Breast Cancer Screening, Incidence, and Mortality Across US Counties

Charles Harding, AB; Francesco Pompei, PhD; Dmitriy Burmistrov, PhD; H. Gilbert Welch, MD, MPH; Rediet Abebe, MAST; Richard Wilson, DPhil

IMPORTANCE Screening mammography rates vary considerably by location in the United States, providing a natural opportunity to investigate the associations of screening with breast cancer incidence and mortality, which are subjects of debate.

OBJECTIVE To examine the associations between rates of modern screening mammography and the incidence of breast cancer, mortality from breast cancer, and tumor size.

DESIGN, SETTING, AND PARTICIPANTS An ecological study of 16 million women 40 years or older who resided in 547 counties reporting to the Surveillance, Epidemiology, and End Results cancer registries during the year 2000. Of these women, 53,207 were diagnosed with breast cancer that year and followed up for the next 10 years. The study covered the period January 1, 2000, to December 31, 2010, and the analysis was performed between April 2013 and March 2015.

EXPOSURES Extent of screening in each county, assessed as the percentage of included women who received a screening mammogram in the prior 2 years.

MAIN OUTCOMES AND MEASURES Breast cancer incidence in 2000 and incidence-based breast cancer mortality during the 10-year follow-up. Incidence and mortality were calculated for each county and age adjusted to the US population.

RESULTS Across US counties, there was a positive correlation between the extent of screening and breast cancer incidence (weighted $r = 0.54; P < .001$) but not with breast cancer mortality (weighted $r = 0.00; P = .98$). An absolute increase of 10 percentage points in the extent of screening was accompanied by 16% more breast cancer diagnoses (relative rate [RR], 1.16; 95% CI, 1.13-1.19) but no significant change in breast cancer deaths (RR, 1.01; 95% CI, 0.96-1.06). In an analysis stratified by tumor size, we found that more screening was strongly associated with an increased incidence of small breast cancers ($\leq 2$ cm) but not with a decreased incidence of larger breast cancers ($>2$ cm). An increase of 10 percentage points in screening was associated with a 25% increase in the incidence of small breast cancers (RR, 1.25; 95% CI, 1.18-1.32) and a 7% increase in the incidence of larger breast cancers (RR, 1.07; 95% CI, 1.02-1.12).

CONCLUSIONS AND RELEVANCE When analyzed at the county level, the clearest result of mammography screening is the diagnosis of additional small cancers. Furthermore, there is no concomitant decline in the detection of larger cancers, which might explain the absence of any significant difference in the overall rate of death from the disease. Together, these findings suggest widespread overdiagnosis.
The goal of screening mammography is to reduce breast cancer mortality by detecting and treating cancer early in the course of the disease. If screening detects tumors early, the diagnosis of smaller and more treatable cancers should increase, while the diagnosis of larger and less treatable cancers should decrease. The associations between screening, incidence, mortality, and tumor size can be investigated at the population level by comparing areas of the United States that have different rates of screening.

One may expect that these associations are already understood. However, there are increasing concerns that screening unintentionally leads to overdiagnosis by identifying small, indolent, or regressive breast tumors that would not otherwise become clinically apparent. In addition, although screening mammography showed favorable efficacy in most randomized trials, these trials were conducted decades ago. There are concerns that the benefits and harms may have changed as treatments improved and screening was applied in general practice.

Local data on rates of mammography screening and breast cancer diagnosis are available for approximately one-fourth of the US population. We used these data to examine the associations between rates of modern screening mammography and the incidence of breast cancer, mortality from breast cancer, and tumor size.

Methods

Data Source
We analyzed all regions reporting to the Surveillance, Epidemiology, and End Results (SEER) cancer registries from January 1 to December 31, 2000, with the following exceptions: Louisiana, because Hurricane Katrina (2005) affected data quality; the Alaska Native Tumor Registry because it does not include counties; and Kalawao County, Hawaii, because of insufficient data (population of <200 individuals). The 547 remaining counties serve as the units of our analysis, which was performed between April 2013 and March 2015. The Dartmouth College Institutional Review Board has deemed studies using de-identified, publicly available data to be exempt from review.

Participants
Because screening mammography is not generally recommended for younger women, our study population was restricted to the 16 120 349 women who were 40 years or older in the SEER counties in 2000. Elderly women were not excluded; doing so would overlook any reductions in diagnoses at older ages owing to cancers that had been detected early by screenings at younger ages.

In the 547 counties, 55 809 women 40 years or older were diagnosed with breast cancer in 2000 (approximately one-fourth of all cases diagnosed in the United States10), of whom 53 207 (95.3%) had 10 years of follow-up and were included in the primary analysis (Figure 1). During follow-up, fewer deaths were attributed to breast cancer (7729 [42.4%]) than to other causes (10 511 [57.6%]).

Exposure
The study exposure was the percentage of women 40 years or older in each county who had a mammogram in the past 2 years, as of 2000 (median, 62.2%; range, 39.1%-77.8%). For all counties, estimates of screening mammography were published by the National Cancer Institute's Small Area Estimates (NCI-SAE) for Screening Behaviors program. The NCI-SAE combined self-reported mammogram histories from the National Health Interview Survey and Behavioral Risk Factors Surveil-
lance System survey, as supplemented by census data and synthesized using hierarchical modeling to adjust for survey nonresponse and noncoverage errors. To provide estimates for the year 2000, we took the mean of the NCI-SAE estimates for the periods 1997-1999 and 2000-2003. During this time, screening in counties increased by a median of 6 percentage points (interquartile range, 3-10 percentage points). Although NCI-SAE values potentially include some misclassified diagnostic mammograms, the practical consequences would be small, as Breast Cancer Surveillance Consortium data on women 40 years or older show that 91% of those with any breast imaging during 1998-2001 received routine screening.

Main Outcome Measures
The primary analysis evaluated the incidence of breast cancer, including ductal carcinoma in situ, among women 40 years or older in each county. For context, we also evaluated 10-year incidence-based mortality, which is defined as the proportion of women 40 years or older in a given county who were diagnosed with breast cancer in 2000 and died of the disease during the 10-year follow-up period. Forms of incidence-based mortality were first developed at the NCI in the early 1990s. Incidence-based mortality is common in studies of screening because, unlike the conventional form of mortality, it excludes cases diagnosed before the study period and, unlike 10-year survival, it is unaffected by overdiagnoses. Despite these differences, conventional and 10-year incidence-based mortality have similar magnitudes: in the 547 counties, the overall 10-year incidence-based mortality was 47.2 per 100,000 for cases diagnosed in 2000 and conventional mortality was 53.3 per 100,000 for deaths occurring in 2000-2010.

In stratified analyses, we investigated breast cancer incidence by tumor size (≤2 cm vs >2 cm), American Joint Commission on Cancer stage (0-II vs III-IV), and surgical treatments during first-course therapy (breast-conserving vs non-breast-conserving therapy). The American Joint Commission on Cancer stage was derived from the extent of disease data by SEER. In all analyses, incidence and incidence-based mortality in each county were age-adjusted to the year 2000 US population by the direct method.

Statistical Analysis
Spline methods were used to model smooth, curving associations between screening and cancer rates. We believed it would be inappropriate to assume that associations were linear, especially since nonlinear associations often arise in ecological data. In detail, univariate thin-plate regression splines (negative binomial model to accommodate overdispersion, log link, and person-years as offset) were specified in the framework of generalized additive models and fitted via restricted maximum likelihood, as implemented in the mgcv package in R. Quantile-quantile plots showed good agreement with some deviation in the tails, and residuals appeared patternless.

To summarize cross-sectional changes in incidence and mortality, we evaluated the mean rate differences and geometric mean relative rates (RRs) associated with a 10-percentage point increase in the extent of screening across the range of data (39%-78% screening). The 95% CIs were calculated by directly simulating from the posterior distribution of the model coefficients (50,000 replicates conditional on smoothing parameters).

Although the statistical analysis focused on nonparametric regression, we also calculated weighted Pearson correlation coefficients via population-weighted covariance but are popular because they are familiar to readers from all medical disciplines. To counterbalance, we include spline regression results with each correlation.

Results
Primary Analysis
Figure 2 shows a clear correlation between the extent of screening mammography and breast cancer incidence across US counties (weighted $r = 0.54$; $P < .001$). However, there is no evident correlation between the extent of screening and 10-year breast cancer mortality (weighted $r = 0.00$; $P = .98$).

We used the regressions shown in Figure 2 to summarize the mean change in incidence or mortality that accompanies...
an absolute increase of 10 percentage points in screening at the county level (eg, from 60% to 70% or 65% to 75%). A 10–percentage point increase in screening is associated with a 16% mean increase in breast cancer incidence (RR, 1.16; 95% CI, 1.13-1.19, or 35–49 cases per 100 000 as an absolute difference [AD]). However, there is no commensurate change in 10-year breast cancer mortality (RR, 1.01; 95% CI, 0.96-1.06, or -2 to +3 deaths per 100 000 as an AD).

**Tumor Size**

By stratifying on tumor size, Figure 3 demonstrates that the association between the extent of screening mammography and breast cancer incidence is largely confined to small cancers (≤2 cm), with no reduction in women presenting with larger cancers (>2 cm). A 10–percentage point increase in screening is accompanied by a 25% mean increase in the incidence of small cancers (RR, 1.25; 95% CI, 1.18-1.32, and 26–40 cases per 100 000 as an AD) and a 7% increase in the incidence of larger cancers (RR, 1.07; 95% CI, 1.02-1.12, and 2-9 cases per 100 000 as an AD). Examining results for individual centimeters of tumor size (eg, ≤0.9, 1-1.9, 2-2.9 cm) shows that our conclusions are not sensitive to the dichotomization into tumors of 2 cm or less vs those more than 2 cm (eAppendix 1 in the Supplement).

At the county level, screening mammography is associated with an increased incidence of small cancers and no significant reduction is seen in the incidence of breast cancer at any tumor size.

**Disease Stage and Surgery**

Since even some small tumors are aggressive, we considered disease stage as a proxy for aggressiveness (Figure 4A). A 10–percentage point increase in screening is associated with increased incidence of early disease (stage 0–II: RR, 1.22; 95% CI, 1.17-1.28) and no change in the incidence of locally advanced and metastatic disease (stage III–IV: RR, 1.02; 95% CI, 0.97-1.07). Given that one of the rationales for screening is to spare women more aggressive surgical procedures, we also examined first-course surgical treatment (Figure 4B). While a 10–percentage point increase in screening is associated with more breast-conserving surgical procedures (RR, 1.24; 95% CI, 1.15-1.34), there is no concomitant reduction in non–breast-conserving surgical procedures (eg, total and radical mastectomies).

**Discussion**

For the individual, screening mammography should ideally detect harmful breast cancers early, without prompting overdiagnosis. Therefore, screening mammography ideally results in increased diagnosis of small cancers, decreased diagnosis of larger cancers (such that the overall risk of diagnosis is unchanged), and reduced mortality from breast cancer. Across US counties, the data show that the extent of screening mammography is indeed associated with an increased incidence of early disease.
small cancers but not with decreased incidence of larger cancers or significant differences in mortality. In addition, although it has been hoped that screening would allow breast-conserving surgical procedures to replace more extensive mastectomies, we saw no evidence supporting this change.

What explains the observed data? The simplest explanation is widespread overdiagnosis, which increases the incidence of small cancers without changing mortality, and therefore matches every feature of the observed data. Indeed, our cross-sectional findings are supported by the prior longitudinal analyses of Esserman et al. in which excess incidences of early-stage breast cancer were attributed to overdiagnosis. However, 4 alternatives are also logically possible: lead time, reverse causality, confounding, and ecological bias.

In the absence of overdiagnosis, periods of increasing screening could result in rising incidence as cancer diagnoses advance in time. For example, it is well known that first screenings are especially likely to catch many presymptomatic cancers, temporarily increasing incidence. However, once screening is in a steady state, these lead-time effects on incidence diminish and then disappear. To consider the contribution of lead time, we compared counties where screening had increased from 1997-1999 to 2000-2003 with counties where screening rates had been stable (eAppendix 2 in the Supplement) while controlling for the current extent of screening. Incidence was not elevated in counties where the extent of screening had recently increased, suggesting that lead-time effects do not explain our results.

Counties could also have very different incidence rates of true breast cancers, which might be associated with the extent of screening because high-risk regions are especially targeted for screening (reverse causation) or because risk factors for true breast cancer happen to be associated with screening (confounding)—for example, through income’s associations with both age at first birth and participation in screening. Either way, the existence of more true breast cancers would result in deaths—yet we observe no increase in breast cancer mortality at the county level. Even where there are 1.8 times as many cancers being diagnosed, mortality is the same (Figure 2, counties at right). To maintain that these additional diagnoses are nevertheless true breast cancers, we would have to suppose that counties with more screening have better outcomes and, oddly, that the counterbalance is so precise that no (r = 0.00; RR = 1.01) association between screening and mortality remains. We see no reason for such a balance except coincidence. Therefore, although the balance is not impossible, we consider it improbable. It seems addition-
ally improbable to see more than twice as many diagnoses of true breast cancers at earlier stages without concomitant changes to the incidence of locally advanced and metastatic disease (Figure 4A).

Finally, our conclusions are inherently vulnerable to ecological biases.\(^2,3^{23-36}\) However, when we evaluated the association between screening and incidence within individual areas, we found similar results in all 9 states and 2 metropolitan areas (eAppendix 3 in the Supplement). Furthermore, the association between screening and breast cancer does not vary substantially by county population (eAppendix 4 in the Supplement). In summary, although ecological bias cannot be excluded without individual data, neither of the analyses suggested threats to validity.

To investigate the limits of the 10-year follow-up, we also analyzed cumulative mortality rates. However, there was no evidence to suggest that results observed during the first 10 years after diagnosis would change substantially thereafter (eAppendix 5 and eAppendix 6 in the Supplement). Other study limitations include loss to follow-up; missing tumor sizes; the absence of analyses of adjuvant therapy, risk factors, ultrasound, and spatial autocorrelation; the use of NCI-SAE estimates; and the quality of cause of death adjudication in registries.

Clinicians are correct to be wary of ecological studies because of the ecological fallacy. Ultimately, however, decisions must be made based on the evidence that is available, not unachievable ideals. Ecological studies are especially suitable for investigating overdiagnosis because overdiagnosis is currently not observable in individuals, only in populations. Indeed, the recent methodologic review by Carter et al\(^2\) concluded that the best designs for investigating overdiagnosis are high-quality ecological and cohort studies with multiple settings. Here, we examined 547 counties with diverse screening rates. It appears that no previous study of breast cancer overdiagnosis compared 12 or more counties, nations, or other regions.\(^2,3^{24-36}\) Other researchers consider mathematical modeling and simulation studies more reliable,\(^9\) but of course clinicians are also right to be wary of the untested assumptions required to model the fundamental unknown—the natural history of screen-detected breast cancers.\(^2\)

**Conclusions**

Our analysis shows that, when directed toward the general US population, the most prominent effect of screening mammography is overdiagnosis. Nonetheless, we do not believe that the right rate of screening mammography is zero. As is the case with screening in general, the balance of benefits and harms is likely to be most favorable when screening is directed to those at high risk, provided neither too frequently nor too rarely, and sometimes followed by watchful waiting instead of immediate active treatment.\(^37\)

Beyond these conclusions, the county data show 2 other troubling features, which we have avoided discussing heretofore because they are more tentative. First, screening is not associated with reduced presentation of larger breast cancers. However, it is not clear whether this county-level result means that screening is failing to catch true breast cancers before they become large, or if reductions in the presentation of large true breast cancers are concealed by increased presentation of large overdiagnoses. Second, and perhaps relatedly, screening was not associated with reduced mortality from breast cancer during the 10-year follow-up. However, observed mortality from breast cancer may be too rare and too noisy to reliably detect the 20% reduction at 13 years of follow-up that was estimated in a comprehensive meta-analysis of screening mammography trials.\(^1\) In summary, both of these features are promising topics for future research. This is also the right time to begin investigating whether all women undergoing screening mammography have the same risk of overdiagnosis, or if overdiagnosis is especially likely in some groups.


