Overdiagnosing Overdiagnosis

MARTIN J. YAFFE,^{a,c,d} KATHLEEN I. PRITCHARD^b

^aSunnybrook Research Institute and ^bSunnybrook Odette Cancer Centre, Toronto, Ontario, Canada; ^cDepartment of Medical Biophysics and Medical Imaging, University of Toronto, Toronto, Ontario, Canada; ^dSmarter Imaging Program, Ontario Institute for Cancer Research, Toronto, Ontario, Canada

Disclosures of potential conflicts of interest may be found at the end of this article.

There is solid evidence from both randomized trials and observational studies that screening mammography for women between the ages of 40–69 years contributes to a substantial reduction in mortality from breast cancer [1–3]; however, mammography screening is far from a perfect test. Sensitivity is often reduced in women with very dense breasts, and lack of specificity results in recall of women who do not have breast cancer for further imaging investigation. In addition, concerns have been expressed regarding overdetected and overtreatment. In this issue, Drs. Bleyer and Kopans express very strongly held differences of opinion regarding the degree of overdetected associated with mammography screening. It is clear from their articles that some of the differences in their positions are based on acceptance of different data as being factually accurate, such as the level of mortality reduction attributable to screening. Other differences are related more to values, for example, the weight of the benefit of a life saved through earlier detection versus the harms caused by stress after having been called back after a positive screening examination when cancer is then ruled out or having treatment for ductal carcinoma in situ (DCIS) that may or may not eventually progress to become lethal.

Dr. Bleyer implies that the benefits of screening have been oversold. Dr. Kopans points to a concentrated effort to reduce access to screening. They are both correct. Certainly, in the nature of the oversimplified messages presented to the public, there is usually little balance. Proponents of screening tend to skip over its limitations, whereas others like Gøtzsche proclaim outrageously that “[t]he best method we have to reduce the risk of breast cancer is to stop . . . screening” [4]. There are complexities associated with screening that cannot be adequately explained to the public in brief sound bites, and even many health professionals may miss some of the subtleties.

Perhaps a useful analogy for the lay public is that screening for breast cancer is like having a smoke detector in the house. You spend some money to install the detector (cost and time invested in screening). For most people, there will never be a fire in the house (only about 12% of women will develop breast cancer). If you are lucky in that way, then after the fact, you can think of the investment in the smoke detector as a waste of money. It is also possible that a fire will start and the detector will not go off for some reason (sensitivity of mammography is approximately 80%) or that the fire gets large

so quickly that the alarm does not really help you (a fraction of cancers are too aggressive for screening to be useful). Alternatively, the fire might start in the fireplace, so it does not really matter that the alarm went off because the fire would not have hurt you (some cancers are indolent; these are the overdetected cancers). The alarm might sound in the middle of the night and wake you up, but there is no fire (false-positive screening examinations). The benefits, of course, are that the alarm could wake you up just in time and save your life (mortality reduction attributable to screening) or could even wake you so quickly you can extinguish the fire before any real damage has been done to your house (morbidity and mortality reduction). The equivalents of all these situations occur with mammography screening.

Despite the limitations described, we buy smoke detectors not because we expect a fire but because, by making a modest investment and incurring a certain level of nuisance, we reduce the risk of dying in a fire. The benefits to an individual are unlikely but are very large should the situation happen to occur. The downside is manageable, and, averaged over the population, there is a significant net benefit of lives and property saved. We buy smoke detectors for the overall benefit.

Overdetected is the detection of breast cancers that would not otherwise have surfaced in a woman’s lifetime and, therefore, would not have threatened her life [5]. Overdetected is frequently labeled incorrectly as “overdiagnosis,” but the actual diagnosis of breast cancer is performed by a pathologist after a lesion is detected and usually following additional imaging workup. Overdetected, like any detection of cancer, results in stress, in additional testing, and almost always in at least some degree of treatment that, at least in principle, is unnecessary. The overtreatment that accompanies overdetected constitutes the main harm.

Screening advances the point at which breast cancer can be detected. If the tumor is not detected by screening, then it can continue to grow and possibly metastasize before it eventually surfaces. In general, we expect detected cancers to be more advanced in an unscreened population compared with a screened population. The implementation of screening should result in a shift in the distribution of cancers in the population to lower stages. This is seen clearly in Table 1, which shows the American Joint Committee on Cancer stage distribution of breast cancers detected in 2002 in British Columbia, Canada, for women

Correspondence: Martin J. Yaffe, Ph.D., Sunnybrook Research Institute, 2075 Bayview Avenue, Room S657, Toronto, Ontario, Canada M4N 3M5. Telephone: 416-480-5715; E-mail: martin.yaffe@sri.utoronto.ca. Kathleen I. Pritchard, M.D., Medical Oncology Program, Sunnybrook Odette Cancer Centre, 2075 Bayview Avenue, Room T2 107, Toronto, Ontario, Canada M4N 3M5. Telephone: 416-480-4616; E-mail: kathy.pritchard@sunnybrook.ca Received January 29, 2014; accepted for publication February 3, 2014. ©AlphaMed Press 1083-7159/2014/\$20.00/0 http://dx.doi.org/10.1634/theoncologist.2014-0036

Table 1. Distribution of cancer stages (American Joint Committee on Cancer) in screened and never-screened women

Stage	Screened (%)	Unscreened (%)
I	62	39
II	30	46
III	8	8
IV	0.3	7

Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: British Journal of Cancer, advance online publication, 3 Dec 2013 (doi: 10.1038/bjc.2013.732) [11].

who participated in routine screening in an organized program and for those who were never screened. These distributions are quite different, and downstaging in the screened group is evident.

If a cancer were sufficiently indolent, it might never surface during the woman's lifetime. If such cancers occurred frequently, the increase in the detection of such cancers by screening would result in more early cancers being detected, without an offsetting reduction in the incidence of more advanced disease (because those cancers never had the potential to advance).

If all other factors were constant, a surplus of the increase in early cancers over the decrease in advanced cancers after implementation of screening would constitute proof of overdetected; however, all other factors are not constant.

The underlying incidence of breast cancer has varied over time (Fig. 1), as have the demographics. In the U.S., incidence rose steadily between 1980 and 2002, then fell sharply over a few years, and more recently has begun to gradually increase again [6, 7]. To estimate the amount of overdetected, an accurate correction for changes in baseline incidence is critical; any error in this correction will bias the estimate. In addition, the lead time associated with screening will shift both the age of incidence and the calendar year in which a cancer is detected, often by several years, and if uncorrected, this provides another potential source of error.

In their paper, Bleyer and Welch estimated the amount of overdetected using historical data on breast cancer diagnosis by stage from the Surveillance, Epidemiology, and End Results (SEER) registries [8]. They extrapolated the temporal trend in incidence rates from 1976–1978, a time when screening was not widely used, to 2006–2008, both for early stage and more advanced cancers. Because they expected the incidence rates to be markedly increased by the onset of screening mammography, they inferred the temporal increase using incidence data from women younger than age 40, the age at which screening typically commences. They compared the predicted values (from the extrapolation) with the actual incidence of cancers and found that the actual incidence of earlier cancers exceeded the predictions, presumably because of screening. They also observed that although there was a drop in advanced cancers, it was not as large as the increase in earlier disease, and the disparity between the two differences was ascribed to overdetected.

In fact, the Bleyer-Welch method is actually simpler than they described. If the box at left in Figure 2 represents overall incidence in women older than age 40 in 1976–1978 and the inferred underlying increase suggests that the incidence without screening in 2006–2008 would be represented by the larger (unshaded) box at the upper right, then according to their method, any observed surplus (the shaded region) beyond this value would be attributed to overdetected regardless of change in stage distribution or characteristics within a stage.

This is where the extrapolation becomes key. We observe from Figure 1 that there is not a smooth change in cancer incidence over time; rather, it fluctuates because of factors such as those mentioned by both authors: the uptake and tapering off of participation in screening and possibly patterns of use of hormone replacement therapy. As both authors also indicated, the patterns have geographical differences, varying among the SEER registries, and there are sociodemographic differences in incidence in a population with composition that has changed over time. There is also an implied assumption that the temporal trends of incidence in women younger than age 40 reliably predict what would happen in the absence of screening over the age range 40–85 years, which includes pre-, peri- and postmenopausal years. Finally, the SEER data include multiple incidences of breast cancer for a woman if they occur over her lifetime, and it appears that these were not accounted for in the Bleyer-Welch analysis.

In his article, Dr. Kopans expresses concerns about the extrapolation technique used, the data on which the extrapolation was based, and the lumping together of DCIS data with those for invasive cancer. It may be useful to consider overdetected related to DCIS separately from that associated with invasive cancer because the management issues are somewhat different, but in fact there is a 28% risk of recurrences after conservative treatment of DCIS without radiation; about half of these recurrences surface as invasive disease.

Puliti et al. [9] examined 16 estimates of overdetected from 13 European observational studies. They found that the estimates varied widely depending on whether corrections for baseline incidence trends and lead time were performed. High estimates were obtained when these corrections were not applied, but after correction, overdetected estimates fell into the range of 1%–10%. It is not really possible to correct for screening lead time in the method used by Bleyer and Welch because, in the SEER data, it is not known which cancers (or even what fraction of the cancers) were detected through screening.

Not knowing who was screened raises another, perhaps more important issue. Dr. Bleyer points out that the hypothesis underlying screening is that, through earlier detection, therapy can be more effective in reducing deaths from breast cancer. But screening can only contribute to reduced mortality for those women who are actually screened.

The SEER database entries can be considered as having two components: cancers diagnosed in women who participated in regular screening (left side of the box at lower right in Figure 2) and those found in women who were unscreened or who attended screening only sporadically (right side). The distributions of earlier and later stage cancers should be different in the two groups, but the cancers are not identified by the group to which they pertain. Consequently, inferences made from the database can be applied only to this composite. If we use the stage

Figure 1. Surveillance Epidemiology and End Results 9 incidence and U.S. mortality, 1975–2010, for females of all races. Rates are age adjusted.

Figure 2. The size of box (left) represents breast cancer incidence from 1976 to 1978 for women older than age 40. It is assumed that very little screening is occurring. The unshaded region (upper right) represents the expected incidence (by extrapolation of data for younger women using the methods of Bleyer and Welch) that would occur in the absence of screening. Excess of observed incidence compared with expected incidence (shaded region) is attributed to overdiagnosis. Consideration of well-screened and unscreened or poorly screened subcomponents in the Surveillance Epidemiology and End Results data (lower right).

Abbreviations: E, early stages; L, later stages.

distributions for screened and unscreened women from Table 1 and perform a weighted fit to try to reproduce the distribution of cancer stage seen in SEER for 2002, the best fit is consistent with a population in which screening participation is 47% rather than the 70% suggested by the self-reported data quoted by Bleyer and Welch [8]. Furthermore, screening participation never appears to reach a higher level. This suggests that half of the women in the analysis by Bleyer and Welch would not have had access to the potential benefits typical of organized screening. A substantial reduction in later stage cancers in the screened component of the SEER population could easily be diluted by a lesser reduction or even possibly by an increase in such cancers in the unscreened component.

Inferences from distributions of cancers in the registry do not necessarily reflect what is happening in a screened population in that the data are influenced by cancers detected in women who do not receive screening and some who may not even avail themselves regularly of health care. What the SEER data do tell us and what can legitimately be concluded from the paper by Bleyer and Welch [8] is that the demonstrated benefits from cancer downstaging seen in randomized trials and in studies of organized screening programs have not been uniformly experienced in the U.S.

Where might we find common ground? Drs. Bleyer and Kopans agree that there is room for improvement in screening, certainly for improving the sensitivity in detecting invasive cancers and the specificity of the screening examination so that there are fewer abnormal recalls of women without cancer. Better ability to distinguish between cancers with the potential to become lethal and those destined to be indolent would allow overtreatment to be reduced. This is especially important with respect to DCIS. These goals motivate further research in methods for detection and characterization of breast cancer.

Until new developments have been established, it is critical that we make effective use of the tools that we have. At the same time, we must find better ways to communicate candidly both the strengths and the limitations or harms associated with screening, so women can make informed decisions regarding participation in screening.

For example, women should become familiar with the idea that, to rule out the possibility of breast cancer with high confidence using currently available tests, it would not be unusual to require additional imaging to be performed after the initial screening examination and that this does not indicate a high likelihood of cancer. Normalization of this message would be a good first step to reduce the stress associated with screening recalls.

Meanwhile, the combination of earlier detection through routine screening and modern therapy provides a means of reducing breast cancer mortality and/or morbidity. We and others [10] are convinced that the benefits of screening outweigh the risks. Furthermore, we feel that some of the charged language used and the extreme positions taken in the discussion of screening distract from coherent communication of the complex issues surrounding this topic. More balanced discussion would better serve the public.

DISCLOSURES

Martin J. Yaffe: Ontario Breast Screening Program (C/A); GE Healthcare (RF); Sunnybrook Health Sciences Centre, Ontario Institute for Cancer Research (E); Mammographic Physics Inc. (OI); **Kathleen I. Pritchard:** Novartis, Roche, AstraZeneca, Pfizer, Boehringer-Ingelheim, GlaxoSmithKline, Sanofi, Ortho-Biotech, Amgen, Bristol-Myers Squibb (C/A); Novartis, Roche, AstraZeneca, Pfizer, Boehringer-Ingelheim, GlaxoSmithKline, Sanofi, Ortho-Biotech, Amgen, Bristol-Myers Squibb (H).

(C/A) Consulting/advisory relationship; (RF) Research funding; (E) Employment; (ET) Expert testimony; (H) Honoraria received; (OI) Ownership interests; (IP) Intellectual property rights/inventor/patent holder; (SAB) Scientific advisory board

REFERENCES

1. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380:1778–1786.
2. Duffy SW, Yen AM, Chen TH et al. Long-term benefits of breast screening. *Breast Cancer Manage* 2012;1:31–38.
3. Nelson HD, Tyne K, Naik A et al. Screening for breast cancer: An update for the U.S. Preventive Services Task Force. *Ann Intern Med* 2009;151:727–737, W237–42.
4. Gøtzsche PC. Time to stop mammography screening? *CMAJ* 2011;183:1957–1958.
5. Breast Cancer Screening. Vol 7: Lyon, France: International Agency for Research on Cancer, 2002.
6. DeSantis C, Ma J, Bryan L et al. Breast cancer statistics, 2013. *CA Cancer J Clin* 2014;64:52–62.
7. National Cancer Institute, Division of Cancer Control and Population Science. Surveillance, Epidemiology, and End Results (SEER) program, SEER9 Registries, 1975–2010. Available at <http://seer.cancer.gov/statfacts/html/breast.html>. Accessed February 3, 2014.
8. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med* 2012;367:1998–2005.
9. 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(suppl 1):42–56.
10. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: An independent review. *Lancet* 2012;380:1778–1786.
11. Coldman AJ, Phillips N. Breast cancer survival and prognosis by screening history. *Brit J Cancer* 2013 [Epub ahead of print].

EDITOR'S NOTE: See the related articles by Daniel J. Kopans on pages 107-112 and Archie Bleyer on pages 113-126.