Submitted Electronically

United State Preventive Services Taskforce
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

Re: Draft USPSTF Recommendations on Breast Cancer Screening

Dear Chairman Siu and Task Force Members,

The American College of Radiology (ACR)[1] and the Society of Breast Imaging (SBI)[2] have a long history of advocating for quality in mammographic screening, and of encouraging women and their health care providers to utilize proven screening methods to save lives. Therefore, we are gravely concerned by the United States Preventive Services Task Force (USPSTF) draft recommendations on breast cancer screening, which we believe greatly overstate the potential harms of breast cancer screening and greatly underestimate the benefit of mammography in reducing morbidity and mortality from this disease. Notwithstanding what we consider to be an imbalanced view of the available evidence, the Task Force still recognized that annual mammography screening for women forty and older saves the most lives and results in the greatest number of life-years gained, based on the RCTs, observational trials and CISNET models. Accordingly, we respectfully urge you to reconsider the evidence as detailed below and adopt final recommendations ascribing a B grade for annual mammographic screening of women age forty and older.

It is important to note that the C rating in the draft recommendations for women age 40-49 reflects a value judgment on the part of the Task Force that is contrary to survey-supported evidence on women's attitudes toward "false positive" mammograms. Moreover, in making this value judgment and assigning a "C" rating for this population of women, the Task Force could ultimately impair women's access to insurance coverage under the Affordable Care Act, thus severely limiting a woman's right to choose the screening schedule that best reflects her personal choice and values about screening.

The ACR and SBI are extremely disappointed at the methodology used by the USPSTF to generate this draft. Many of the types of significant errors made by the Task Force were foreseen and outlined by the Institute of Medicine (IOM) when they formulated their document "Clinical Practice Guidelines We Can Trust" (1). Similarly, the Task Force report does not reflect the transparency and accountability protections afforded by the Administrative Procedures Act and the Federal Advisory Committee Act.

The IOM recommends that trustworthy guideline development should include a knowledgeable *multidisciplinary panel of experts* and representatives from *key affected groups*. In fact, the IOM

[1] ACR is a professional organization representing more than 35,000 radiologists, radiation oncologists, interventional radiologists, nuclear medicine physicians, and medical physicists

[2] SBI's core mission is to save lives through early detection, quality education, and trusted information provided to patients, physicians and organizations worldwide

report states that, "…there is broad international consensus that GDGs should be multidisciplinary, with representation from all key stakeholders (ACCF and AHA, 2008; AGREE, 2003; NICE, 2009; SIGN, 2008)" (1). The USPSTF has failed on this account. The USPSTF panel did not include a single expert in breast cancer diagnosis or care. This is unreasonable for a guideline with such important implications, and could have been easily achieved if proper planning had prevailed. Failure to include knowledgeable experts hampered the ability of the USPSTF to understand and review the evidence. The IOM report suggests that such guideline development cannot assess the evidence in the same way that a multidisciplinary group can. Nuances are missed and data is sometimes misunderstood. This has resulted in critical omissions and errors in the draft report.

The USPSTF has failed to be transparent in its evaluation of the evidence and in the processes used to make the draft recommendations. We agree with the IOM that a transparent guideline should give users confidence that guidelines are based on best available evidence, largely free from bias, clear about the purpose of recommendations to individual patients, and therefore trustworthy (1). The USPSTF draft guidelines lack a clear indication of the panel's clinical expertise, their biases, and where consensus was apparent or lacking in their deliberations.

Further, the external review performed under the Task Force's directive was inappropriately opaque. Were the comments of the external reviewers taken into account by the panel? If so, what comments were incorporated and which were discounted and why? There should be a systematic process for responding and noting ***all*** external review comments and this must be transparent. A table should have been published concurrent with this draft showing each commentary from every reviewer, explaining how the guideline was or was not modified accordingly and describing the rationale for these actions (1). The public availability of this information is important for the transparency needed to establish trust. In viewing the draft document presented, it is clear that either there were no expert reviewers who understand breast cancer screening or that there were experts but these external reviewers were ignored.

These draft recommendations have utilized tremendous Federal resources yet the product is not trustworthy because of the lack of stakeholders and experts, the dismissal of expert external review, the overall lack of transparency of its panel and processes, and the numerous critical factual errors and omissions in the draft and supporting AHRQ documents. We urge the USPSTF to review the comments herein and seriously reconsider its draft recommendations to better reflect the current and complete evidence available for breast cancer screening.

Both the SBI and the ACR independently visited with the Task Force leadership before the development of the guidelines and reminded them that we would be looking for such crucial input by breast cancer experts, adherence to the IOM methodology and transparency. We advised the Task Force leaders that the literature on the subject of breast cancer screening is enormous and that there are many published studies that passed peer review but are not scientifically valid. We emphasized that expert participation was absolutely critical in the Task Force review of breast cancer screening because otherwise a review of the literature by individuals naïve to the subject might result in an inappropriate assessment of the benefits and harms of screening.

A detailed review of evidence and analysis of the Task Force is provided below.

Evidence of Mortality Reduction from Screening Mammography

The USPSTF relies almost entirely on randomized trial data to assess the mortality reduction from screening mammography. Given the existence of randomized trials, albeit using obsolete mammography technology in an era when good systemic breast cancer treatment also was unavailable, it is reasonable to cite the trials as demonstrating the *existence* of mortality reduction as a benefit of screening. However, there are numerous reasons why these trials (or any trials) underestimate the *magnitude* of mortality reduction, magnitude being of great consequence in assessing benefits versus harms. Most important among these are: [a] non-compliance in the study cohort (women counted in the study group who do not undergo screening dilute the observed benefit); and [b] contamination of the control group (women in the control group who undergo screening or diagnostic imaging for signs or symptoms) outside of the study actually do experience the mortality reduction of screening but are counted in the control group, thereby further diluting the observed benefit).

The magnitude of the effects of both [a] and [b] is readily demonstrated by cohort (incidence-based mortality) studies and case-control studies, within which the magnitude of mortality reduction observed for invitation to screening (invited versus not invited) is very substantially lower than that observed for exposure to screening (screened versus not screened). This is because the invitation to screening group differs from the exposure to screening group by the effects of [a] and [b]. This has been documented by EUROSCREEN systematic reviews of organized screening programs in Europe: for incidence-based mortality studies, invitation to screening yielded a 25% mortality reduction (RR 0.75, 95% CI 0.69-0.81), while exposure to screening yielded a 38% mortality reduction (RR 0.62, 95% CI 0.56-0.69); for case-control studies after adjustment for self-selection, invitation to screening yielded a 31% mortality reduction (RR 0.69, 95% CI 0.57-0.83), while exposure to screening yielded a 48% mortality reduction (RR 0.52, 95% CI 0.42-0.65) (2).

The USPSTF recommendation statement pays lip service to these data, acknowledging in a single sentence only the invitation to screening (invited versus not invited) data. In the United States, however, screening mammography is opportunistic and not centrally organized, so the effectiveness of screening is based on whether an individual woman actually attends screening. Non-compliance and contamination, meaningful in the context of invitation-to-screen trials, are meaningless in the context of opportunistic screening. Therefore, for the United States, mortality reduction is best measured, and screening guidelines best based, on exposure to screening (screened versus not screened). As stated above, the EUROSCREEN pooled data from 20 incidence-based studies showed a 38% mortality reduction for screened women compared to unscreened women and from 8 case-control studies, a 48% mortality reduction for screened compared to unscreened women after adjustment for self-selection. There also are robust incidence-based mortality-study data for exposure to screening (screened versus not screened) from within North America, based on service screening data from the organized screening programs throughout the Canadian provinces. These particularly relevant data, involving 20.2 million person-years and not even acknowledged by the USPSTF recommendation statement, indicate an overall 40% mortality reduction for exposure to screening (RR 0.60, 95% CI 0.52-0.67) (3), almost exactly mirroring comparable EUROSCREEN data.

The USPSTF states that its decision-model-based estimate of mortality reduction is higher than that from meta-analysis of the randomized trials, because at least in part, meta-analysis evaluated the impact of screening across a decade whereas the decision models evaluated screening across an entire lifespan. However, the decision models still apparently underestimated the magnitude of mortality reduction because there is convincing evidence from the trial with longest reported follow-up that less than half of the total mortality reduction (for the

7 years of screening in this trial) was observed after only 10 years of follow-up; continued substantial mortality reduction was observed all the way out to 29 years of follow-up (4). From these data it is reasonable to expect that substantially greater mortality reductions would have been reported for all the other trials that showed mortality reduction, given that none of these other trials reported follow-up beyond 18 years. Therefore, the USPSTF estimates of mortality reduction were limited by the years of follow-up reported by most trials.

In conclusion, the USPSTF has failed to consider the much larger magnitude of mortality reduction pertinent to the practice of screening mammography in the United States, an important omission because the magnitude of benefit is of great consequence in balancing benefit versus harms. The USPSTF should be telling the American woman the amount of mortality reduction she should expect if she chooses to be screened in 2015, not the mortality reduction she would have received had she been randomized to the invited group in a trial conducted in the 1970's and 1980's. This is best estimated from observational studies, which were largely ignored in this review.

The USPSTF could inform the American woman about the amount of mortality reduction she could expect from modern mammography by including in Table 1 the mortality reduction estimated by CISNET models. Below is Draft Table 1, which has been extended to include 2009 and 2015 CISNET results on lives saved by screening mammography.

Draft Table 1, USPSTF Guidelines: Modified to Compare RCT Results with CISNET 2009 and CISNET 2015 Results

	Ages 40–49 Years	**Ages 50–59 Years**	**Ages 60–69 Years**	**Ages 70–74 Years**
Breast cancer deaths avoided based on RCT meta-analysis	4 (95% CI: 0-9)	8 (95% CI: 2-17)	21 (95% CI: 11-32)	13 (95% CI: 0-32)
Breast cancer deaths avoided based on 2009 CISNET modeling, annual screening; mean (median)	13 (9) [range: 4-28]	29 (23) [range: 18-48]	43 (41) [range: 36-54]	15 (14) [range: 12-18]
Breast cancer deaths avoided based on 2009 CISNET modeling, biennial screening; mean (median)	10 (8) [range: 2-18]	22 (18) [range: 15-33]	33 (33) [range: 27-39]	15 (15) [range: 12-18]
Breast cancer deaths avoided based on 2015 CISNET modeling, annual screening; mean (median)	14 (12) [range: 9-18]			
Breast cancer deaths avoided based on 2015 CISNET modeling, biennial screening; mean (median)	11 (12) [range: 7-13]			

Note: 2015 CISNET results did not provide mortality reduction by 10-year age intervals, so 2015 CISNET data could be used only to estimate breast cancer deaths avoided for women ages 40-49 years. CISNET should be able to provide these data and verify included results.

CISNET estimates of lives saved are more relevant than RCT data not just because they are based on the performance of modern screening mammography in the United States, but also because they reflect benefit to women actually screened compared to unscreened women, while RCT data reflect benefit to women invited to screening compared to uninvited women. Since Draft Table 2 lists "harms" to women actually attending screening based on modern U.S. data, Table 1 should describe the benefits to women screened based on modern data, not simply women invited to screen in outdated RCTs. To do otherwise would be to make the same error the Task Force made in their 2009 recommendations of confusing number needed to invite (NNI) with number needed to screen (NNS) to prevent one breast cancer death.

Women Age 40-49: Magnitude of Mortality Reduction

The USPSTF has decided to ignore almost all observational data in determining the magnitude of mortality reduction in women age 40-49. The USPSTF recommendation statement makes no mention of, nor is it clear whether USPSTF members are even aware of, a key observational study performed in Sweden, a study that comes as close as possible to a randomized trial, a study limited to women ages 40-49 that involved far more person-years of exposure than all the randomized trials combined for women ages 40-49. Hellquist, et al, using (intention-to-screen) service screening data from the various Swedish counties that did and did not offer screening for women age 40-49, after adjustment for self-selection bias, reported a statistically significant 26% mortality reduction for women age 40-49 who live in screening counties compared to women who live in non-screening counties (RR 0.74, 95% CI 0.66-0.83).(5) This result occurred in a country where treatment guidelines are uniform and closely adhered to across all counties, where all women have equal access to breast cancer treatment, so that this mortality reduction was achieved *in addition to the benefits of modern therapeutic advances*. The 26% statistically significant mortality reduction demonstrated in this study with more than 7 million woman-years of observation is more than twice the 12% mortality reduction used by the USPSTF in its deliberations.

The pan-Canadian service screening study (3) screened women ages 40-49 having a 44% mortality reduction (RR 0.56, 95% CI 0.45-0.67), ages 50-59 having a 40% mortality reduction (RR 0.60, 95% CI 0.49-0.70), ages 60-69 a 42% mortality reduction (RR 0.58, 95% CI 0.50-0.67), and screened women 70-79 a 35% mortality reduction (RR 0.65, 95% CI 0.56-0.74). Each decade-specific mortality reduction was statistically significant. This study comes as close as possible to screening performed in the United States based on exposure to screening (screened versus not screened). Also note the small differences in mortality reduction for different age decades, justifying the conclusion that *modern mammography screening appears to be effective in reducing breast cancer deaths by approximately the same amount in women 40-79 regardless of age decade*. The substantially different age-specific mortality reductions observed in the randomized trials, and used by the USPSTF to develop age-specific screening recommendations, no longer apply to modern service screening. These two studies, the former (3) close in design to that of a randomized trial, the latter (5) producing age-specific data for screening within North America, demonstrate compelling evidence of the life-saving capabilities of modern screening mammography, including women ages 40-49.

In addition, the new CISNET models show a median 47/1000 LYG benefit for annually screened women age 40-49 years. This benefit is substantially higher than 2009 CISNET estimates which used outdated film-screen mammography. For decision making, women should be informed that the per-decade LYG benefit of annual screening ages 40-49 is the same as USPSTF-recommended per-decade benefit (49/1000) for biennial screening of women ages 50-

74. The 2015 CISNET models estimate that annual mammography in the 40-49 age decade provides a 58.5% improvement in LYG compared to biennial screening.

USPSTF requested CISNET to model starting ages of 40 and 45 (Tables 10a, b). Data in these tables, however, erroneously show 28 data points that are identical for LYG and QALY for starting ages 40 (10a) and 45 (10b). This is erroneous based upon the differences of mortality reduction in the same tables. We are very concerned that incorrect information was used in formulating the "C' level recommendation for women ages 40-49, given that this apparent error was not noted by a single member of the Task Force or CISNET.

Women Age 40-49: "C" Recommendation

The USPSTF explains its "C" recommendation for women ages 40-49 by indicating that women should individually decide whether they will undergo screening based on an informed *personal* decision of whether the benefits of screening exceed the harms. If the final recommendation retains the "C" rating, however, the recommendation will likely limit patient choice, not empower it. The Affordable Care Act requires private insurers to cover screening tests with a USPSTF grade of "B" or above at no cost to the patient. There is no such requirement for screening tests with a "C" grade. If the draft recommendations are adopted as final, 17 million women ages 40-49 could be forced to make a financial decision about breast cancer screening and many will not be able to benefit from the shared decision making process with their physicians, as recommended by the Task Force.(6) We strongly believe that the USPSTF's rating should not become a barrier to a woman's access to care or limit her informed choice about breast cancer screening.

To be clear, the "C" rating is not dictated by the evidence; it is a value judgment based on the Task Force's opinion of what constitutes benefits and harms of mammography and its subjective weighting of the net benefits of screening for this population. There is ample support, not only in the evidence contained in these comments but also in the draft recommendations, on which the Task Force can support a "B" rating in this population.

The USPSTF's decision not to recommend screening mammography to women ages 40-49, but rather to let the decision of when a woman begins screening be an informed *personal* decision of whether the benefits of screening exceed the harms, relies heavily on outdated RCT data on benefit and CISNET modeling results on harms. This advice to women from the USPSTF ignores the caveat stated in the Executive Summary of the 2015 CISNET Collaborative Modeling document: "… these analyses were designed to provide modeling data for use in public health decision making for populations of women; the results are not intended to guide individual screening decisions." Yet that is exactly what the USPSTF is asking women 40-49 to do. The irony of this inconsistency is not lost on the ACR and the SBI; this cavalier disregard of the context should not be allowed to impact women's lives.

USPSTF Recommended Screening Strategy versus Other Strategies

Based in large part on CISNET modeling of benefit and harms with various screening strategies, the USPSTF has maintained their 2009 recommendation for biennial screening mammography in women ages 50-74 (B50). This is in spite of the fact that median 2015 CISNET models show annual screening from ages 40-74 (A40) will save 46.5% more lives and 57% more life-years than B50 (Collaborative Modeling, Tables 7a and 7b). To inform women and providers, the USPSTF should provide a clearly labeled benefits table that compares CISNET mortality reduction, LYG and deaths averted for biennial screening ages 50-74 with annual screening of

ages 40-79. Failure to provide a comparison benefit analysis while providing harms data demonstrates bias and non-transparency. In addition, CISNET modeling of the UK Age Trial of annual screening in women 40-49 with 100% compliance and 13 years follow-up yielded a median mortality reduction of 28% (range 25% to 35%), in good agreement with the observed mortality reduction of 24% for screened versus unscreened women.

Although unstated by the USPSTF, their strategy is clearly to maximize lives saved per mammogram performed, not to save the most women's lives. While the USPSTF is careful to mask this decision in terms of benefits versus harms, it amounts to benefit (in terms of lives or life-years saved) versus cost, with number of mammograms performed as a surrogate for cost. In this vein, the CISNET modelers go so far as to explicitly list "number of mammograms" as a "harm" of screening (Collaborative Modeling, p. iv). This comes perilously close to resembling the "death panel" approach to health care access that critics of the Affordable Care Act (and general government involvement in health care administration) fear.

The USPSTF relied on an obsolete method for analyzing outcomes data for biennial (2-year) screening, incorrectly using a 1-year period instead of 2-year period for cancer ascertainment. Use of a 2-year ascertainment period, the current standard approach in the United States (7), will likely show substantially higher false-negative outcomes for 2-year versus 1-year screening; randomized trial data indicate that far more than half of interval cancers are identified in the second half of the 2-year period between screens (8).

There are two major misconceptions about the differences in outcomes between annual and biennial screening. [a] Annual screening will identify the same total number of cancers as biennial screening (untreated cancers do not completely regress spontaneously during the course of one year). However, what is important is that annual screening will identify approximately half of the cancers at smaller size, more likely node-negative, and more likely at earlier stage, all of which are surrogate markers for longer survival and mortality reduction. [b] Annual screening also will have more false positives. But it is overly simplistic to think that this is due to twice the number of screening exams – a mammographic finding that is false positive at annual screening will be false positive on only one examination (after which it will be acknowledged as benign); the same mammographic finding will also be false positive at biennial screening (again on only one examination – very few false-positive findings completely regress during the course of one year). The real reason for increased false positives at annual screening is that it is more difficult to detect small, early stage cancer than it is to detect larger, more advanced cancer. Hence, the threshold for identifying potential malignancy at annual screening is different, requiring abnormal assessment (recall and occasionally biopsy) of more findings in order to identify smaller, more subtle cancers.

Women Age 75 Years and Older: "I" Recommendation

Far too little data from observational studies are presented in the USPSTF recommendation statement on screening mammography outcomes for women age ≥ 70. According to the most current SEER data, approximately 30% of invasive breast cancers in United States women are diagnosed at ages ≥ 70 (http://seer.cancer.gov/csr). Given that high-quality (randomized trial) data are too sparse to have the statistical power to indicate a statistically significant mortality reduction, the USPSTF should rely on existing observational data from the United States. It would be unconscionable to ignore potentially life-extending benefits of screening mammography for 30% of invasive cancers simply because the randomized trials studied too few elderly women. Consider the following table of data on the performance of screening mammography, stratified by age, published in a large-scale study from one of the

mammography registries that participates in the BCSC (9). These population-based data involve screening mammography from 1996-2006, representing actual United States practice.

Performance of Screening Mammography

Age	Exams	Cancer Detection Rate	Sensitivity	Positive Predictive Value
50-59	186,944	3.7 per 1000	77.3%	22.2%
60-69	116,362	4.9 per 1000	80.1%	29.3%
70-79	75,692	6.2 per 1000	80.4%	37.6%
80-89	23,409	7.9 per 1000	83.4%	40.7%
90-101	1,041	14.1 per 1000	93.8%	55.6%

As demonstrated by the presented data, as well as other U.S. data (10,11), the high cancer detection rate, sensitivity, and positive predictive value of mammography in middle-aged women continues to increase as women become more elderly. This, combined with the fact that age-specific incidence of breast cancer remains high in elderly women, strongly suggests that screening elderly women will be at least as effective in reducing disease-specific mortality as it is in women age 50-69, except as affected by limited life expectancy.

It is especially troublesome for the USPSTF to suggest (even worse to recommend) a policy of stopping screening at a specific age that applies to *all* women. Rather, more appropriate guidance (and guidelines) would be to advise women and their healthcare providers to consider stopping screening on a case-by-case basis, at whatever age is judged appropriate for each woman. This age will vary widely from woman to woman, because there is great variation in life expectancy and co-morbidity at each age (12). For example, among 75-year-old women, approximately 25% will live 17 years, approximately 50% will live 12 years, and approximately 25% will live 7 years. CISNET models and narratives show continued substantial benefit in LYG when healthy women are screened past age 74. This screening is considered “efficient.”

The brief two-sentence paragraph in the USPSTF recommendation statement that discusses the “I” rating given for screening women after age 74 is inconsistent with the “B” recommendation given for women ages 70 to 74 years. In the first sentence of this paragraph, the USPSTF asserts that trial data are inconclusive for women ages 70 to 74 years, implying that trial data evidence is non-helpful. That leaves CISNET modeling data to justify the USPSTF’s previously stated conclusion that screening mammography at ages 70 to 74 deserves a “B” rating. However, the second sentence of this paragraph indicates that the same CISNET modeling data used to justify the “B” recommendation for women ages 70 to 74 also show a net benefit for screening mammography after age 74 years. It is inconsistent to provide a “B” rating for women ages 70 to 74 and an “I” rating for women after age 74 when the determinative data for both groups of women are similar (CISNET modeling). Based on the observational data from United States described above, as well as to achieve internal consistency and improved clinical relevance, the USPSTF should qualify use of the “B” rating for elderly women by indicating that this is meant for women without substantial comorbidity / limited life expectancy, as already done for women ages 70-74, but instead apply this recommendation to elderly women of all ages. After all, performance outcomes for screening mammography are excellent for all elderly women, CISNET modeling suggests that mortality reduction continues beyond age 74, and the harms of screening are higher for women after age 74 only as affected by an increased frequency of co-morbidity and more limited life expectancy (these qualifications already being embedded within the “B” rating).

Benefits of Screening Other than Mortality Reduction

The USPSTF recommendation statement concludes that "the effect of screening mammography on associated adverse effects of treatment or their intensity is not currently clear from the literature." The recommendation statement apparently arrives at this conclusion by juxtaposing the results of a meta-analysis of five of the randomized trials from the 1970s and 1980s that showed women randomized to screening mammography were statistically significantly more likely to have a mastectomy and other surgical therapy than women in the control groups with the counter-argument that these randomized trials do not reflect modern treatment standards and may therefore not be representative of modern practice, citing four more current case-series that compared breast cancer treatments in women who had previous mammography screening with those who did not, each of which reported statistically significantly more breast-conserving surgeries, fewer mastectomies, and less chemotherapy among screened women. Note that the systematic review by Nelson et al. that informed the USPSTF on this issue, citation (2) within the performance statement, failed to include 7 additional current case-series that also reported the same findings (13-19). Given that [a] all 11 of the modern case-series reported that compared to palpation-detected cancer, the surgery provided for women with screening-detected cancer is significantly more frequently breast conservation (lumpectomy) and significantly less frequently mastectomy, and that chemotherapy is provided significantly less frequently and less aggressively for screening-detected cancers, [b] these case series came from different countries, most from the United States, [c] these case-series have been compiled for all women regardless of age, or specifically for women ages 40-49, ages 50-59, and ages ≥ 70, and [d] these case-series involved data on surgery, chemotherapy, or both, it is reasonable to conclude that the case-series data supersede the obsolete randomized trial data, and that current evidence indicates that screening indeed results in less morbidity from cancer treatment, an important secondary benefit of screening.

There are additional benefits of screening, also not acknowledged in the recommendation statement. To parallel the increased anxiety from false-positive results listed as a screening harm (but involving only about 10% of screening examinations) is the reduced anxiety from true-negative results experienced for approximately 90% of screening examinations. There also is reduced anxiety among women with true-positive screening results, women who have chosen to undergo screening because they value the many benefits of screening. Among these women, knowledge that their cancers were detected, diagnosed, and treated earlier and less extensively than would have occurred without screening indeed provides crucial peace of mind that helps them get through the rigors of cancer treatment. Finally, in focusing entirely on the harm of diagnosing (and treating) very early cancers at screening because some otherwise would not have been discovered or caused health problems, the USPSTF does not acknowledge that there is a parallel benefit to the harm that is called "overdiagnosis": the identification of high-risk lesions at biopsy of screen-detected abnormalities that may be addressed by risk reduction strategies, such as treatment with successful chemoprevention regimens (USPSTF "B" recommendation), for lesions that would not have been identified, hence not successfully treated, without screening mammography.

Harms – False Positives

The vast majority of "false positive" results refer to recalls or call-backs from screening for additional imaging with mammography or ultrasound to clarify a finding. Most recalls will be normal and women returned to routine screening. Furthermore, these extra views may occur at the same time of screening so there is rapid temporal resolution of initial finding. While the

USPSTF has chosen to refer to these events using the pejorative term of “false positives”, they are formally considered “incomplete”. The harms of false-positive results are discussed, concentrating on the psychological harms (anxiety/apprehension) that may occur after learning of the need for additional testing, especially when this involves a biopsy. However, the USPSTF recommendation statement omits current evidence indicating that there is no long-term anxiety and no measurable health utility decrement. The recommendation statement also omits current evidence indicating that screened women who experience false-positive results are more likely to undergo subsequent screening mammography than screened women who do not experience false-positive results (OR 2.12, 95% CI 1.54-2.93) (20). This demonstrates that the durable effect of a false-positive result is to encourage rather than discourage subsequent screening. The USPSTF recommendation statement should be revised to include the omitted consequences of false-positive results.

Harms – Overdiagnosis (General Comments)

The USPSTF recommendation statement does not explain to a sufficient extent or with sufficient clarity the limitations on the quality of data on overdiagnosis. The recommendation statement should be revised to add or expand on the following. It is currently impossible to measure overdiagnosis directly, just as it is currently impossible to determine on a case-by-case basis whether a specific screening-detected cancer would never have been detected or caused health problems in the absence of screening. This causes healthcare providers to treat all detected cancers as malignancies, not knowing which ones may be overdiagnosed, resulting in overtreatment (of those cancers that are overdiagnosed). The recommendation statement also should indicate that overdiagnosis is an issue principally concerning ductal carcinoma in situ (DCIS), especially DCIS of low nuclear grade (which accounts for approximately 20% of all DCIS cases) (21). Given that current surgical and oncologic therapy for DCIS is usually less extensive and intensive than therapy for invasive cancer, the harm of overtreatment is relatively diminished. Evidence of overdiagnosis for invasive carcinoma is sparse, at best, the most likely examples being cases of tubular carcinoma (a rare but particularly indolent type of invasive ductal carcinoma). Given that current treatment of tubular carcinoma involves no systemic therapy, the harm of overtreatment also is limited. More fundamental, the concept of overdiagnosis rests upon the assumption that the level of diagnosis absent screening is optimal for women. But this level is not optimal due to the higher breast cancer death rate. Furthermore, breast cancer incidence has exceeded the mortality rate long before screening mammography began in the United States.

Harms – Overdiagnosis (Frequency of Occurrence)

All current approaches to estimating the frequency of overdiagnosis are subject to substantial bias, some more so than others. As described in the USPSTF recommendation statement, estimates of overdiagnosis have ranged widely, from 0% (no overdiagnosis at all) to 54%, depending on methodology, patient population, and other factors. Not described in the recommendation statement is a systematic review of a large series of service screening experience that explains the reasons for observed wide variations in estimates of overdiagnosis (22). To the extent that overdiagnosis exists, estimating its frequency must include methods that adjust for breast cancer risk, lead time, and underlying cancer incidence trends. The systematic review of overdiagnosis studies by Puliti et al comprehensively and clearly indicates that the most plausible estimates of overdiagnosis range from 1% to 10% (on average, approximately 5%), and that higher estimates of overdiagnosis reported in the literature are due to the lack of adjustment for breast cancer risk and/or lead time (22). Indeed, several recently published studies also use estimates (or guesses) rather than observed data in determining

underlying cancer incidence trends, causing these studies to overestimate, often to a great extent, the true magnitude of overdiagnosis.(23)

The USPSTF recommendation statement estimates the frequency of overdiagnosis at 19% based on data reported from three of the randomized trials as provided in the accompanying Systematic Review by Nelson et al, in the recommendation statement: "CNBSS 2, 16% (95% CI, 12.5% to 19.5%); Malmö I, 18.7% (95% CI, 15.1% to 22.4%); and CNBSS 1, 22.7% (95% CI, 18.4% to 27.0%)".

While the USPSTF recognized the need for "follow-up beyond the screening period to distinguish between earlier diagnosis and over diagnosis", the cited overdiagnosis rate is from the screen period ("short-case accrual") and not after follow-up beyond the screening period ("long-case accrual"). The review narrative draft (Nelson, AHRQ p22) appears to have erroneously transposed the results for short case accrual with long case accrual. However, Table 21 correctly identified these values associated with short- and long-term accrual (reproduced below) and is consistent with the source UK review document. The USPSTF considered overdiagnosis the most significant harm, yet confusion regarding this key issue raises grave concerns about the entire process by which its magnitude was estimated. As seen below, the long-case accrual value from these selected studies is 10.7%, not 19%. This value includes DCIS; the invasive cancer rate is even lower.

Table 21. Overdiagnosis Estimates From RCTs Without Screening of Control Groups Trial (reference)	**Age, *years***	**Overdiagnosis, *%* (95% CI) Short-case accrual***	**Overdiagnosis, *%* (95% CI) Long-case accrual***
Malmö I158	55-69	18.7 (15.1 to 22.4)	10.5 (8.4 to 12.7)
CNBSS- I 76	40-49	22.7 (18.4 to 27.0)	12.4 (9.9 to 14.9)
CNBSS- 278	50-59	16.0 (12.5 to 19.5)	9.7 (7.5 to 11.9)
Meta-analysis162	40-69	19.0 (15.2 to 22.7; *I2*=64.8%; p=0.058)	10.7 (9.3 to 12.2; *I2*=22.3%; p=0.276)

In addition, the Systematic Review suffered from poor scholarship by using old data from the CNBSS on overdiagnosis. Table 1 of the 25-year follow-up data from the combination of both CNBSS trials shows 3250 cancers in the screened group and 3133 cancers in the unscreened group, representing a long-term excess of only 3.7% screen-detected cancers that might be termed overdiagnosis (24). Since the USPSTF places great weight on the CNBSS trials as being the least biased in estimating overdiagnosis, and since CNBSS results involve two-thirds of the data used to estimate the overall frequency of overdiagnosis, the USPSTF estimate should be reduced substantially.

We ask that the USPSTF reconsider their estimate of overdiagnosis based on these methodological and reporting errors in the Systematic Review. We also remind the USPSTF of the uncertainty of these estimates, as noted by the UK Independent Review "The issue for the UK screening programmes is the magnitude of overdiagnosis in women who have been in a screening programme from age 50 to 70, then followed for the rest of their lives (25). There are no data to answer this question directly. Any estimate will therefore be, at best, provisional."

One other point about the three RCTs used to estimate the extent of overdiagnosis needs to be mentioned. These 3 trials were designed and conducted to study breast cancer mortality, not

overdiagnosis, as an endpoint. There was no data collected beyond the intervention (screening) period on the frequency of breast cancer screening outside the trial in either invited or uninvited (control) groups. It is well known that service screening started in several Canadian provinces shortly after the end of the CNBSS trials. It is also well documented that the older population studied to estimate overdiagnosis in the Malmo trial had other-cause mortality that may have affected the accuracy of overdiagnosis estimates. Hence, these RCTs may be no more accurate in estimating overdiagnosis than non-RCT-based estimates.

CISNET modeling of invasive cancer overdiagnosis (Table 11) shows median values of only 2-3%, almost all overdiagnosis is attributed to DCIS. The extreme range of invasive cancer overdiagnosis of "1.4% to 24.9%" (note, no median value provided) undermines confidence in this assessment. The upper range occurred in a single model which assumed no temporal incidence change since the 1970s, in contrast to all other models and to known worldwide breast cancer incidence increases over the last 60 years.

In several parts of its text, the USPSTF recommendation statement emphasizes the harm of overdiagnosis as being most important among all harms, but there is no parallel emphasis on the frequency with which a screened woman may experience overdiagnosis. The only indication of frequency is buried in Table 3, at the very end of the recommendation statement, in which the USPSTF calculates that 20 women per 1000 who undergo a lifetime of screening will have an overdiagnosed lesion. To provide appropriate balance, the very low (2%) frequency of this lifetime risk should be indicated in each part of the recommendation statement where the harm of overdiagnosis is described as being the most important (or most serious) harm of screening mammography. This may be done effectively by changing "the most important (or serious) harm" to "the most important (or serious), although infrequent harm".

Harms – Overdiagnosis (How To Mitigate)

Given that the USPSTF has concluded that overdiagnosis is an important (serious) problem, Task Force members must consider how to mitigate the problem. Some have advocated that overdiagnosis may be mitigated by educating radiologists to raise the threshold for considering findings sufficiently abnormal to justify recall for additional imaging or to justify biopsy recommendation after complete diagnostic imaging (additional diagnostic mammography and ultrasound examination) (26, 27). Unfortunately, none of these approaches have been shown to be successful. One cannot define the mammographic appearance (or appearance at any other breast imaging modality) of overdiagnosed cancer versus correctly diagnosed cancer, in no small part because it is impossible to know, even in retrospect, whether a specific screening-detected cancer was overdiagnosed. Furthermore, the American College of Radiology's BI-RADS Atlas defines and illustrates all of the mammographic features of breast cancer, each of which has been established as being suspicious for malignancy based on robust observational studies (7). There is an essentially zero likelihood that one or more of these mammographic features will subsequently be attributed only to overdiagnosed cancer.

Some have advocated that overdiagnosis may be mitigated by delaying the age at which screening starts, and/or lengthening the screening interval from annual to biennial (26). If these simplistic approaches factored into the USPSTF decisions to provide a "C" recommendation for women ages 40-49 and a "B" rating for biennial instead of annual screening for women ages 50-74, then Task Force members must understand that such "solutions" cannot succeed; such strategies will only defer rather than eliminate overdiagnosis. In considering an overdiagnosed lesion that would have been detected at a screening examination deferred by less frequent screening, one must realize that the lesion will still be there at the next mammogram. There is

no published evidence, not even a single case report, of a mammographically visible lesion sufficiently suspicious as to warrant biopsy and ultimately diagnosed as DCIS or invasive carcinoma, that disappeared spontaneously without any treatment. There are occasional examples of cancers not identified at screening mammography due to error in image interpretation, but for each of these cases the lesion is still visible, almost always larger and only rarely the same size, when identified at the next examination. The only breast-imaging-based approach that will mitigate overdiagnosis is to completely eliminate screening mammography, for every woman at any age. But eliminating all screening mammography would mean forgoing the benefits of screening for all women, and the USPSTF continues to endorse screening for all average-risk women as old as age 74.

The USPSTF should understand that there is real danger in placing blame for the harm of overdiagnosis squarely on screening, because uninformed attempts to mitigate that harm by reducing the frequency of screening will inevitably fail, and excessive zeal in mitigating the harm (by recommending no screening at all) will produce the opposite and unintended consequence of underdiagnosis, with resultant delayed cancer detection and missed opportunity for mortality reduction and all other screening benefits. A far preferable approach is for the USPSTF to acknowledge in the recommendation statement that *there is no acceptable breast imaging solution to the problem of overdiagnosis*, and then recommend that a major direction of future research should be to revise the definition of cancer made by pathologists and/or to identify less aggressive/extensive yet successful treatment options for likely-overdiagnosed cancers. Overdiagnosis really has three components: overdetection by the radiologist, overdiagnosis by the pathologist, and overtreatment by the surgeon and oncologist.

Harms – Radiation Risk

The USPSTF estimates of radiation risks appear consistent with prior estimates, except perhaps for the case of women with breasts too large to fit on current digital image receptors. For these women, the risk estimates appear to be overstated, in part because not all women with large breasts require a doubling (or more than doubling) of the number of views acquired. In many cases, a woman with large breasts may require only an additional view in a single view projection. In addition, the USPSTF guidelines fail to advise the public about the fraction of women who might fall into the category of having breasts too large to fit on digital image receptors. At the time of the DMIST screening trial accrual, digital mammography was being performed on prototype or first-generation digital mammography systems and a reasonable fraction (19%) of women included in DMIST required some additional exposures due primarily to breast size. Today, every digital mammography manufacturer (and the great majority of breast screening sites in the U.S.) have large field-of-view digital detectors that equal or exceed the size of large screen-film detectors. Thus, the fraction of women who would require any additional views during screening is approximately the same as the fraction of women requiring additional views on screen-film detectors in DMIST, which is 5-6%. The USPSTF should include this information in their guideline.

In addition, the estimated numbers of radiation-induced cancers and radiation-caused cancer deaths should be expressed using the same denominator as used for absolute mortality reduction from screening (not per 100,000 women), to facilitate comparison of the extremely infrequent occurrence of radiation-induced breast cancers or breast cancer deaths with the relatively much more frequent occurrence of breast cancer deaths averted by screening. General observations are that for mammography examinations performed in women age 40 and older, the estimated frequencies of radiation-induced breast cancer and radiation-induced breast cancer mortality are negligible compared to the demonstrated frequencies of screening-

detected cancers and screening-associated breast cancer deaths averted. These observations are true for both a single screening mammography examination and for multiple screening examinations over a lifetime from age 40 years.

The American Association of Physicists in Medicine has stated that “Risks of medical imaging at effective doses below 50 mSv for single procedures or 100 mSv for multiple procedures over short time periods are too low to be detectable and may be nonexistent(28). Predictions of hypothetical cancer incidence and deaths in patient populations exposed to such low doses are highly speculative and should be discouraged. These predictions are harmful because they lead to sensationalistic articles in the public media that cause some patients and parents to refuse medical imaging procedures, placing them at substantial risk by not receiving the clinical benefits of the prescribed procedures. For reference, the mean effective dose of the typical mammography exam (consisting of two views of each breast) is about 0.5 mSv, so even 40 years of annual screening exams does not approach the effective dose at which the the relationship between radiation exposure to the breast and cancer risk is a significant concern.

Harms of Not Screening

Women at any age who choose not to be screened, as well as women who are unable to be screened if constrained by personal cost considerations that may flow from the “C” recommendation for women ages 40-49 or the “I” recommendation for women above age 74, will forego both the benefits and harms of screening. However, malignancies still will be diagnosed in non-screened women, detected by palpation instead of screening. The USPSTF recommendation statement does not (but should) include discussion of the harms of cancer detection by palpation relative to the harms of cancer detection by screening mammography.

The harms analysis of “false-positive tests “(recalls) and "unnecessary" biopsy recommendations is seriously flawed. Harms of screening have not been compared to harms of non-screened women as stated. The harms analysis incorrectly assumes non-screened women will not undergo false positive tests (such as clinical physical exam), diagnostic breast imaging or “unnecessary” breast biopsies independent of screening. In fact, non-screened women frequently present to their clinician for diagnostic evaluation and biopsy of what eventually proves to be a benign finding. Barton showed that 23% of women (32% of women in their 40s) had a clinical visit for a breast problem in a 10 year period (29). In addition, 6.5% underwent an invasive procedure; nearly all proved benign (or in USPSTF terminology, “unnecessary”). More germane to screening harms judgment was the observation that screened women had significantly fewer symptomatic visits and subsequent work-ups. In addition, Blanchard (in a different study, showed annually screened women’s risk of undergoing biopsy that did not reveal cancer decreased over time to a rate lower than women who did not undergo screening (30). Each of the “harms” occurring among non-screened women can provoke anxiety and discomfort.

Harms of screening should be analyzed as net harms compared to a non-screened population, not simply those documented among screened. To state in the text that harms are compared with no screening is erroneous and misleading. More importantly, if the benefits and harms ratio is altered when net harms are reduced, it is logical that subjective judgment about screening schedule would necessarily be changed.

Use of the CNBSS data

The draft recommendations of the USPSTF utilize data included in the article by Miller et al - Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial (24). Numerous criticisms of the two Canadian National Breast Screening Studies (CNBSS) were published at the time of release (31-36). Its use by the USPSTF in estimating overdiagnosis has been discussed in an earlier portion of this document. In addition, publications dispute the authors' conclusion that annual mammography in women 40-59 does not reduce mortality from breast cancer beyond that of physical examination or usual care, when adjuvant therapy for breast cancer is freely available. The major problems with the CNBSS studies, including inadequate quality of mammography and improper randomization of patient subjects, significantly compromise these trials.

It should be noted that the CNBSS was conducted during the same time frame as other RCTs performed in Europe. Unlike the other RCTs, the CNBSS trial had no reduction in the advanced cancer rate and no mortality reduction in the screened group, showing that this study was an outlier. The relative rate of node-positive cases in the mammography group was 50% higher than in the other RCTs (37).

The Task Force also needs to be aware that poor mammography can substantially decrease the benefit of screening, and this kind of concern was a major factor in the passing of the Mammography Quality Standards Act in 1992. The Task Force should also be aware that Canadians have ignored the CNBSS' no-benefit results and have established guidelines which recommend the use of mammography screening. In contradistinction to the CNBSS, another Canadian study recently published by Coldman et al demonstrated an average 40% mortality reduction in 7 provincial screening programs (3).

Due to the lack of proper randomization and the very poor quality of the mammograms obtained during the study, the USPSTF should exclude the CNBSS from its analysis and not use it to formulate their screening recommendations.

Tomosynthesis

The USPSTF has concluded that current evidence is insufficient to assess the benefits and harms of tomosynthesis as a screening modality for breast cancer, giving it an "I" recommendation). However, the observational data reported in the recommendation statement provide only a small part of the available information on the outcomes reported for tomosynthesis. The USPSTF should add text stating that studies comparing outcomes from digital mammography + tomosynthesis with digital mammography alone demonstrate that combination examinations have lower recall rates and higher cancer detection rates (involving only invasive cancers) and that this applies for women of all breast density categories in all age decades ,with equal or slightly improved positive predictive values (38-41). Hence, combined digital plus tomosynthesis has been shown to decrease false-positives and increase true-positives, with additionally detected cancers unlikely to represent overdiagnosis.

The USPSTF document refers to digital mammography as 2-D and digital breast tomosynthesis (DBT) as 3-D. Tomosynthesis is not a true 3-D examination of the breast. This term was coined by a vendor for marketing purposes and should not be used in a scientific discussion of this modality. The appropriate terms for conventional full field digital mammography might be "planar" or "conventional" digital mammography. The appropriate term for tomosynthesis might be shortened to "DBT."

Other comments

In the United States, compliance with screening guidelines is far less than 100%, whether those of the USPSTF or those of other national organizations. Therefore, a major thrust of the USPSTF recommendation statement should acknowledge this lack of compliance and urge those women and healthcare providers who choose to accept USPSTF recommendations to follow them as presented, specifically not to consider a biennial recommendation as being equivalent to screening every 2½ or 3 years, and not to consider the annual recommendation for those women at ages 40-49 who choose to be screened as being equivalent to screening every 1½ to 2 years.

Update of Previous USPSTF Recommendations: In sentence 1 of paragraph 2, the recommendation statement states "USPSTF did not update its recommendation on the additional benefits and harms of the use of digital mammography or MRI instead of film mammography for breast cancer screening in women not at increased risk (I statement)." This text is troubling, because it combines the use of digital mammography with that of MRI and compares both to the now-almost-obsolete use of film mammography. Does the USPSTF really intend to limit its "C" rating to film mammography, given that more than 96% of mammography facilities in the United States are now using digital technology? (http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityScorecard/ucm113858.htm)

Retaining the text as is makes the USPSTF appear to assert that there is insufficient evidence to support the use of virtually all screening mammography now performed in the United States. Retaining a statement that even indirectly implies digital mammography should have an "I" rating (which will be discounted as being out-of-touch) also substantially dilutes the force of the "I" ratings that the USPSTF gives to the much less widely used technologies of MRI, ultrasound, and tomosynthesis. It is far preferable to avoid confusion by rewording the sentence to compare MRI with "mammography" (not "film mammography"), rather than retaining the clinically obsolete statement about film versus digital mammography.

Tables

Table 1: comments in an earlier section. Please see above.

Table 2 contains erroneous data which grossly overestimates the "number of biopsies needed per case of invasive breast cancer diagnosed." The effect is to inflate harms. Nelson (AHRQ Table 12) shows there were 2.2 invasive cancers per 1000 per screening round, and 16.4 biopsies were recommended for women ages 40-49. Hence, the number of recommended biopsies for a single case of invasive breast cancer is 16.4/2.2= 7.5, NOT 100. This represents a greater than 10 fold error in estimating harms. Furthermore, each value in this row by age decade is incorrect. We remain concerned that the USPSTF used erroneous data, resulting in overstated harms when making their subjective assessment of benefit versus harms.

Table 3 enumerates the lifetime benefits and harms of screening mammography per 1000 women screened, as estimated presumably from the RCT data in Table 1.This needs correction as stated in the above discussion of Table 1. Table 3 presents estimates for all the harms but omits estimates for two categories of benefits. Therefore, two rows of data must be added under the heading "Benefits." One added row should be called "True-negative tests (reduced anxiety)"; the data for this row should be available from BCSC or CISNET modeling data, calculated as TN = all exams – (FP + TP + FN). The other added row should be called "Correctly diagnosed breast cancers." By adding the two currently omitted rows, the

recommendation statement will provide complete and balanced information on the benefits and harms, rather than skewed information on some of the benefits and all of the harms. In addition, the table should refine the estimate of overdiagnosis (see the overdiagnosis discussion above), and state that the overdiagnosis of invasive breast cancers is miniscule, and not as stated. The table does not reflect the current best estimate of the magnitude of overdiagnosis and does not qualify that whatever overdiagnosis exists is most likely DCIS.

Summary

It is the opinion of both the ACR and the SBI that the lack of transparency, lack of breast cancer expertise on the Task Force, and selective analysis of available evidence impair the legitimacy of the draft recommendations. We strongly feel that the conclusions and recommendations of the Task Force will negatively impact public health.

The ACR and SBI believe that all of the benefits of screening should be discussed, and not just the mortality benefit. We also strongly suggest that the magnitude of the mortality benefit is better estimated by inclusion of observational studies. The narrative above also describes why the harms of screening are inappropriately estimated and overemphasized in the draft document. The above errors and comments need to be addressed.

Since the recommendations are based on the judgment of the panel, we suggest that it be made very clear to women in the United States that the greatest number of lives saved is derived from a strategy of annual screening beginning at age 40. Rather than emphasizing that the cumulative harms should preclude a straightforward screening recommendation, women should be informed that for those who rate the harms of screening more important than its benefits, screening might be delayed or less frequent, but that those women would be giving up potential life-saving benefits. We suggest that the appropriate recommendation should read, "**women should start annual screening at the age of 40, unless they put higher value on the potential harms of screening, and that choice should be an individual one." (B recommendation)**

Respectfully Submitted,

Barbara Monsees, MD, FACR
Chair, Commission on Breast Imaging
American College of Radiology

Debra Monticciolo, MD, FACR
Chair, Commission on Quality and Safety
American College of Radiology

Murray Rebner, MD, FACR, FSBI
Director & Immediate Past President
Society of Breast Imaging

References

1. Developing Trustworthy Clinical Practice Guidelines, Board of Health Care Clinical practice guidelines we can trust / Committee on Standards for Services, Institute of Medicine of the National Academies; R. Graham, et al. 2011.

2. Broeders M, Moss S, Nyström L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012; 19 Suppl 1:14-25.

3. Coldman A, Phillips N, Wilson C, et al. Pan-Canadian study of mammography screening and mortality from breast cancer. JNCI J Natl Cancer Inst 2014, 106(11):dju261 doi:10.1093/jnci/dju261.

4. Tabár L, Vitak B, Chen TH-H, et al. Swedish Two-County Trial: impact of mammographic screening on breast cancer mortality during 3 decades. Radiology 2011; 260(3):658-663.

5. Hellquist BN, Duffy SW, Abdsaleh S, et al. Effectiveness of population-based service screening with mammography for women ages 40 to 49 years: evaluation of the Swedish Mammography Screening in Young Women (SCRY) cohort. Cancer 2011; 117(4):714-722.

6. Avalere Health: www.avalere.com

7. Sickles EA, D'Orsi CJ, Bassett LW, et al. ACR BI-RADS Mammography. In: ACR BI-RADS Atlas, Breast Imaging Data and Reporting System. Reston, VA: American College of Radiology, 2013.

8. Bennett RL, Sellars SJ, Moss SM. Interval cancers in the NHS breast cancer screening programme in England, Wales and Northern Ireland. British Journal of Cancer (2011) 104, 571- 577.

9. Sinclair N, Littenberg B, Geller B, Muss H. Accuracy of screening mammography in older women. AJR Am J Roentgenol 2011; 197(5):1268-1273.

10. Faulk RM, Sickles EA, Sollitto RA, Ominsky SH, Galvin HB, Frankel SD. Clinical efficacy of mammographic screening in the elderly. Radiology 1995; 194(1):193-197.

11. Solin LJ, Schultz DJ, Kessler HB, Hanchak NA. Downstaging of breast carcinomas in older women associated with mammographic screening. Breast J 1999; 5(2):94-100.

12. Walter LC, Covinsky KE. Cancer screening in elderly patients: a framework for individualized decision making. JAMA 2001; 285(21):2750-2756.

13. Maibenco D, Daoud Y, Phillips E, Saxe A. Relationship between method of detection of breast cancer and stage of disease, method of treatment, and survival in women aged 40 to 49 years. Am Surg 1999; 65(11):1061-1066.

14. Barth RJ Jr, Gibson GR, Carney PA, Mott LA, Becher RD, Poplack SP. Detection of breast cancer on screening mammography allows patients to be treated with less-toxic therapy. AJR Am J Roentgenol 2005; 184(1):324-329.

15. Malmgren JA, Parikh J, Atwood MK, Kaplan HG. Impact of mammography on the course of breast cancer in women aged 40-49 years. Radiology 2012; 262(3):797-806.

16. Malmgren JA, Parikh J, Atwood MK, Kaplan HG. Improved prognosis of women aged 75 and older with mammographically detected breast cancer. Radiology 2014; 273(3):686-694.

17. Coldman AJ, Phillips N, Speers C. A retrospective study of the effect of participation in screening mammography on the use of chemotherapy and breast conserving therapy. Int J Cancer 2007; 120(10):2185-2190.

18. Zorzi M, Puliti D, Vettorazzi M, et al. Mastectomy rates are decreasing in the era of service screening: a population-based study in Italy (1997-2001). Brit J Cancer 2006; 95(9):1265-1268.

19. Spillane AJ, Kennedy CW, Gillett DJ, et al. Screen-detected breast cancer compared to symptomatic presentation: an analysis of surgical treatment and end-points of effective mammographic screening. ANZ J Surg 2001; 71(7):398-402.

20. Tosteson ANA, Fryback DG, Hammond CS, et al. Consequences of false-positive screening mammograms. JAMA Intern Med 2014; 174(6):954-961.

21. Ponti A, Lynge E, James T, et al. International variation in management of screen-detected ductal carcinoma in situ of the breast. Eur J Cancer 2014; 50(15):2695-2704.

22. Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen 2012; 19 Suppl 1:42-56.

23. Bleyer A, Welch HG. Effective Screening Mammography on Breast Cancer Incidence. N Engl J Med 2013; 368:679.

24. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomized screening trial. BMJ 2014; 348:g366 doi: http://dx.doi.org/10.1136/bmj.g366.

25. Marmot MG, Altman DG, Cameron DA et. al British Journal of Cancer (2013 108, 2205-2240

26. Esserman LJ, Thompson IM Jr, Reid B. Overdiagnosis and overtreatment in cancer. JAMA 2013; 310(8):797-798.

27. Flowers CI, O'Donoghue C, Moore D, et al. Reducing false-positive biopsies: a pilot study to reduce biopsy rates for BI-RADS 4A/B assessments through testing risk stratification and new thresholds for intervention. Breast Cancer Res Treat 2013; 139(3):769-777.

28. AAPM Position Statement on Radiation Risks from Medical Imaging Procedures. PP 25-A12/13/2011 through 12/31/2016. Available on-line at: http://www.aapm.org/org/policies/details.asp?id=318&type=PP&current=true

29. Barton MB, Elmore JG, Fletcher SW. Breast symptoms among women enrolled in a health maintenance organization: frequency, evaluation, and outcome. Ann Intern Med. 1999;130: 651-657.

30. Blanchard, Karen, et al. "Long-term risk of false-positive screening results and subsequent biopsy as a function of mammography use 1." Radiology (2006); 240 (1): 335-342.

31. Boyd NF, Jong RA, Yaffe MJ, Trichtler D, Lockwood G, Zylak CJ. A critical appraisal of the Canadian National Breast Cancer Screening Study. Radiology 1993; 189(6):661-663.

32. Kopans DB, Feig SA. The Canadian National Breast Screening Study: a critical review. AJR Am J Roentgenol 1993; 161(4):755-760.

33. Warren Burhenne LJ, Burhenne HJ. The Canadian National Breast Screening Study: a Canadian critique. AJR Am J Roentgenol 1993; 161(4):761-763.

34. Tarone RE. The excess of patients with advanced breast cancer in young women screened with mammography in the Canadian National Breast Screening Study. Cancer 1995; 75(4):997-1003.

35. Kopans DB. The Canadian Screening Program: a different perspective. AJR Am J Roentgenol 1990; 155(4):748-749.

36. Yaffe, MJ “Correction: Canada Study.” J Natl Cancer Inst 85, no. 2 (1993): 94.

37. Tabar L, Gad A, Holmberg LH, et al. Reduction in Mortality from Breast Cancer after Mass Screening with Mammography. The Lancet, Volume 325, Issue 8433, 13 April 1985, pages 829-832.

38. Rose SL, Tidwell AL, Gujnoch LJ, et al.Implementation of breast tomosynthesis in a routine screening practice: an observational study. AJR Am J Roentgenol 2013; 200(6):1401-1408.

39. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology 2013; 267(1):47-56.

40. Friedewald SM, Rafferty EA, Rose SL, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. JAMA 2014; 311(24):2499-2507.

41. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol 2013; 14(7):583-589.