Screening Mammography:

Update and Review of Publications Since Our Report in the New England Journal of Medicine on the Magnitude of the Problem in the United States

Archie Bleyer, MD

Rationale and Objectives: After a half century of clinical trials, expansive observations, vigorous advocacy and debate, screening mammography could not be in a more controversial condition, especially the potential harm of overdiagnosis. Despite a simple rationale (catch the cancer early and either prevent death or at least decrease the amount of therapy needed for cure), the estimates to date of overdiagnosis rates are conflicting and the interpretations complex.

Materials and Methods: Since the author's 2012 publication in the New England Journal of Medicine (NEJM), the peer-reviewed publications on overdiagnosis caused by screening mammography are reviewed and the NEJM analyses updated with three additional calendar years of results.

Results: The recent peer-reviewed medical literature on screening mammography induced overdiagnosis of breast cancer has increased exponentially, nearly 10-fold in 10 years. The average estimate of overdiagnosis is about 30%, but the range extends from 0% to 70+%. An update of the NEJM report estimates that in the US, 78,000 women and 30%–31% of those diagnosed with breast cancer at the age of 40 years or older during 2011 were overdiagnosed.

Conclusions: Until we have better screening procedures that identify who really has cancer and needs to be treated, the risk of overdiagnosis relative to the benefit of screening merits more effective public and professional education. Radiologists, pathologists, and other professionals involved with screening mammography should recognize that the potential harm of overdiagnosis is downplayed or not discussed with the patient and family, despite agreement that the objective is informed choice.

Key Words: Screening mammography; overdiagnosis; United States; literature review.

The author of this report was privileged to study the magnitude of overdiagnosis of breast cancer in the United States (US) with H. Gilbert Welch that was published in the New England Journal of Medicine (NEJM) on Thanksgiving Day, 2012 (1) and that evoked yet more controversy on the topic. Since then and as of August 1, 2014, there have been, according to the U.S. National Library of Medicine (PubMed), 46 publications on overdiagnosis in breast cancer in peer-reviewed journals in English with abstracts. Since the first report that alluded to overdiagnosis in 1982, the number of publications in peer-reviewed journals has increased exponentially, from an average of 3.8 during the 5 years before 2004 to 47 in 2013 and 45 in 2014 (Fig 1).

Acad Radiol 2015; 22:949–960

From the Oregon Health and Science University, Portland, Oregon. Received August 15, 2014; accepted March 8, 2015. Grant funds or other sources of funding were not required for the conduct of this research. Disclosures: The author has no potential conflicts of interest of which he is aware with the health screening industry or other related private or government sources. He is a consultant for Sigma-Tau Pharmaceuticals on an entirely different topic (acute lymphoblastic leukemia). **Address correspondence to:** A.B. e-mail: ableyer@gmail.com

http://dx.doi.org/10.1016/j.acra.2015.03.003

PURPOSE

This article summarizes and updates our report and the subsequent medical literature, knowing that, such as our national political schism, opinions on the topic are so polarized that any summary will be met with substantial skepticism. The controversy may explain in part why, in general, overdiagnosis is given short shrift by the radiology community, and yet regardless of its magnitude should be communicated as a risk to each woman being screened and her family. Thus, the review begins with reports on personalized approaches and educational needs. Reports of high and low overdiagnosis rates follow, and the literature review is concluded with those reporting professional biases and vested interests. The problem of false-positive (biopsy negative) mammograms and their adverse effects on quality life and financial costs is not included in this review.

MATERIALS AND METHODS

Publications in PubMed (2) with abstracts since November 2012 identified with search word "breast cancer overdiagnosis" were reviewed as of August 12, 2014 and summarized. Surveillance, Epidemiology, and End Results (SEER) (3) and U.S. Census data (4) for 2009, 2010, and 2011 were added to the

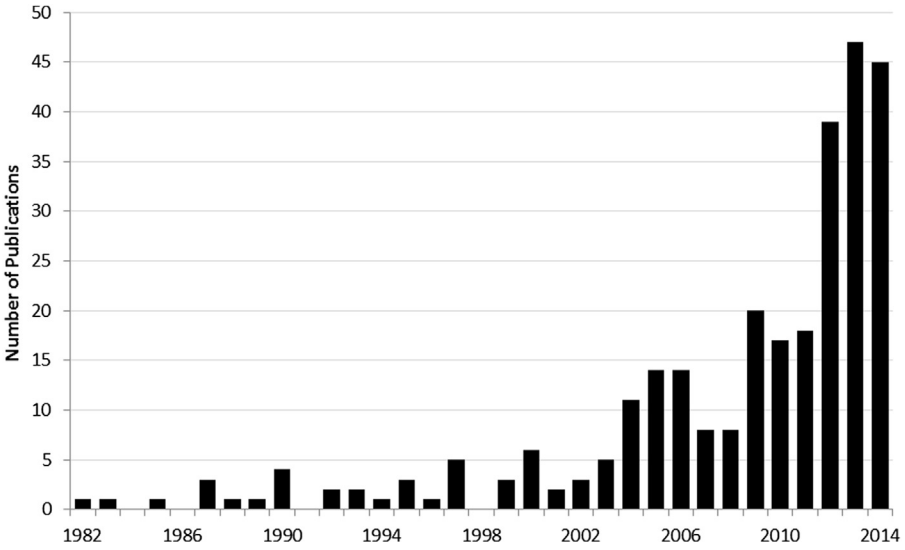

Figure 1. Annual number of peer-reviewed publications identified by PubMed with search terms "breast cancer overdiagnosis," 1982–2014.

original NEJM report since it had been 3 years when the analyses were performed for our original report, thereby updating the concluding year of our report from 2008 to 2011. In the context of breast cancer screening, "lead time" refers to the additional time an earlier diagnosis adds to the subsequent survival simply because the diagnosis is made earlier. "Lead time bias" occurs when screening appears to prolong survival, when in fact it only resulted in an earlier diagnosis in comparison to women who are not screened and did not prolong their survival. "Length bias" occurs when mammography detects a cancer with longer preclinical durations that are, by definition, present during more opportunities for discovery and therefore are more likely to be detected by screening; these cancers tend to be slow growing, have better prognoses, and are more likely identified by screening than clinically without screening. Overdiagnosis is an extreme form of length bias, in that screening finds cancers that are so slow growing or can regress spontaneously and never become manifest clinically in the woman's lifetime. "Aggregate incidence" is the incidence of a population whether individual subjects were screened, which if used for breast cancer screening may lead to overestimation of overdiagnosis rates (5).

RESULTS

Recent Peer-Reviewed Original Publications with Abstracts

Fifteen of the 46 reports are reviews or personal opinions and of the 30 original articles, four are primarily on biopathology, three pertain to women's perspectives, and eight describe country-specific programs (Denmark, Norway, United Kingdom, Portugal), one of which describes an ongoing national study without endpoint results. Most of the rest of the original articles focused on quantifying benefits and harms of screening mammography.

Personalized Approach and Educational Needs. Investigators at the Dartmouth Institute for Health Policy and Clinical Practice emphasize that breast cancer screening should be personalized, with the women's risks and preferences incorporated and a conceptual model to function at the level of the patient, provider, facility, health care system, and population/policy arena. Working with the National Cancer institute's initiative entitled Population-Based Research Optimizing Screening Through Personalized Regimens, their model builds on prior breast cancer screening models and may serve to identify new measures to optimize the benefits to harm ratio in population-based screening (6).

Radiologists in Spain express their concern about overdiagnosis and overtreatment by suggesting that patients be stratified according to their level of risk and that other radiologic procedures (tomosynthesis, ultrasonography, and magnetic resonance imaging) be considered in selected cases (7).

Faculty at the University of Porto reviewed 200 websites that provided information on breast cancer in Portuguese (8). They found that although 80% mentioned mammography as a breast cancer screening method, only 28% referred to it as the only recommended method and very few addressed the potential for overdiagnosis. In Australia, investigators at the School of Public Health, University of Sydney demonstrated how little women in Sydney suburbs knew about overdiagnosis associated with screening mammography (9). When in focus groups overdiagnosis was explained, women generally reacted with surprise, but most came to understand the issue. Responses to overdiagnosis and the different estimates of its magnitude were diverse. The highest estimate (50%) made some women perceive a need for more careful personal decision making about screening. In contrast, the lower and intermediate estimates (1%–10% and 30%) had limited impact on attitudes and intentions, with many women remaining committed to screening.

In community and University settings in London, epidemiologists at the University College London used a qualitative