ORIGINAL ARTICLE

Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness

H. Gilbert Welch, M.D., M.P.H., Philip C. Prorok, Ph.D., A. James O'Malley, Ph.D., and Barnett S. Kramer, M.D., M.P.H.

ABSTRACT

BACKGROUND

The goal of screening mammography is to detect small malignant tumors before they grow large enough to cause symptoms. Effective screening should therefore lead to the detection of a greater number of small tumors, followed by fewer large tumors over time.

METHODS

We used data from the Surveillance, Epidemiology, and End Results (SEER) program, 1975 through 2012, to calculate the tumor-size distribution and size-specific incidence of breast cancer among women 40 years of age or older. We then calculated the size-specific cancer case fatality rate for two time periods: a baseline period before the implementation of widespread screening mammography (1975 through 1979) and a period encompassing the most recent years for which 10 years of follow-up data were available (2000 through 2002).

RESULTS

After the advent of screening mammography, the proportion of detected breast tumors that were small (invasive tumors measuring <2 cm or in situ carcinomas) increased from 36% to 68%; the proportion of detected tumors that were large (invasive tumors measuring ≥2 cm) decreased from 64% to 32%. However, this trend was less the result of a substantial decrease in the incidence of large tumors (with 30 fewer cases of cancer observed per 100,000 women in the period after the advent of screening than in the period before screening) and more the result of a substantial increase in the detection of small tumors (with 162 more cases of cancer observed per 100,000 women). Assuming that the underlying disease burden was stable, only 30 of the 162 additional small tumors per 100,000 women that were diagnosed were expected to progress to become large, which implied that the remaining 132 cases of cancer per 100,000 women were overdiagnosed (i.e., cases of cancer were detected on screening that never would have led to clinical symptoms). The potential of screening to lower breast cancer mortality is reflected in the declining incidence of larger tumors. However, with respect to only these large tumors, the decline in the size-specific case fatality rate suggests that improved treatment was responsible for at least two thirds of the reduction in breast cancer mortality.

CONCLUSIONS

Although the rate of detection of large tumors fell after the introduction of screening mammography, the more favorable size distribution was primarily the result of the additional detection of small tumors. Women were more likely to have breast cancer that was overdiagnosed than to have earlier detection of a tumor that was destined to become large. The reduction in breast cancer mortality after the implementation of screening mammography was predominantly the result of improved systemic therapy.

From the Dartmouth Institute for Health Policy and Clinical Practice, Lebanon (H.G.W., A.J.O.), and the Departments of Medicine (H.G.W.) and Biomedical Data Science (A.J.O.), Geisel School of Medicine, Hanover — both in New Hampshire; and the Division of Cancer Prevention, National Cancer Institute, Bethesda, MD (P.C.P., B.S.K.). Address reprint requests to Dr. Welch at the Dartmouth Institute for Health Policy and Clinical Practice, 35 Centerra Pkwy., HB 7251, Lebanon, NH 03766, or at h.gilbert.welch@dartmouth.edu.

N Engl J Med 2016;375:1438-47.
DOI: 10.1056/NEJMoa1600249
Copyright © 2016 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

Although it may be possible to show the efficacy of screening mammography in reducing cancer-specific mortality in the relatively controlled setting of randomized trials, those trials may not accurately reflect the actual effectiveness of screening when it is used in clinical practice. Differences between efficacy and effectiveness with respect to the benefit of screening may be particularly stark when the treatments administered in practice have markedly changed from those administered in the trials that led to the implementation of widespread screening. Furthermore, although trial data may provide an assessment of some negative consequences of screening, such as false positive results and associated diagnostic procedures, such assessments may understate what actually occurs when screening is implemented in the general community. The collection of data regarding other harms, such as overdiagnosis (i.e., tumors detected on screening that never would have led to clinical symptoms), requires additional long-term follow-up of trial participants, and those data are often either not available or they reflect patient follow-up and testing practices from decades earlier.

One response to these challenges in the assessment of the population effects of screening mammography has been microsimulation modeling. The output of statistical models has the appeal of quantitative precision, but the precision may be more apparent than real.[1] Modeling is only as good as its data inputs and underlying assumptions, particularly those regarding the (unobserved) natural history of tumors detected with the use of screening mammography. Not surprisingly, different models elicit a wide range of results; in the models used by the Cancer Intervention and Surveillance Modeling Network, for example, the estimates of the contribution of screening to the observed reduction in breast-cancer mortality ranged from as little as 28% to as much as 65%.[2]. Furthermore, the complexity of modeling limits the ability of peer reviewers and journal readers to assess the validity of the approach.

To assess the potential mortality benefit and the potential harm of overdiagnosis associated with breast-cancer screening, we used a transparent approach in which the objective was to approximate the magnitude of these effects rather than to attempt precise estimation. We used population data on tumor size, a variable that has been collected for decades and is a direct proximate indicator of screening effect.

Although the biologic characteristics of a tumor are now recognized to be more relevant to breast-cancer prognosis than the size of the tumor, tumor size is more relevant to the assessment of the proximate effect of screening. Screening mammography is not an assessment of functional gene expression; rather, it is an anatomy-based search for small structural abnormalities that are too small to be felt. Thus, the ultimate goal of reduced cancer-specific mortality must be mediated through tumor size at diagnosis. In this analysis, we used trends in malignant breast-tumor size to approximate the contribution of screening mammography to the reduction in breast-cancer mortality and to estimate the magnitude of overdiagnosis.

METHODS

OVERVIEW

To assess the effectiveness of screening mammography, we examined trends in breast-tumor size at diagnosis. We started with the assumption that the underlying probability that clinically meaningful breast cancer would develop was stable, an assumption we believe was warranted given the stable incidence of metastatic breast cancer for more than three decades, despite spanning the era of increasing prevalence of screening-mediated breast cancer and changing patterns of hormone therapy (Fig. 1).

All the analyses were performed in the same study population: women 40 years of age or older at nine long-standing sites of the Surveillance, Epidemiology, and End Results (SEER) program, which represents approximately 10% of the population of the United States.[4] The SEER program is the population-based registry for incident cancers in the United States. It is broadly representative of the nation as a whole; SEER-based estimates of breast-cancer mortality are virtually identical to those ascertained from U.S. mortality data,[5] and the SEER program has had virtually complete case ascertainment and reporting for decades.[6] The study period, 1975 through 2012, spans the time periods before and after the widespread dissemination of screening mammography. All population rates are age-adjusted to the standard population of the United States in 2000.

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

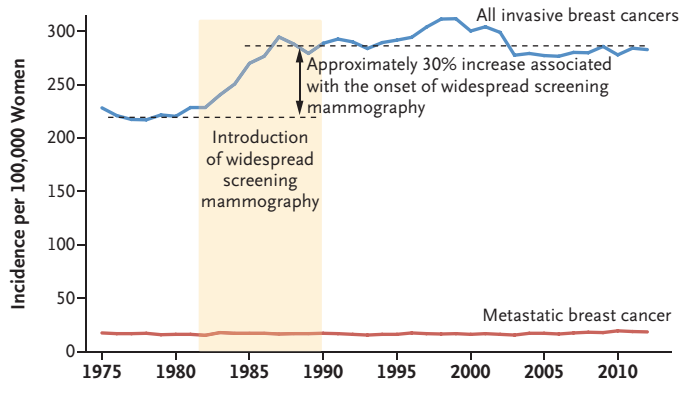

Figure 1. Temporal Relationship between the Introduction of Screening Mammography and Increased Incidence of Invasive Breast Cancer.

Shown are the incidences of overall invasive breast cancer and metastatic breast cancer among women 40 years of age or older at nine sites of the Surveillance, Epidemiology, and End Results (SEER) program, during the period from 1975 through 2012. The use of screening mammography was rare before 1980 (as evidenced by the rarity of ductal carcinoma in situ — an abnormality that is nearly always detected by mammography rather than by breast self-examination, physical examination, or the development of symptoms), yet its use had disseminated to over half of women 40 years of age or older by 1990 (as determined by responses to a National Health Interview Survey question in which women were asked if they had had a mammogram in either 1988 or 1989[3]).

MEASURES

Tumor-Size Distribution and Size-Specific Incidence

We classified the recorded size of invasive breast tumors in five categories (details regarding these categories are provided in Section 1 in the Supplementary Appendix, available with the full text of this article at NEJM.org). In situ carcinomas were included as a separate category. The denominator for the determination of tumor-size distribution was the number of women with a diagnosis of breast cancer; the denominator for the determination of size-specific incidence was the number of women in the study population. In both cases, the numerator was the number of women with breast cancer within each size category.

Missing data with respect to tumor size decreased with time; missing data were common in the early years (33% of tumors were of unknown size in the period from 1975 through 1979), then became less common (5% of tumors were of unknown size in the period from 2008 through 2012). If we had directly calculated size-specific incidence by excluding tumors of unknown size, this decreasing frequency of missing data on size would have produced a spuriously low baseline incidence followed by a spuriously large increase, which would have led us to overestimate overdiagnosis and underestimate the contribution of screening to lowering mortality. To avoid this bias, we used inverse-probability weighting to calculate the tumor-size distribution. The data for each woman with a known tumor size were weighted by the reciprocal of the probability that similar women — those with identical values of observed characteristics — had tumors that were of a known size (details of this analysis are provided in Section 2 in the Supplementary Appendix). Size-specific incidence was then calculated by multiplying the proportion of tumors in the specific size category by the overall incidence of invasive breast cancer.

Ten-Year Risk of Death from Breast Cancer

We calculated the 10-year risk of death from breast cancer (case fatality rate) for two time periods: a baseline period before the advent of widespread screening mammography (1975 through 1979) and a period encompassing the most recent years for which 10 years of follow-up data were available (2000 through 2002). The denominator for the determination of case fatality rate was the number of women who received a diagnosis of breast cancer at the beginning of a 10-year period, and the numerator was the number of deaths from breast cancer within 10 years after diagnosis.

APPROXIMATIONS

Magnitude of Overdiagnosis

An increased incidence of small tumors is an early indicator of screening effect that could be the result of either effective screening or overdiagnosis. Assuming a stable underlying incidence of disease burden and no overdiagnosis of tumors, the additional detection of small tumors should be accompanied by a corresponding decrease in large tumors over time. In other words, the potential benefit of screening is to identify women in whom larger tumors are destined to develop and to make the diagnosis of the cancer earlier, when their tumors are still small. A decrease in the incidence of larger tumors suggests that earlier detection is occurring — a necessary, but not sufficient, condition for screening to result in lower mortality (with the second condition being that earlier treatment of these

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

tumors must be more effective than treatment after clinical presentation). The extent to which diagnosis of additional smaller tumors exceeds the decrease in the incidence of larger tumors approximates the magnitude of overdiagnosis in the population.

Relative Contribution of Improved Cancer Treatment versus Screening

Analyses of the relative contribution of improved treatment of breast cancer versus screening to lowering breast-cancer mortality were limited to the larger tumors (invasive tumors measuring ≥2 cm). This restriction was established for two reasons: the potential benefit of screening in lowering mortality should be mediated largely through the reduction in the incidence of the larger tumors, and the case fatality rate is a relatively unbiased estimate of the treatment effect in larger tumors because those tumors are detected predominantly clinically, which minimizes biases associated with lead time, length, and overdiagnosis. Lead-time bias refers to the overestimation of the duration of survival among women with screening-detected tumors (relative to tumors detected by signs and symptoms) when survival is measured from diagnosis. Length bias refers to the overestimation of the duration of survival among women with screening-detected tumors, with the overestimation caused by the relative excess of cases that progress slowly; these cases are disproportionately identified by screening because the probability of detection is directly proportional to the length of time during which they are detectable. Overdiagnosis bias refers to the overestimation of the duration of survival among women with screening-detected tumors, with the overestimation caused by the inclusion of "pseudodisease" — subclinical disease that would not become overt before the patient dies of other causes. (Further details regarding these biases can be found at http://ecp.acponline.org/journals/ecp/marapr99/primer.pdf.)

The contribution of improved treatment to lowering breast-cancer mortality in the absence of screening was approximated by holding size-specific incidence constant at prescreening levels and applying the reduction in size-specific case fatality rate over time periods of increasingly effective systemic therapy. The contribution of screening was approximated by applying the reduction of size-specific incidence to a constant size-specific case fatality rate. To explore the effect of screening both before and after improvement in therapy, we performed the latter calculation twice, according to the case fatality rate associated with older therapy (1975 through 1979) and the rate associated with more recent therapy (2000 through 2002).

RESULTS

TUMOR-SIZE DISTRIBUTION AND SIZE-SPECIFIC INCIDENCE

The shift in the size distribution of breast tumors associated with the widespread use of screening mammography is shown in Figure 2A. Although large tumors predominated in the period before the advent of screening, small tumors predominated after its implementation. From 1975 to 2012, the proportion of breast tumors that were small (invasive tumors measuring <2 cm or in situ carcinomas) increased from 36% to 68%, and the proportion of large tumors (invasive tumors measuring ≥2 cm) decreased from 64% to 32%.

However, this shift in size distribution was less the result of a substantial decrease in the incidence of large tumors and more the result of substantial increases in the detection of small tumors(Fig. 2B). Nevertheless, modest decreases were seen in the incidence of large tumors. The changes in size-specific incidence of breast cancer after the introduction of screening mammography are shown in Table 1. The incidence of large tumors decreased by 30 cases of cancer per 100,000 women (from 145 to 115 cases of cancer per 100,000 women), and the incidence of small tumors increased by 162 cases of cancer per 100,000 women (from 82 to 244 cases of cancer per 100,000 women). Assuming that the underlying burden of clinically meaningful breast cancer was unchanged, these data suggest that 30 cases of cancer per 100,000 women were destined to become large but were detected earlier, and the remaining 132 cases of cancer per 100,000 women were overdiagnosed (i.e., 30 subtracted from 162).

SIZE-SPECIFIC CASE FATALITY RATE

The size-specific case fatality rates during the baseline period before the introduction of screening mammography and the period encompassing the most recent years for which 10 years of

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

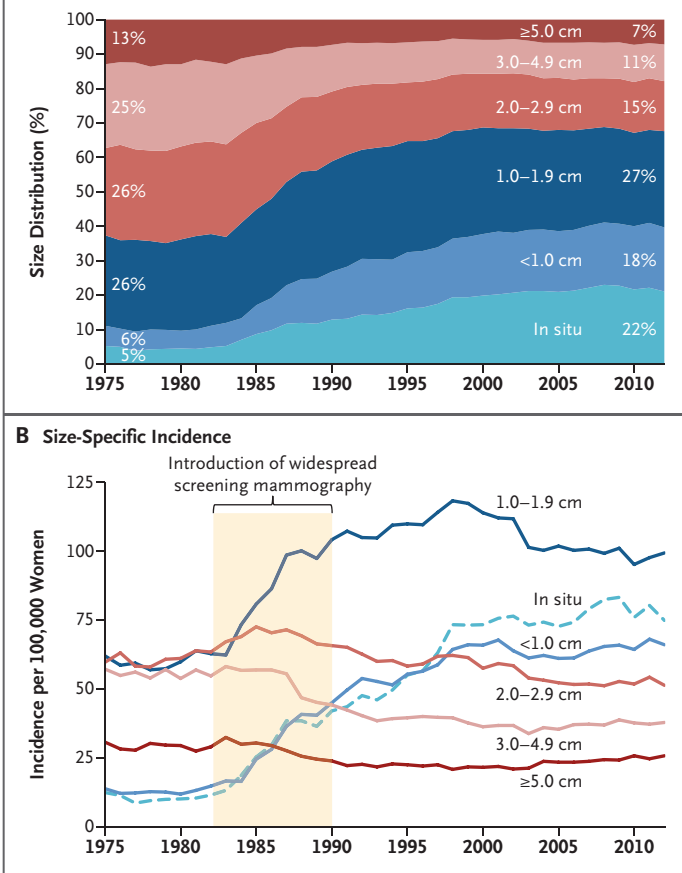

Figure 2. Breast-Cancer Tumor-Size Distribution and Size-Specific Incidence among Women 40 Years of Age or Older in the United States, 1975–2012.

Panel A shows the shift in the size distribution of breast tumors over time. The percentages along the left side of the panel represent the size distribution during the period from 1975 through 1979 (before the widespread use of mammography screening) and those along the right side represent the period from 2008 through 2012. Larger tumors are shown in shades of red, and smaller tumors in shades of blue. Percentages may not sum to 100 because of rounding. Panel B shows the size-specific incidence of breast cancer per 100,000 women.

follow-up data were available are shown in Figure 3. For large tumors, the declining case fatality rate predominantly reflected improved treatment. For small tumors, however, the declining case fatality rate was biased by the combined effect of lead time, length, and overdiagnosis. In fact, during the period from 2000 through 2002, women with in situ carcinomas or those with invasive tumors measuring less than 1 cm had 10-year relative survival rates that exceeded 100% — meaning that they were more likely than age-matched women in the general population to survive. Relative survival refers to the ratio of the proportion of survivors in a cohort of patients with cancer to the proportion of survivors in a comparable set of persons free from cancer. Details of this analysis are provided in Section 3 in the Supplementary Appendix.

The approximate effect of improved treatment of breast cancer on mortality had screening mammography not occurred is shown in Table 2. The estimated reduction in mortality owing to treatment alone was approximately 17 deaths per 100,000 women.

EFFECTS OF SCREENING MAMMOGRAPHY ON MORTALITY

The effect of screening mammography on mortality given previously available therapies and more recent available therapies is also shown in Table 2. In this analysis, the reduction in the incidence of large tumors was attributed to screening, and this reduction was assumed to translate directly to a reduction in mortality. During the period of previously available therapy, the reduction in mortality as a result of screening mammography was approximately 12 deaths per 100,000 women. As treatment improved, the benefit of early detection of tumors necessarily diminished: the reduction in mortality as a result of screening during the period of more recent therapy was approximately 8 deaths per 100,000 women. Thus, improved treatment was responsible for at least two thirds (i.e., 17 divided by the sum of 17 and 8) of the reduction in breast cancer mortality.

DISCUSSION

Decisions about cancer prognosis and therapy have historically been guided by anatomy — the size of the tumor and the extent of disease. However, it has become increasingly clear that the biologic characteristics of the tumor are more relevant to breast-cancer prognosis than the size of the tumor.[7] Tumor size is, at best, a very crude manifestation of underlying biologic characteristics. A recent prospective trial involving women with breast cancer showed that the prognoses of those whose tumors had favorable molecular features were similar regardless of whether their tumors measured greater than or less than 2 cm.[8] Although few clinicians would question that nodal status is a far better indica-

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

Table 1. Change in Size-Specific Incidence of Breast Cancer among Women 40 Years of Age or Older after the Introduction of Screening Mammography.*

Tumor Size	Size-Specific Incidence per 100,000 Women†		
	1975–1979	2008–2012	Change
Large tumors			
≥5.0 cm	29	25	–4
3.0 to 4.9 cm	56	38	–18
2.0 to 2.9 cm	60	52	–8
Total (95% CI)	145 (144 to 147)	115 (114 to 116)	–30 (–28 to –33)
Small tumors			
1.0 to 1.9 cm	59	99	40
<1.0 cm	13	66	53
In situ	10	79	69
Total (95% CI)	82 (81 to 83)	244 (243 to 245)	162 (160 to 164)

* The increase in the incidence of small tumors exceeded the decrease in the incidence of large tumors by 132 per 100,000 women (95% confidence interval [CI], 129 to 135), which suggests overdiagnosis.

† Tumor-size data were missing for a third of cases during the period from 1975 through 1979 and 5% of cases during the period from 2008 through 2012. Inverse-probability weighting was used to address missing data on tumor size. The data for each woman with a known tumor size were weighted by the reciprocal of the probability that similar women — those with identical values of observed characteristics — had tumors that were of a known size (details of this analysis are provided in Section 2 in the Supplementary Appendix). The period from 1975 through 1979 was the baseline time period, which refers to the period before the advent of widespread screening.

tor of metastatic potential and biologic aggressiveness than tumor size, some even question whether advances in tumor biology will supplant the need to determine lymph-node status.[9] However, while clinicians have moved on to focus on tumor biology, breast-cancer screening has remained rooted in anatomy.

The immediate focus of screening continues to be the detection of small lesions; in fact, the detection rate of so-called "minimal tumors" (i.e., invasive tumors measuring <1 cm or in situ carcinomas) is used as an indicator of the quality of mammography.[10] However, the detection of additional small tumors is helpful only when it is accompanied by a reduction in the presentation of larger tumors. A reduction in the number of large tumors may, in turn, reduce the incidence of late-stage disease. Because tumor stage is a more important predictor of mortality than tumor size, trends in stage-specific incidence have been used in previous research that evaluated the effectiveness of screening.[11-13] A reasonable concern was identified, however, that the disease stage (e.g., node-positive disease) may be subject to "upstaging" over time as technology and practice change (e.g., sentinel-node biopsy).[14] To avoid the potentially confounding effect of disease-stage migration, we focused on tumor size, which has the advantage of being a metric that remains constant over time.

Our analysis of size-specific incidence highlights the fact that the introduction of screening mammography has produced a mixture of effects. A modest decrease in the incidence of large tumors (≥2 cm) was observed, which suggests that screening has had the desired effect of advancing the time of diagnosis of some tumors that were destined to become large. At the same time, a much larger increase in the incidence of small tumors (<2 cm) was observed, which suggests that screening has had the undesired effect of detecting tumors that otherwise would not have become clinically apparent. The magnitude of the imbalance indicates that women were considerably more likely to have tumors that were overdiagnosed than to have earlier detection of a tumor that was destined to become large.

Screening can result in the harm of overdiagnosis yet simultaneously result in the benefit of lower breast-cancer mortality. To focus on the effect of screening that would most likely reflect its benefit — a decline in the incidence of large tumors — we approximated the relative contri-

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

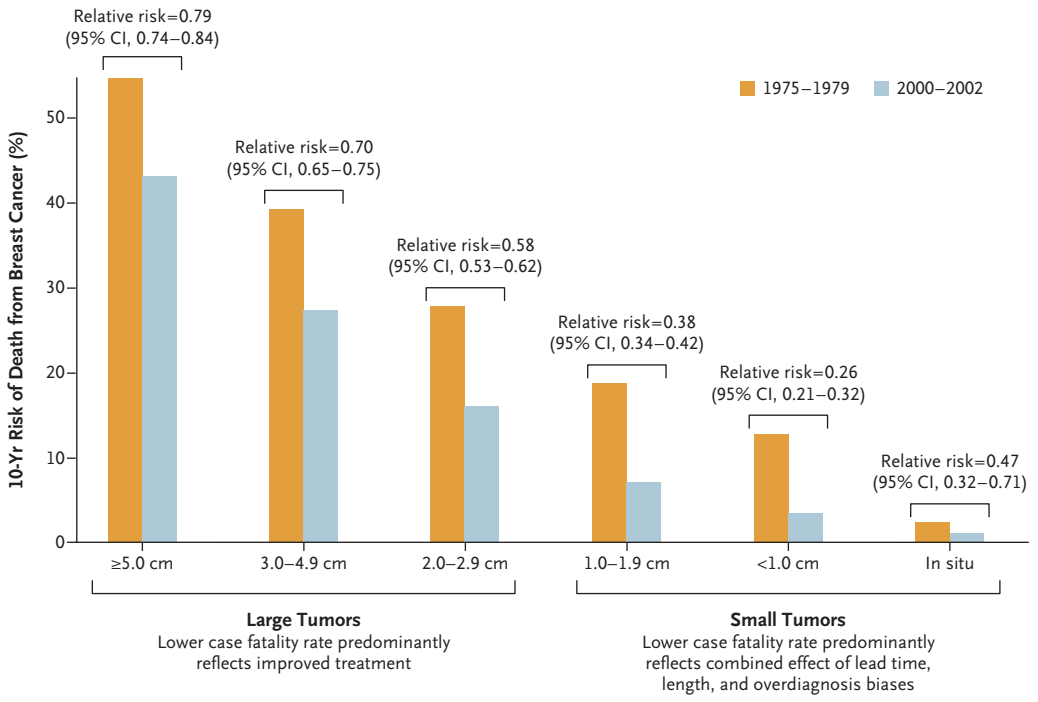

Figure 3. Change in Size-Specific Case Fatality Rate.

Data are shown for the size-specific 10-year risk of death from breast cancer (case fatality rate) among women 40 years of age or older with breast cancer that was treated before the introduction of screening mammography (diagnosis during the period from 1975 through 1979) and during the most recent years for which 10 years of follow-up were available (diagnosis during the period from 2000 through 2002). The relative risk is for the risk in 2000 through 2002 versus 1975 through 1979. In the period from 2000 through 2002, women with tumors less than 1 cm in size were more than 4 times as likely to die from causes other than breast cancer than from breast cancer. In that same period, women with in situ carcinoma were more than 10 times as likely to die from causes other than breast cancer than from breast cancer. Details of this analysis are provided in Section 3 in the Supplementary Appendix.

bution of screening versus improved treatment to declining breast-cancer mortality. Because the rate of deaths averted as a result of a declining size-specific case fatality rate was larger than the rate of those averted as a result of the reduced incidence of large tumors, we conclude, as others have,[15-19] that improved treatment must explain a majority of the reduction in mortality.

We should emphasize that our approximation of the contribution of screening mammography to the reduction in breast-cancer mortality is likely to be a highly favorable estimate. The screening-mediated reduction in the incidence of large tumors is unlikely to translate perfectly to a reduction in mortality. Screening can advance the time of diagnosis of a tumor, thereby detecting the tumor when it is still small, without changing its prognosis, a phenomenon termed "biologic predeterminism." In particular, screening is known to selectively identify tumors that have favorable molecular features.[20] Because tumors with favorable molecular features grow more slowly, they are disproportionately available to be detected by screening (so-called length-biased sampling). Thus, the expectation is that some tumors that are detected by screening when they are small would have favorable biologic characteristics and could have been treated equally effectively at clinical presentation.[8] For this subset of tumors, earlier detection at a smaller size would not translate into a mortality reduction.

Furthermore, there are other reasons to believe that we have overstated the effect of screening relative to improved breast-cancer treatment. We attributed the observed shift in tumor-size

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

Table 2. Approximations of the Effects of Improved Breast-Cancer Treatment and Screening Mammography on Breast-Cancer Mortality among Women 40 Years of Age or Older.*

Effect	Tumor Size			Total
	≥5.0 cm	3.0–4.9 cm	2.0–2.9 cm	
Approximate effect of improved treatment had screening not occurred				
Size-specific case fatality rate				
Baseline	55%	39%	28%	
Recent	43%	27%	16%	
Absolute reduction from baseline (percentage points)	12	12	12	
Baseline size-specific incidence of breast cancer per 100,000 women	29	56	60	
Mortality reduction per 100,000 women, calculated as absolute reduction from baseline × baseline size-specific incidence (95% CI)	3 (2–4)	7 (6–8)	7 (6–8)	17 (15–19)
Approximate effect of screening				
Size-specific incidence of breast cancer per 100,000 women				
Baseline	29	56	60	
Recent	25	38	52	
Absolute reduction from baseline	4	18	8	
Effect given previously available therapy				
Baseline case fatality rate	55%	39%	28%	
Mortality reduction per 100,000 women, calculated as absolute reduction from baseline × baseline case fatality rate (95% CI)	2 (2–3)	7 (7–8)	2 (2–3)	12 (11–13)†
Effect given more recent therapy				
Recent case fatality rate	43%	27%	16%	
Mortality reduction per 100,000 women, calculated as absolute reduction from baseline × recent case fatality rate (95% CI)	2 (1–2)	5 (5–6)	1 (1–1)	8 (7–9)

* This analysis was limited to large tumors (measuring ≥2 cm), where earlier detection could exert its beneficial effect. Incidence of breast cancer and mortality reduction are expressed as diagnoses of breast cancer and deaths from breast cancer, respectively, per 100,000 women in the population of women 40 years of age or older. The baseline time period refers to the period before the advent of widespread screening (1975 through 1979). The recent time period for this analysis refers to the period encompassing the most recent years for which 10 years of follow-up data were available (2000 through 2002).

† Values do not sum to 12 as a result of rounding. All calculations were made using full precision and were then rounded.

distribution solely to screening mammography and did not take into consideration the possibility that women may have sought care earlier in the course of their disease. Indeed, evidence suggests that increased breast-cancer awareness in the general population has led women to present earlier with clinically evident disease than they did in the past.[21] Finally, our analysis does not credit treatment for any improved prognosis in women with invasive tumors smaller than 2 cm. Although women with these small tumors have a generally favorable prognosis, 15 to 20% have node-positive disease, the category in which improved systemic therapy has exerted the greatest effect in declining breast-cancer mortality.[22]

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

Our findings are limited because a fundamental variable, the underlying true incidence of breast cancer (i.e., the true burden of disease), is unobservable. Observed breast-cancer incidence is influenced by observational intensity,[23] which represents the combined effect of the frequency of screening, the resolution of the screening examination, and the threshold that is used to label the examination as abnormal. We assumed that the underlying incidence of breast cancer was unchanged and that the observed increase reflected the increased observational intensity associated with screening (information regarding a sensitivity analysis with alternative assumptions is provided in Section 4 in the Supplementary Appendix). Proponents of screening mammography have contended that the increase instead reflects genuine disease and that overdiagnosis has been greatly exaggerated.[24] Those who postulate such substantial increases in underlying incidence, however, must explain why the increase coincides temporally with the introduction of screening, why the incidence of the most aggressive form of the disease — metastatic breast cancer — remains essentially unchanged,[25] and why overdiagnosis is also evident in analyses that are based on a single point in time.[26]

There is no perfectly precise method to assess the population effects of cancer screening. Screening mammography performed in an asymptomatic population that has an average risk of cancer can, at best, have only a small absolute effect on cancer-specific mortality because the vast majority of women are not destined to die from the target cancer. Because the mortality effect is necessarily delayed in time, the availability of improving cancer treatment over time further complicates the assessment of the contribution of screening. Inferences regarding overdiagnosis are equally imprecise since overdiagnosis cannot be measured directly. Studies in which minimal overdiagnosis is reported are typically conducted under the assumption that all abnormalities that are given the pathologic diagnosis of cancer will, in fact, progress.[27,28] Modeling all cancers as having lead time (i.e., as invariably progressing) ignores an important possibility, namely that some cancers are quiescent — or, in fact, regress.[29] To avoid these problems, a recent review concluded that high-quality ecologic and cohort studies represent the best designs for investigating overdiagnosis.[30]

We do not pretend to present a precise estimate of either the amount of overdiagnosis or the contribution of screening mammography to the reduction in breast-cancer mortality. The data regarding size-specific incidence, however, make clear that the magnitude of overdiagnosis is larger than is generally recognized. Furthermore, the data regarding size-specific case fatality rate clarify that decreasing breast-cancer mortality largely reflects improved cancer treatment.

The views expressed in this article are those of the authors and do not constitute official positions of the U.S. Government or the National Cancer Institute.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

REFERENCES

1. Kramer BS, Elmore JG. Projecting the benefits and harms of mammography using statistical models: proof or proofiness? J Natl Cancer Inst 2015;107(7).
2. 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.
3. Use of mammography among women 40 years of age and over, by selected characteristics: United States, selected years 1987–2008. Atlanta: Centers for Disease Control and Prevention (http://www.cdc.gov/nchs/data/hus/2010/086.pdf).
4. SEER*Stat Database: (1975-2012). Bethesda, MD: National Cancer Institute Surveillance Research Program, 2015 (http://www.seer.cancer.gov).
5. Merrill RM, Dearden KA. How representative are the Surveillance, Epidemiology, and End Results (SEER) program cancer data of the United States? Cancer Causes Control 2004;15:1027-34.
6. Zippin C, Lum D, Hankey BF. Completeness of hospital cancer case reporting from the SEER Program of the National Cancer Institute. Cancer 1995;76:2343-50.
7. Hudis CA. Biology before anatomy in early breast cancer — precisely the point. N Engl J Med 2015;373:2079-80.
8. Sparano JA, Gray RJ, Makower DF, et al. Prospective validation of a 21-gene expression assay in breast cancer. N Engl J Med 2015;373:2005-14.
9. Carlson RH. Lymph node status losing importance in breast cancer? Oncology Times 2011;33:31.
10. Rauscher GH, Murphy AM, Orsi JM, Dupuy DM, Grabler PM, Weldon CB. Beyond the mammography quality standards act: measuring the quality of breast cancer screening programs. AJR Am J Roentgenol 2014;202:145-51.
11. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. JAMA 2009;302:1685-92.
12. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 2012;367:1998-2005.
13. Lousdal ML, Kristiansen IS, Møller B, Støvring H. Effect of organised mammography screening on stage-specific incidence in Norway: population study. Br J Cancer 2016;114:590-6.
14. Lannin DR. Effect of screening mammography on breast cancer incidence. N Engl J Med 2013;368:678.
15. Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales,

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.

1990-8: comparison of observed with predicted mortality. BMJ 2000;321:665-9.
16. Paci E, Duffy SW, Giorgi D, et al. Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990-96. Br J Cancer 2002;87:65-9.
17. Kalager M, Zelen M, Langmark F, Adami H-O. Effect of screening mammography on breast-cancer mortality in Norway. N Engl J Med 2010;363:1203-10.
18. Park JH, Anderson WF, Gail MH. Improvements in US breast cancer survival and proportion explained by tumor size and estrogen-receptor status. J Clin Oncol 2015;33:2870-6.
19. Bleyer A, Baines C, Miller AB. Impact of screening mammography on breast cancer mortality. Int J Cancer 2016;138:2003-12.
20. Baré M, Torà N, Salas D, et al. Mammographic and clinical characteristics of different phenotypes of screen-detected and interval breast cancers in a nationwide screening program. Breast Cancer Res Treat 2015;154:403-15.
21. Rostgaard K, Vaeth M, Rootzén H, Lynge E. Why did the breast cancer lymph node status distribution improve in Denmark in the pre-mammography screening period of 1978-1994? Acta Oncol 2010;49:313-21.
22. Narod SA, Iqbal J, Miller AB. Why have breast cancer mortality rates declined? J Cancer Policy 2015;5:8-17.
23. Wennberg JE, Staiger DO, Sharp SM, et al. Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims. BMJ 2013;346:f549.
24. Kopans DB. Arguments against mammography screening continue to be based on faulty science. Oncologist 2014;19:107-12.
25. Welch HG, Gorski DH, Albertsen PC. Trends in metastatic breast and prostate cancer — lessons in cancer dynamics. N Engl J Med 2015;373:1685-7.
26. Harding C, Pompei F, Burmistrov D, Welch HG, Abebe R, Wilson R. Breast cancer screening, incidence, and mortality across US counties. JAMA Intern Med 2015;175:1483-9.
27. Baker SG, Prorok PC, Kramer BS. Lead time and overdiagnosis. J Natl Cancer Inst 2014;106(12).
28. Zahl P-H, Jørgensen KJ, Gøtzsche PC. Lead-time models should not be used to estimate overdiagnosis in cancer screening. J Gen Intern Med 2014;29:1283-6.
29. Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168:2311-6.
30. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. BMJ 2015;350:g7773.
Copyright © 2016 Massachusetts Medical Society.

The New England Journal of Medicine
Downloaded from nejm.org on June 19, 2021. For personal use only. No other uses without permission.
Copyright © 2016 Massachusetts Medical Society. All rights reserved.