

# UC Davis

## UC Davis Previously Published Works

### Title

Quantifying performance thresholds for recommending screening mammography: a revealed preference analysis of USPSTF guidelines.

### Permalink

<https://escholarship.org/uc/item/3m15x06v>

### Journal

Breast Cancer Research and Treatment, 172(2)

### Authors

Narayan, Anand  
Elkin, Elena  
Lehman, Constance  
et al.

### Publication Date

2018-11-01

### DOI

10.1007/s10549-018-4917-5

Peer reviewed



Published in final edited form as:

*Breast Cancer Res Treat.* 2018 November ; 172(2): 463–468. doi:10.1007/s10549-018-4917-5.

## Quantifying performance thresholds for recommending screening mammography: a revealed preference analysis of USPSTF guidelines

Anand K. Narayan<sup>1</sup>, Elena B. Elkin<sup>2</sup>, Constance D. Lehman<sup>3</sup>, Elizabeth A. Morris<sup>2</sup>

<sup>1</sup>Department of Radiology, Massachusetts General Hospital, 55 Fruit St. Wang 219H, Boston, MA 02114, USA

<sup>2</sup>Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA

<sup>3</sup>Massachusetts General Hospital, 55 Fruit St., Boston, MA 02114, USA

### Abstract

**Purpose**—During ongoing controversies about mammography screening, many investigators have stated that performance improvements in screening mammography may mitigate concerns about harms. However, there have been few attempts to quantify performance improvements required to recommend mammography screening. Based on USPSTF benchmarks, we utilized revealed preference methods to ascertain quantitative thresholds at which screening mammography would be recommended beyond biennial screening in women 50 and older.

**Methods**—Benefits of routine screening mammography (breast cancer deaths averted) were from published USPSTF meta-analyses. Potential harms (10-year cumulative probability of at least one false-positive) were from published Breast Cancer Surveillance Consortium estimates. We identified the implicit threshold (benefit/harm ratio) to recommend biennial screening starting at age 50. Using this threshold, we ascertained reductions of false-positives required to recommend more frequent screening and screening initiation under age 50 using revealed preference analyses.

**Results**—Using USPSTF implied benefit/harm ratio, routine biennial screening would be recommended starting at 40 if false-positives declined by at least 62%. Reductions of false-positive proportions of 74% would be required to recommend annual screening starting at 40 and reductions of false-positive proportions of 31% would be required to support annual screening starting at 50.

**Conclusions**—Using USPSTF revealed preferences, 31–74% reductions in false-positives would be required to recommend mammography screening beyond biennial screening starting at age 50. Widespread implementation of tomosynthesis and reducing recall rates to the lower end of recommended recall rates (5–12%) would provide support for expanding screening beyond biennial screening in women age 50.

---

Anand K. Narayan aknarayan@mgh.harvard.edu.

Compliance with ethical standards

**Conflict of interest** The other authors declare no conflicts of interest.

## Keywords

Revealed preferences; Economics; Breast cancer; Screening; Guidelines

---

## Purpose

Controversies about the benefits and harms of breast cancer screening in women at average risk have led to differing recommendations about when to begin and how often to have routine screening mammograms. For other screening tests, policy makers can increase thresholds for defining test results as positive thereby reducing false positives [1]. For example, in lung cancer screening, radiologists can change the threshold for defining a lung cancer screening CT as positive from 4 to 8 mm, thereby reducing false positives. Interpretation of screening mammograms, however, involves subjective analysis of breast parenchymal patterns without discrete quantitative thresholds that can be adjusted to decrease false positives. Although many investigators have stated that improvements in screening mammography performance may mitigate concerns about the harms of mammography screening [2, 3], there have been few attempts to quantify performance thresholds at which mammography screening would be recommended.

Revealed preference methods have been used in economics to quantify consumer preferences by analyzing actual choices [4]. Implicit behind breast cancer screening guidelines are the preferences of guideline makers for acceptable tradeoffs between the benefits and harms of screening mammography. The US Preventive Services Task Force guideline makers have identified two principle harms of screening mammography—overdiagnosis/treatment and false positives. While the task force identifies overdiagnosis/treatment as the most important potential harm, false positives are the primary harm quantified by the Task Force in determining the benefits/harms of various screening strategies. The Task Force used studies from the Breast Cancer Surveillance Consortium estimating that approximately 61% of women undergoing 2D mammography screening every year will have at least one false-positive examination over a 10 year time period, compared with 42% of women undergoing 2D mammography screening every other year [5]. Although the Task Force uses these numbers in considering the benefits/harms of recommending routine screening mammography, they did not explicitly state their quantitative preferences for what would be an appropriate benefit/harm ratio for recommending screening mammography. Our objective was to ascertain the implicit quantitative thresholds at which routine screening mammography would be recommended earlier than age 50 and more frequently than biennially.

## Materials and methods

We used revealed preference theory to quantitatively assess the preferences of guideline makers for recommending breast cancer screening. Revealed preference theory assumes that the preferences of consumers can be revealed by their actual purchasing habits. By analyzing the choices actually made by consumers, this analysis can be used to compare the influence of policies on consumer behavior.

Revealed preference theory assumes a given budgetary constraint. For example, if a consumer purchases item A over item B when item A and item B are both affordable, it is revealed that the consumer directly prefers item A over item B. Additionally, the theory assumes that if item A is preferred over item B but item B is preferred over item C, item A will be preferred over item C (assuming that items A, B and C are comparably affordable). Finally, the theory assumes that preferences are stable over the observed time period (A will always be preferred over B when both are affordable).

Just as consumers have preferences for goods that can be revealed by analyzing actual purchasing decisions, guideline developers have preferences about different types of health programs and policies that can be revealed by analyzing their actual recommendations. In 2009, and again in 2016, the U.S. Preventive Services Task Force (USPSTF) recommended routine screening mammography every 2 years starting at age 50 in women at average risk of developing breast cancer [6]. Revealed preference theory would suggest that the USPSTF found the tradeoff s associated with this screening strategy acceptable, and that given a similar set of constraints and ratio of harm to benefit, they would always recommend screening over not screening.

In the revealed preference framework, we can use the USPSTF's actual recommendation as an indication of its implicit threshold for the acceptable trade-off between the benefits and harms of routine screening mammography, and apply that to other age groups and screening intervals. For example, the USPSTF did not recommend routine biennial screening starting at 40; instead it recommended individualized decision making, due to the smaller expected benefit and higher risk of false-positive findings and unnecessary diagnostic workups among women in their 40's [7]. The USPSTF recommended against annual screening starting at age 50 for similar reasons. However, using the revealed preference framework, we can identify scenarios under which the USPSTF would recommend starting routine screening at age 40 or screening annually, based on the threshold ratio of harm to benefit implied by the actual recommendation to screening biennially starting at age 50.

## Variables and data sources

In their review of evidence, the USPSTF characterized the harms of routine screening mammography as the number of false-positive exams and their associated diagnostic workups, as well as the possibility of being diagnosed with or receiving treatment for an indolent or non-lethal breast cancer, a phenomenon often referred to as overdiagnosis or overtreatment [8]. Although overdiagnosis is a well accepted consequence of population-based cancer screening, it is difficult to estimate, and estimates of the rate of overdiagnosis associated with breast cancer screening vary widely depending on data source, assumptions, and methods [9]. However, estimates of overdiagnosis based on similar analytic methods tend to find similar rates when comparing screening strategies starting at age 40 versus age 50 [10]. Thus, for this analysis, we considered false positives to be the principal quantifiable harm of routine screening mammography, defined as the cumulative probability of a false-positive screening exam over 10 years. The principal benefit of screening mammography cited by the USPSTF is the reduction in deaths due to breast cancer, defined as breast cancer

deaths averted per 10,000 women screened by repeat screening mammography over 10 years [6].

We used available data from previously published studies commissioned and cited by the USPSTF. Specifically, estimates of the harms of routine screening mammography were from the Breast Cancer Surveillance Consortium (BCSC) [5]. Estimates of the expected benefit of routine screening mammography were from previously published USPSTF meta-analyses of randomized trials [6].

## Statistical methods

We calculated the reductions in 10-year cumulative false-positive exams per breast cancer death averted in each age group (40–49, 50–59) and for each screening interval (biennial, annual) required to achieve the threshold associated with routine biennial screening in women 50–59.

Changes in recall rates and false-positive probabilities could theoretically influence cancer detection rates and breast cancer-specific mortality. Therefore, we varied breast cancer mortality rates by increasing and decreasing the number of breast cancer deaths averted ( $\pm 1$  death averted per 10,000 women) in order to evaluate the sensitivity of our conclusions to changes in these assumptions.

## Results

The USPSTF meta-analysis found that routine screening mammography in women 40–49 averts 3 breast cancer deaths per 10,000, while breast cancer screening in women 50–59 averts 8 breast cancer deaths per 10,000. The BCSC estimated a 10-year cumulative false-positive proportion of 42.0% in women 50–59 (4200 false positives per 10,000 women) with biennial screening. The proportion was similar in women 40–49 (4160 false positives per 10,000 women). The 10-year cumulative false-positive proportion was 61.3% (6130 false positives per 10,000 women) for annual screening in both age groups.

Based on these inputs, biennial screening prevents 1 breast cancer death for every 525 false positives in women 50–59 (4200 false positives per 8 breast cancer deaths per 10,000 women). Biennial screening prevents 1 breast cancer death for every 1387 false positives in women 40–49 (4160 false positives per 3 cancer deaths per 10,000 women). Annual screening prevents 1 breast cancer death for every 2043 false positives (6130 false positives per 3 per 10,000) in women 40–49 and prevents 1 breast cancer death for every 766 false positives (6130 false positives per 8 per 10,000) in women 50–59.

The USPSTF's implicit threshold for recommending routine screening mammography is 1 breast cancer death prevented for every 525 false positives—the ratio of harm to benefit associated with biennial screening in women 50–59. Achieving this threshold for biennial screening in women starting at age 40 would require a decrease in the cumulative false-positive proportion of 62.1% [(1387 false positives—525 false positives)/1387 false positives] (Fig. 1). To achieve the implicit threshold for annual screening in women starting at age 40, we would require a decrease in the cumulative false-positive proportion of 74.3%

[(2043–525)/2043]. To achieve the threshold for annual screening in women starting at age 50, we would require a decrease in the cumulative false-positive proportions of 31.5% [(766–525)/766].

In order to assess the sensitivity of our findings to assumptions about the impact of false positives on cancer detection rates, we both increased and decreased the number of breast cancer deaths prevented by breast cancer screening. To assess the possibility of reduced benefit for women 40–49, we decreased the number of breast cancer deaths averted by screening from 3 per 10,000 to 2 per 10,000 in women 40–49, and for women 50–59 we decreased the number of breast cancer deaths prevented from 8 per 10,000 to 7 per 10,000 in women 50–59. With these assumptions of reduced benefit, in order to achieve the implied threshold for recommending routine screening mammography, we would need to decrease the false-positive proportion by 40.1% for annual screening starting at age 50, 74.8% for biennial screening starting at age 40, and 82.9% for annual screening starting at age 40.

Alternatively, improvements in technology or radiologist performance that reduce false positives could be associated with decreased breast cancer deaths. To assess the impact of this assumption, we increased the number of breast cancer deaths averted from 3 per 10,000 to 4 per 10,000 in women 40–49, and from 8 per 10,000 to 9 per 10,000 in women 50–59. With these assumptions, we would achieve the implicit threshold for recommending routine screening mammography if the false-positive proportion declined by 22.9% with annual screening starting at age 50, 49.5% with biennial screening starting at age 40, and 65.8% with annual screening starting at age 40.

## Conclusions

Revealed preference analyses of USPSTF guidelines suggest that 525 false-positive examinations for every 1 additional death prevented from breast cancer is an acceptable threshold for recommending routine screening mammography. In order to recommend starting routine screening at age 40, or screening annually instead of biennially, the false-positive proportions associated with those strategies would have to be 31–74% lower to meet the same threshold of harm to benefit. These types of reductions in false-positive proportions may be achievable with modern mammography technology and improved interpretive skills by specialized radiologists.

To maximize cancer detection rates while minimizing false positives, expert panels recommend abnormal interpretation (recall) rates ranging between 5 and 12% [10]. Recent national estimates from the BCSC suggest an overall abnormal interpretation rate (BI-RADS Category 0, 4, or 5) of 11.6% with 59.0% of radiologists achieving recommended abnormal interpretation rates ranging between 5 and 12% [11]. 31–74% reductions in the mean abnormal interpretation rate of 11.6% would result in mean abnormal interpretation rates ranging from 3.0 to 8.0%. Assuming similar cancer detection rates, this range of abnormal interpretation rates overlaps with the lower range of acceptable abnormal interpretation rates recommended by the BCSC, suggesting that improvements in the performance of screening mammography can provide support for expanding mammography screening beyond biennial screening in women > 50. These reductions in false positives are most readily achievable for

recommending annual screening in women starting at age 50 (31% reduction in false positives, abnormal interpretation rate of 8.0%), followed by biennial screening in women starting at age 40 (62% reduction in false positives, abnormal interpretation rate of 4.4%), followed by annual screening starting at age 40 (74% reduction in false positives, approximate abnormal interpretation rate of 3.0%).

While some investigators have raised concerns that decreases in recalls would be associated with decreases in cancer detection, Smith-Bindman et al. compared outcomes from two large scale mammography registries based in the United States with one large scale registry in the United Kingdom [12]. Comparing outcomes among women receiving their first screening mammogram in each country, recall rates were twice as high in the US (12.5–14.4%) as the UK (7.6%) while cancer detection rates were similar in the US (5.8–5.9 per 1000 mammography screens) as the UK (6.3 per 1000 mammography screens). Targeted, practice-based interventions can reduce recall rates while maintaining cancer detection rates [13].

In addition to improvements in radiologist performance, reductions in recall rates may be possible with recent advances in technology. The recent development and expansion of tomosynthesis has reduced false-positive rates by 15–30% while increasing cancer detection rates by approximately 29% [14]. Although overdiagnosis is not well-quantified in existing guidelines, concerns about overdiagnosis, particularly of low grade in situ cancers, were important considerations in both the USPSTF and ACS recommendations. Although the impact of tomosynthesis on breast cancer mortality has not been studied, higher proportions of breast cancers detected with tomosynthesis are invasive cancers compared with in situ cancers, suggesting that tomosynthesis screening may reduce overdiagnosed/treated cancers [15]. Modern mammography technology that reduces false positives, increases cancer detection, and shifts cancers diagnosed from DCIS to invasive cancers may support more effective screening programs for a larger population of women. In our study, we found that similar reductions in false-positive proportions and small increases in cancer detection rates would be sufficient to achieve implicit USPSTF quantitative thresholds for recommending breast cancer screening beyond biennial screening in women greater than 50 years old. The development of newer technologies, including contrast-enhanced mammography [16] and machine learning algorithms [17], offers the possibility of reducing false-positive rates and increasing cancer detection rates, particularly of more biologically aggressive cancers.

Limitations of our study include inherent assumptions under revealed preference theory and assumptions about the estimates our analysis is based on. Revealed preference theory assumes constant preferences over time, an assumption which may not be valid, given changing opinions about breast cancer screening [18]. Nevertheless, our analysis attempts to provide a quantitative evaluation of current thinking regarding the benefits and harms of breast cancer screening.

Our study was also limited by inherent subjectivity underlying thresholds for recommending breast cancer screening. While the USPSTF emphasizes harms of breast cancer screening in recommending biennial breast cancer screening starting at age 50, the American Cancer Society notes that women are more likely to present with more aggressive breast cancers



before menopause and therefore recommends annual screening starting at age 45. Alternatively, the American College of Radiology and the Society of Breast Imaging recommend annual mammography screening starting at age 40, because this strategy is associated with the largest decreases in breast cancer mortality. Although our study provides quantitative estimates about the USPSTF's implicit threshold for recommending routine screening mammography, similar analyses can be performed for other guideline producing organizations and for other types of cancer screening. If we had conducted similar analyses using other societies' recommendations for screening mammography, we would expect even lower thresholds for recommending screening mammography as other guidelines recommend screening mammography at earlier ages and higher frequency screening intervals compared with the USPSTF.

Additionally, the USPSTF emphasizes false positives as one of the primary quantitative harms of screening mammography; however, there is limited evidence that false positives inflict long-term damage on patients. Randomly selected participants in the Digital Mammographic Imaging Screening Trial (DMIST) surveyed after screening mammography were found to have increases in short-term anxiety after false positives but no long-term changes in anxiety or declines in health status [19]. Additionally, as the vast majority of false-positive recalls from screening mammography do not result in biopsies, some authors have suggested that benign biopsy rates would be a much more accurate measure of the harms of screening mammography [20]. Using benign breast biopsy rates as an alternative benchmark to false positives would suggest a much lower frequency of harms from screening mammography.

Finally, thresholds for evaluating the benefits and risks of screening mammography were derived mostly from studies that included negligible proportions of racial/ethnic minorities [6]. Prior studies have found that black women develop breast cancer at an earlier age and are much more likely to die from breast cancer [21]. Subgroup analyses evaluating benefit/harm ratios of breast cancer screening should consider the possibility that black women may benefit from starting breast cancer screening at an earlier age and more frequent screening interval, as they are much more likely to die from breast cancer at an earlier age.

In summary, revealed preference analysis suggests that reductions of false-positive proportions between 31 and 74% would be required for USPSTF guideline makers to recommend routine screening mammography annually or starting at age 40. These reductions are consistent with abnormal interpretation (recall) rates at the lower end of recommended abnormal interpretation rates (3.0–8.0%) and are consistent with wider adoption of tomosynthesis. Clinicians and policy makers should continue to develop and evaluate new technologies for breast cancer screening that will increase the detection of early stage breast cancers and reduce false-positive results.

## Acknowledgments

**Funding** This study did not receive any funding.

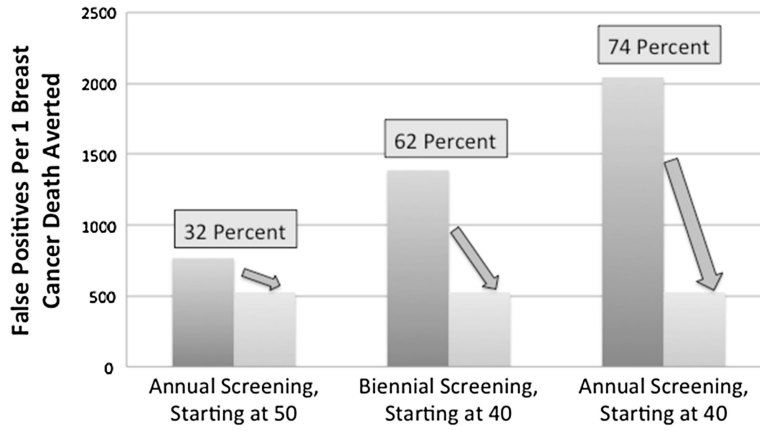
CDL reports research support from General Electric and is on a health care advisory board for General Electric.



## References

1. Fintelmann FJ, Bernheim A, Digumarthy SR, Lennes IT, Kalra MK, Gilman MD, Sharma A, Flores EJ, Muse VV, Shepard JA (2015) The 10 pillars of lung cancer screening: rationale and logistics of a lung cancer screening program. *Radiographics* 35(7):1893–1908 [PubMed: 26495797]
2. Elmore JG, Barton MB, Mocerri VM, Polk S, Arena PJ, Fletcher SW (1998) Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 338(16):1089–1096 [PubMed: 9545356]
3. Welch HG, Black WC (2010) Overdiagnosis in cancer. *J Natl Cancer Inst* 102(9):605–613 [PubMed: 20413742]
4. Samuelson P (1937) A note on measurement of utility. *Rev Econ Stud* 4(2):155–161
5. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL (2011) Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med* 155(8):481–492 [PubMed: 22007042]
6. Siu AL, U.S. Preventive Services Task Force (2016) Screening for breast cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* 164(4):279–296 [PubMed: 26757170]
7. Nelson HD, Cantor A, Humphrey L, Fu R, Pappas M, Daeges M, Griffin J (2016) Screening for breast cancer: a systematic review to update the 2009 U.S. Preventive Services Task Force Recommendation [Internet]. Agency for Healthcare Research and Quality (US), Rockville
8. Carter SM, Barratt A (2017) What is overdiagnosis and why should we take it seriously in cancer screening? *Public Health Res Pract* 27(3):2731722 [PubMed: 28765855]
9. de Gelder R, Heijnsdijk EA, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ (2011) Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 33:111–121 [PubMed: 21709144]
10. Rosenberg RD, Yankaskas BC, Abraham LA, Sickles EA, Lehman CD, Geller BM, Carney PA, Kerlikowske K, Buist DS, Weaver DL, Barlow WE, Ballard-Barbash R (2006) Performance benchmarks for screening mammography. *Radiology* 241(1):55–66 [PubMed: 16990671]
11. Lehman CD, Arao RF, Sprague BL, Lee JM, Buist DS, Kerlikowske K, Henderson LM, Onega T, Tosteson AN, Rauscher GH, Miglioretti DL (2017) National performance benchmarks for modern screening digital mammography: update from the breast cancer surveillance consortium. *Radiology* 283(1):49–58 [PubMed: 27918707]
12. Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K (2003) Comparison of screening mammography in the United States and the United Kingdom. *JAMA* 290(16):2129–2137 [PubMed: 14570948]
13. Mullen LA, Panigrahi B, Hollada J, Panigrahi B, Falomo ET, Harvey SC (2017) Strategies for decreasing screening mammography recall rates while maintaining performance metrics. *Acad Radiol* 24(12):1556–1560 [PubMed: 28760363]
14. Morris E, Feig SA, Drexler M, Lehman C (2015) Implications of overdiagnosis: impact on screening mammography practices. *Popul Health Manag* 18(Suppl 1):S3–S11 [PubMed: 26414384]
15. Bahl M, Gaffney S, McCarthy AM, Lowry KP, Dang PA, Lehman CD (2017) Breast cancer characteristics associated with 2D digital mammography versus digital breast tomosynthesis for screening-detected and interval cancers. *Radiology* 22:171148
16. Jochelson M (2014) Contrast-enhanced digital mammography. *Radiol Clin North Am* 52(3):609–616 [PubMed: 24792660]
17. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD (2017) High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. *Radiology* 17:170549
18. Narayan A, Fischer A, Zhang Z, Woods R, Morris E, Harvey S (2017) Nationwide cross-sectional adherence to mammography screening guidelines: national behavioral risk factor surveillance system survey results. *Breast Cancer Res Treat* 164(3):719–725 [PubMed: 28508184]

19. Tosteson AN, Fryback DG, Hammond CS, Hanna LG, Grove MR, Brown M, Wang Q, Lindfors K, Pisano ED (2014) Consequences of false-positive screening mammograms. *JAMA Intern Med* 174(6):954–961 [PubMed: 24756610]
20. Arleo EK, Hendrick RE, Helvie MA, Sickles EA (2017) Comparison of recommendations for screening mammography using CISNET models. *Cancer* 123(19):3673–3680 [PubMed: 28832983]
21. Stapleton SM, Oseni TO, Bababekov YJ, Hung YC, Chang DC (2018) Race/ethnicity and age distribution of breast cancer diagnosis in the United States. *JAMA Surg* 153(6):594–595 [PubMed: 29516087]



**Fig. 1.** Percentage reductions in false positives required to achieve USPSTF revealed preference threshold for recommending different mammography screening schedules. This figure characterizes the percentage reductions in false positives required in order to achieve the USPSTF revealed preference threshold for recommending routine screening mammography by age group and screening interval. For each screening strategy, the dark grey bar (left) represents the ratio of false positives for each breast cancer death averted and the light grey bar (right) represents the USPSTF revealed preference threshold for recommending routine screening (525 false positives for every 1 breast cancer death averted). Above the two bars is the percentage reduction in false positives required to achieve the USPSTF threshold. For example, the USPSTF’s implicit threshold (revealed preference) for recommending routine screening mammography is 1 breast cancer death prevented for every 525 false positives—the ratio of harm to benefit associated with biennial screening in women 50–59 (recommended screening schedule by the USPSTF). Achieving this threshold for biennial screening in women starting at age 40 would require a decrease in the cumulative false-positive proportion of 62.1% [(1387 false positives—525 false positives)/1387 false positives]