen and Gøtzsche conclude that the rate of overdiagnosis with the screening program in the United Kingdom from 1993 to 1999 was 57%. The flaws in this estimation include a broad range of unreliable methodologic approaches as well as misplaced or casual assumptions, including:

1. The analysis was not based on actual data; that is, rather than obtaining data from cancer registries, the authors estimated rates from published graphs, rendering the incidence data prone to serious inaccuracy. They also estimated trends with use of the wrong method of analysis (ie, linear rather than Poisson regression).

2. The authors assumed that underlying incidence rates were stable. They observed a 41% increase in invasive cancers in 1999, compared with expected rates based on trends from 1971 to 1984, before screening was initiated. Two observations render this comparison incorrect. First, the authors ignored the fact, noted in the source of their graphical data (26), that an exponential increase in incidence began in 1984—several years before the screening program started. Second, there was a 7% increase in incidence in the age group below the screening age range. Thus, their method underestimated the expected rates in 1993–1999 in the absence of screening.

3. In addition to excluding the later years in the prescreening period, the authors based their overdiagnosis estimate not on their "screening" period of 1993–1999 but on 1999 alone. Thus, they excluded the years of highest incidence in the prescreening period and the years of lowest incidence in the screening period. This extreme selection, in the presence of a well-documented underlying trend of increased incidence, seriously inflates the difference between observed and expected incidence.

4. The authors assumed that 10% of all breast cancer diagnoses are DCIS and that all DCIS is overdiagnosed. Thus, their estimate of a 57% overdiagnosis rate was obtained by combining the 41% increase in invasive cancers and augmenting it with an additional 10% rate of estimated detection of DCIS.

5. The authors made no adjustment for lead time. The authors assumed that all additional incidences after the initiation of screening were overdiagnosed rather than simply diagnosed earlier. Their justification is that they could see no compensatory reduction in incidence above the age range for screening. This is likely due to the absence of direct data and to the fact that their analysis pertains to the early years of the screening program, before a sufficiently large cohort has been through the screening program and emerged into the age group above it. Other investigators who did have access to the direct data and/or to data up to 2004 found a substantial compensatory decrease in incidence above the age range for screening (27,28).

6. The 57% overdiagnosis estimate implies that almost all screening-detected cancers were overdiagnosed (29). Given that a number (fortunately small) of screening-detected tumors prove fatal, the inaccuracy in their overdiagnosis estimate must be considerable.

What Is the True Level of Overdiagnosis?

The most reliable way to evaluate and quantify whether mammography depicts cancers that would never become clinically evident is to observe long-term incidence in an RCT, where there is one cohort of women that is randomly divided and observed during the same period of time. If the division is truly random, and there is no overdiagnosis, approximately the same number of women should be diagnosed with breast cancer over time in both groups. The only difference should be that the cancers in the group of women invited to screening should be found earlier than those in the control group; however, the women in the control group should "catch up" some time after the trial ends and screening has ceased in the study group (19). However, this kind of analysis of the existing breast cancer screening RCT data is difficult because, with the exception of the Malmö trial (30), women in the control arms of the RCTs were offered screening at the conclusion of the trials.

In the Malmö trial, the investigators estimated that screening appeared to have helped detect 10% more cancers than became clinically evident in control subjects with 15 years of follow-up (30). This figure, however, depends on the extent to which women in the study arm stopped being screened as per protocol, which is unknown. Thus, the estimate of 10% overdiagnosis in the Malmö trial would appear to be the maximum amount of "overdiagnosis" that might occur as a result of screening.

In general, estimates of overdiagnosis that take into account lead time and the other factors influencing incidence noted in the previous section are in agreement with or lower than those of the Malmö trial (30). Studies that do not take these factors into account tend to have much higher estimates of overdiagnosis, which is due to some clear, and other perhaps less intuitive, flaws rendering them implausible (13,31). In a review of the data from two RCTs—the Two-County and Gothenburg Trials—Duffy et al (19) estimated an overdiagnosis rate of only 1%. Olsen et al (32) found a similarly low rate in the Copenhagen program, as did Puliti and colleagues (33) in Italy. Thus, RCT estimates and service screening estimates, which properly take into account the complexities of cancer incidence in the presence of screening, tend to find levels of overdiagnosis of 10% or less, which is likely to be closest to the true level.

Some commentators have suggested that overdiagnosis is among the reasons to forgo the potential lifesaving benefits of screening (11,12). Duffy et al (27) evaluated this in a study that looked at the decrease in deaths versus overdiagnosis in the Two-County Trial and the U.K. National Breast Screening Programme in England. The authors found an approximately 30% reduction in mortality and concluded that 2–2.5 lives were saved for every case that was estimated to have been "overdiagnosed."

DCIS

The analyses referred to earlier have dealt either with invasive cancers alone or with combined invasive cancers and DCIS. Before the use of mammography, DCIS made up only 2%–5% of breast

cancers. The only time these lesions were diagnosed was when they were so large that they became palpable. In mammographic screening programs, DCIS now makes up 20%–30% of the cancers detected. There is still debate over the importance of these lesions (34), with interlacing issues of the underlying pathologic characteristics, progressive potential, overdiagnosis, overtreatment, patient awareness, and anxiety driving the debates. Although a thorough discussion of the natural history of DCIS is beyond the scope of this article, the fact remains that there is no direct way to determine which cases of DCIS, if left alone, would not progress to become invasive.

In a retrospective study, Page et al (35) identified 28 women with DCIS who were not treated beyond the diagnostic biopsy. Among these women, seven (25%) developed an invasive breast cancer—with four (14% of the original 28 patients) developing metastatic disease during 15-year follow-up. Because some of the lesions were likely completely excised, and these were low-grade (favorable) lesions, this likely represents the minimum number of DCIS lesions that will progress. More recent studies suggest that the progressive potential of DCIS is higher and depends on whether it is detected on the first mammogram or on subsequent mammograms. Yen et al (36) estimated that 37% of DCIS discovered at the first (prevalence) screening examination and only 4% detected at subsequent screenings would not progress. The authors concluded that "overdiagnosis" of DCIS was a minor phenomenon.

If DCIS is the precursor lesion for invasive breast cancers, then the detection and removal of DCIS from a population should result in fewer subsequent invasive cancers. This is consistent with the results of the Gothenburg and Two-County Trials, where the excess of DCIS in the study arm was largely compensated for by a deficit in invasive disease (19).

In the Two-County Trial, it was estimated that 12% of all prevented deaths resulted from the detection of DCIS (37).

The RCTs of treatment of DCIS give us clear evidence that DCIS generally does have progressive potential and that its detection and treatment are valuable. In one of the largest trials (38), the recurrence rate for surgically resected DCIS, without radiation therapy, was 26.8% at 8 years. Of concern was that half of the recurrences were invasive cancers. Postsurgical radiation therapy reduced the recurrences to 12.1%, with only 3.9% being invasive. In the United Kingdom, Australia, and New Zealand DCIS Trial (39), surgery alone was compared with surgery plus radiation therapy, surgery plus tamoxifen, and surgery plus both radiation therapy and tamoxifen in women undergoing wide local excision for DCIS. The recurrence rate in patients treated with surgery alone was 32%, which is similar to the rates of recurrence observed in operable invasive tumors (39).

In 2007, Ravdin et al (40) noted that the incidence of invasive breast cancer was declining in the United States. They attributed the decline to the marked reduction in exogenous hormone use that had occurred following publication of the results of the Women's Health Initiative Study in 2002. However, the incidence actually began to decrease several years earlier, in 1999, so it is clear that the cessation of hormone use was not solely responsible for the decrease in incidence (41). The decrease in incidence may be due to a return toward baseline after a long "prevalence" hump because women in the United States did not begin screening all at once but with increasing numbers over many years. It is also possible that the removal of many DCIS lesions during the past decades resulted in fewer subsequent invasive cancers (28). Unfortunately, at this point in time it is not possible to determine the contribution, if any, of the detection and treatment of DCIS lesions.

In summary, we recognize that there are almost certainly some breast cancers that will never be lethal. It is also clear that many women, even those with palpable breast cancers, are treated with systemic therapy without our knowing specifically who will benefit (42). This overtreatment is not, however, confined to those cancers detected with mammography. The data clearly show that there are large, clinically evident, palpable cancers that are not lethal.

We completely support the suggestion that all women should be provided with the information they need to decide whether to participate in screening, including the fact that they may be overtreated if cancer is detected ==because we are unable to differentiate nonprogressive and nonlethal cancers from progressive and potentially lethal cancers.== We also must inform women that early detection, although not perfect, has been repeatedly demonstrated to reduce deaths from breast cancer and that the risk of overdiagnosis is small compared with this benefit.

Disclosures of Potential Conflicts of Interest: **D.H.K.** No potential conflicts of interest to disclose. **R.A.S.** No potential conflicts of interest to disclose. **S.W.D.** No potential conflicts of interest to disclose.

References

1. Kopans DB. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2002;94(2):580–581; author reply 581–583.

2. Altekruse S, Kosary C, Krapcho M, et al. SEER Cancer Statistics Review, 1975–2007. Bethesda, MD: National Cancer Institute, 2010.

3. Jemal A, Ward E, Thun M. Declining death rates reflect progress against cancer. PLoS ONE 2010;5(3):e9584.

4. 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(17):1784–1792.

5. Duffy SW, Tabár L, Chen HH, et al. The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer 2002;95(3): 458–469.

6. Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography. I. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev 2006;15(1):45–51.

7. Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. Cancer 2001;91(9):1724–1731.

8. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service

screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. Cancer 2011;117(4):714–722.

9. Otto SJ, Fracheboud J, Looman CW, et al. Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. Lancet 2003;361(9367):1411–1417.

10. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA 2009;302(15):1685–1692.

11. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2011 (1):CD001877.

12. Welch HG, Black WC. Overdiagnosis in cancer. J Natl Cancer Inst 2010;102(9):605–613.

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

14. Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168(21):2311–2316.

15. Kopans DB, Monsees B, Feig SA. Screening for cancer: when is it valid?—Lessons from the mammography experience. Radiology 2003;229(2):319–327.

16. Smith RA, Duffy S, Tabar L. Screening and early detection. In: Barbiera GV, Esteva FJ, Skoracki R, eds. Advanced therapy of breast disease. 3rd ed. Shelton, Conn: People's Medical Publishing House, 2011.

17. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002;137 (5 Part 1):347–360.

18. Welch HG, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? Ann Intern Med 1997;127(11):1023–1028.

19. Duffy SW, Agbaje O, Tabar L, et al. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res 2005;7(6):258–265.

20. Dussan C, Zubor P, Fernandez M, Yabar A, Szunyogh N, Visnovsky J. Spontaneous re-

gression of a breast carcinoma: a case report. Gynecol Obstet Invest 2008;65(3):206–211.

21. Cole WH, Everson TC. Spontaneous regression of cancer: preliminary report. Ann Surg 1956;144(3):366–383.

22. Krutchik AN, Buzdar AU, Blumenschein GR, Lukeman JM. Spontaneous regression of breast carcinoma. Arch Intern Med 1978; 138(11):1734–1735.

23. Lewison EF. Spontaneous regression of breast cancer. Natl Cancer Inst Monogr 1976;44: 23–26.

24. Adesunkanmi AR, Lawal OO, Adelusola KA, Durosimi MA. The severity, outcome and challenges of breast cancer in Nigeria. Breast 2006;15(3):399–409.

25. von Euler-Chelpin M, Njor S, Lynge E. Answer to Jørgensen & Gøtzsche. BMJ [serial online] 2009. http://www.bmj.com.proxy .library.emory.edu/content/339/bmj.b2587 .full/reply#bmj_el_217871. Accessed March 22, 2011.

26. Johnson A, Shekhdar J. Breast cancer incidence: what do the figures mean? J Eval Clin Pract 2005;11(1):27–31.

27. 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 Screening Programme in England. J Med Screen 2010; 17(1):25–30.

28. McCann J, Treasure P, Duffy S. Modelling the impact of detecting and treating ductal carcinoma in situ in a breast screening programme. J Med Screen 2004;11(3):117–125.

29. Lawrence G, O'Sullivan E, Kearins O, Tappenden N, Martin K, Wallis M. Screening histories of invasive breast cancers diagnosed 1989–2006 in the West Midlands, UK: variation with time and impact on 10-year survival. J Med Screen 2009;16(4):186–192.

30. Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP. Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. BMJ 2006;332(7543):689–692.

31. Zahl PH, Strand BH, Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. BMJ 2004; 328(7445):921–924.

32. Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J 2006; 12(4):338–342.

33. Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. Eur J Cancer 2009;45(18):3166–3171.

34. Allegra CJ, Aberle DR, Ganschow P, et al. National Institutes of Health State-of-the-Science Conference statement: diagnosis and management of ductal carcinoma in situ September 22–24, 2009. J Natl Cancer Inst 2010;102(3):161–169.

35. Page DL, Dupont WD, Rogers LW, Landenberger M. Intraductal carcinoma of the breast: follow-up after biopsy only. Cancer 1982;49(4):751–758.

36. Yen MF, Tabár L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 2003;39(12):1746–1754.

37. Duffy SW, Tabar L, Vitak B, et al. The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. Eur J Cancer 2003;39(12):1755–1760.

38. Fisher ER, Dignam J, Tan-Chiu E, et al. Pathologic findings from the National Surgical Adjuvant Breast Project (NSABP) eight-year update of Protocol B-17: intraductal carcinoma. Cancer 1999;86(3):429–438.

39. Cuzick J, Sestak I, Pinder SE, et al. Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: long-term results from the UK/ANZ DCIS trial. Lancet Oncol 2011;12(1):21–29.

40. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med 2007;356(16):1670–1674.

41. Jemal A, Ward E, Thun MJ. Recent trends in breast cancer incidence rates by age and tumor characteristics among U.S. women. Breast Cancer Res 2007;9(3):R28.

42. Osborne CK. Tamoxifen in the treatment of breast cancer. N Engl J Med 1998;339(22): 1609–1618.