An open letter to panels that are deciding guidelines for breast cancer screening

Daniel B. Kopans

Abstract Panels are presently reviewing breast cancer screening guidelines. It is critical that they understand which publications are scientifically valid, and which analyses are methodologically flawed and not valid. The scientific evidence clearly supports annual mammography screening beginning at the age of 40. The analyses that suggest that screening leads to overdiagnosis of invasive breast cancers are flawed and incorrect. There is little if any overdiagnosis of these cancers. The vast majority of breast cancers occur in women who are not at elevated risk so that excluding them from screening and only screening high risk women will deny the benefits of early detection to most women who develop breast cancer. Guidelines panels should not make decisions that exclude women from screening. Women should be provided with accurate information so that they can make informed decisions and have unimpeded access to screening if that is their preference.

Keywords Breast · Breast cancer · Screening · Mammography

Introduction

Breast cancer screening guidelines are under review by a number of organizations including the United States Preventive Services Task Force (USPSTF), the American Cancer Society (ACS), and the International Agency for Research on Cancer (IARC). At the present time, the American Cancer Society recommends annual mammography beginning at the age of 40 for all women [1], while the USPSTF recommends that high risk women discuss screening in their forties with their physicians, but advises that most women wait until the age of 50, and then be screened every 2 years [2]. The IARC also recommends women wait until the age of 50 and then be screened every 2–3 years until the age of 69 [3].

There has been a great deal of misinformation about screening that has made its way into the medical literature, and has then been passed on through the media to the public. The purpose of this summary is to elucidate some of the misinformation that has been promulgated around breast cancer screening with the hope that the new guidelines for screening will be based on the scientific evidence and not faulty analyses.

Not all articles should have passed peer review

Unfortunately, some previous “guidelines panels” have treated many publications as if they are all credible. Since some of the panels, such as the USPSTF, do not involve experts in breast cancer, they rely on outside analyses that they commission. If a panel is going to issue important guidance, the panel members should know all of the details in the various published studies. Review committees should understand that there are a number of analyses that
have made their way into the literature that are not, scientifically, supported. Much of the misinformation is based on so-called “registry reviews” that focus on summary numbers from tumor registries without direct patient data. The analysts do not know who had mammograms nor do they know which cancers were detected by mammography since they do not track actual patient data. Panels need to carefully review these papers by authors who were not actually in the countries whose data they have reviewed. These authors have made claims that mammography has had little impact [4–10], but careful review shows that these papers are methodologically flawed. For example a number of the papers that claim that there had been little change in deaths from breast cancer than would be expected with the institution of national screening programs, did not take into account the fact that most of the deaths that occur in the years after screening programs are initiated, are among women diagnosed with breast cancer before there was access to screening so they could not possibly have benefited from screening. These deaths should not have been counted. Panels also need to be aware that there may have been considerable screening in the countries prior to the start of national programs so that a decline in deaths may have already occurred prior to the start of the national program reducing the apparent impact following the start of national screening programs. Other analysts claiming little impact from screening have compared women in different regions whose populations have differing risks of developing breast cancer and are not comparable. Another error that has been repeated is assuming that the incidence of breast cancer had been stable prior to the onset of screening. In the US and other countries, the incidence of invasive breast cancer had been increasing steadily for decades prior to the start of screening. Finally, many seem to not understand that that screening should not be expected to cause an immediate decline in deaths. Due to “length bias sampling” a delay of 5 or more years should be expected before there is a decline in deaths due to screening.

One paper that was highlighted by the media claiming little benefit from mammography when screening was introduced nationally in Norway, was published in the prestigious New England Journal of Medicine (NEJM) [11]. Compromising that review was the fact that it had only 2.2 years of follow-up. Since the benefit of breast cancer screening begins to appear after 5–7 years, 2.2 years was far too short to see the impact of screening. Furthermore, the authors’ assessment relied on their claim that few women in Norway were participating in screening prior to the start of the National screening program. In fact, it was subsequently revealed that more than 40 % of women in Norway were participating in screening prior to the program [12]. Screening was already decreasing mortality in Norway before the start of the National program which compromised their findings. A subsequent paper from Norway showed that screening was associated with a more than 40 % reduction in deaths [13].

Observational support for screening

Analysts studying the effects of screening who are actually within the countries, and have used direct patient data, show that the death rate is lower among women with access to screening, and even lower for women who actually participate in screening. These direct studies show that mammography screening is associated with major reductions in breast cancer deaths [14–22]. This also applies to women in their forties [13, 17, 23–25].

There is little if any “overdiagnosis of invasive breast cancer”

Another unsubstantiated claim is that mammography screening leads to massive overdiagnosis of breast cancers. This is the detection of cancers which, if left untreated, would regress or disappear on their own [5, 7]. The paper that has had a major impact was also published in the NEJM [26] in which the authors claimed, based on the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, that in 2008 alone, more than 70,000 women were diagnosed with cancers that would have regressed or even disappeared had they not been detected by mammography. Not only has no one ever seen an invasive breast cancer regress or disappear without therapy (70,000 a year and not a single credible case report), but since they had no data on mammography and no data on which cancers were found by mammography, as one of the authors subsequently admitted, they could not, legitimately, fault mammography [27], which has, nevertheless, been blamed.

Another fundamental flaw in their analysis, also admitted [27], is that it was based on their “best guesses”. Had they used direct data from the Connecticut Tumor Registry (CTR), which has been used by numerous other analyses that evaluated the incidence of breast cancer prior to the start of the SEER program, they would have found that their conclusions are simply false and that there has been no overdiagnosis of invasive breast cancer [28]. Etzioni et al. approached the question using different methods and found that, in order to have such a high rate of “overdiagnosis”, mammography would have to be finding small invasive cancers 9 years before the would become clinically evident [29]. The lead time for mammography has been reported to be, at most 3–4 years. In order for
such massive overdiagnosis, American radiologists would have had to find cancers with lead times that no one has ever seen. The overdiagnosis claims of the NEJM paper are simply not supported by the facts.

The other clam in the NEJM paper was that the rate of advanced cancers had not declined very much in the screening era. It has been argued that screening does not work unless there is a decline in late stage cancers. This is not absolute. A decline in cancer sizes within stages has also been linked to reduced mortality, but certainly a reduction in late stage disease is an indication that screening is having a benefit. Helvie et al. have shown that the contention in the NEJM article is also not supported by the data [30]. Using data from the US, and around the world, they showed that there has been a reduction in late stage cancers that is as high as 48%.

Other analysts have also “guessed” as to what the incidence of breast cancer would have been had screening not been made available, or they have tried to compare results between countries. A thorough analysis by Puliti et al. [31] showed that claims of overdiagnosis in these papers were due to these authors failing to take into account lead time, as well as differing risks of breast cancer in different comparison populations. Puliti et al. concluded that there was little if any overdiagnosis due to mammography.

The Canadian national breast screening studies were compromised

The current panels need to know that the Canadian National Breast Screening Studies (CNBSS), that recently published 25-year results [32], were, unfortunately, hopelessly compromised. It is clearly documented that these two trials of mammography screening utilized poor to unacceptable quality mammography proven in a review conducted by the trial organizers. I was one of the reviewers [33]. Their own reference physicist has stated “…in my work as reference physicist to the NBSS, [I] identified many concerns regarding the quality of mammography carried out in some of the NBSS screening centers. That quality [in the NBSS] was far below state of the art, even for that time (early 1980s)” [34]. There was no training for the radiologists and no training for the technologists. Some sites used obsolete devices. It is indisputable that mammography finds smaller cancers than are evident on clinical examination, yet the cancers in the mammography arm were the same size as those in the “usual care” arm. The suggestion that because the women were examined by highly trained nurses this allowed them to find very small cancers is belied by the fact that most, mammographically detected cancers, cannot be palpated by surgeons even when their location is identified. Furthermore, most of the cancers in the “usual care group” were, supposedly, found by “usual care”. There is no evidence that “usual care” women could have their cancers detected at the same size as cancers detected by high quality mammography further confirming the fact that the mammography was poor quality. The results could only occur because the poor quality mammography failed to detect small cancers.

Furthermore, the CNBSS were trials of volunteers and not population based and had a high contamination rate in the control group (more than 20% of the control women had a diagnostic mammogram that starts with bilateral screening images).

Nonblinded allocation in the CNBSS

Perhaps of greatest concern is the fact that the CNBSS violated the fundamental requirement of a randomized, controlled trial—namely blinded allocation. Ignoring the fundamental rules of an RCT, all the women first underwent a clinical breast examination that identified “lumpy” breasts; obvious cancers; and cancers with associated axillary adenopathy indicative of advanced cancers. Even a review falsely claiming to absolve the CNBSS of allocation imbalances [35], documented the fact that this information was provided to the study coordinators who assigned the women to the mammography or “usual care” arms. Compounding the error was the fact that the women were assigned on open lists so that the coordinator need only skip a line to be certain that a woman was placed in the desired arm of the trial. The allocation violations resulted in an excess of women with advanced cancers assigned to the screening arm in the CNBSS1 [36, 37]. When there were more deaths among the screened women ages 40–49 [38], the investigators made unsubstantiated claims (subsequently withdrawn) [39] that mammography was squeezing cancer cells into the blood. The obvious answer to the excess deaths was that when you assign more women with advanced breast cancers to the screening arm, who are destined to die from breast cancer, it should be no surprise that more women died. The likelihood that women were non-randomly assigned is further supported by the fact that the 5-year survival rate from breast cancer in Canada at the time was only 75%. The women in the “usual care” arm had a better than 90% 5-year survival [38]. The only way that this could happen is if women destined to die, who should have randomly been in the control arm, were non-randomly assigned, instead, to the mammography arm. There is no other explanation. The 5-year survival in the US, a quarter of a century later is still less than 90%.

The CNBSS violated the fundamental rules for RCT’s and its results are unreliable. Just as the Edinburgh trial has
been excluded from analyses because of a socioeconomic imbalance, the CNBSS results are also not reliable, and it should not be included in reviews of screening efficacy.

Panels exclude the experts

Part of the problem is that, in an effort to reduce the possibility that guidelines will be influenced by financial interests; experts have been excluded from the review panels because of “conflicts of interest” (COI). It is almost axiomatic that if you do not have a “conflict of interest” then you have no expertise in a field. By excluding experts, some of the panels that have reviewed screening guidelines in the past have been compromised by the inexperience of the members. Not only does this mean that they will be unaware of many of the important nuances in the debate, but they are also subject to “behind the scenes” guidance by “advisors” who do have biases, who are not on the panel, but can present the data in a way that sways the naive panel members. As a consequence, panels such as the 2009 United States Preventive Services Task Force had no scientific credibility and arrived at conclusions that were not supported by the scientific evidence [40–42] leading to major controversy, confusion, and a lack of acceptance of the panel’s conclusions.

It would be far better if guidelines panels comprised documented experts in the field. All potential conflicts of interest (including the finances of granting agencies and foundations) should be made public. All of the panel deliberations should be in public. If there is not unanimity in the conclusions, then a minority report should also be issued with an explanation of the reasons behind the disagreement.

Ignoring their charter

The US Preventive Services Task Force is prohibited from analyzing information based on the cost of the intervention. In 2009, the Task Force ignored this basic rule and relied on a measure called the “number needed to screen” (NNS). This was a calculation as to how many women needed to be screened for screening to save a single life. NNS is nothing more than a surrogate for the cost to save one life. The individual woman is, probably, not very interested in this figure. The USPSTF made subjective decisions (also a violation of their own guidelines) and concluded that the figure which they used for NNS for women ages 40–49 (1900) was too high to justify recommending screening for these women, yet they felt (completely subjective) that the NNS for women ages 50–59 (1300) was fine for recommending screening. The NNS was determined by the USPSTF using the lowest possible estimate of lives saved (15 %). Even using their faulty methodology, had they used a mortality reduction of 30 % (which is indicated by all the data) the NNS for women ages 40–49 would be 950 which is within their subjective threshold. Compounding the misinformation provided by the 2009 USPSTF, is the fact that they called their numbers “number needed to screen”, when, in fact, they based this figure from the RCT’s which were actually the “number needed to invite to screen” (NNI). In most of the trials, women were invited to be screened. Many refused. Even using their faulty estimate, the actual NNS is much lower than what the USPSTF used which was actually the NNI. The actual NNS was much lower [43], and the number is even lower if CISNET models are used [43]. Further compounding their errors was the fact that the USPSTF grossly underestimated years of life saved [43]. These mistakes made by a panel that lacked experience, has, nevertheless, resulted in convincing doctors and their patients that women in their forties should forego screening [44].

Guidelines panels need to know the fundamental facts

1. The randomized, controlled trials (RCT’s) have proved that early detection saves lives for women who begin screening at the age of 40 [45, 46]. The RCT’s did not include women over the age of 74 so there is no proof of benefit beyond this age, but if a woman has a sufficiently long life expectancy and reasonable quality of life and she does not want to die from breast cancer there is no reason to expect that she cannot benefit from early detection.

2. Observational studies that have looked directly at patient data have shown that when screening is introduced into the general population the breast cancer death rate declines [12–23].

3. Annual mammography saves more lives than screening every 2 years [47]. The CISNET models used by the 2009 USPSTF all show that the most lives are saved by annual mammography beginning at the age of 40 [48].

4. Using the CISNET models Helvie and Hendrick showed that had women who, at that time, were in their thirties followed the 2009 USPSTF guidelines, and waited until the age of 50 to begin screening, and then participated every 2 years, as many as 100,000 lives would be lost that could have been saved by annual mammography beginning at the age of 40 [42].

5. There is no biological or scientific support for using the age of 50, originally chosen as a surrogate for
6. The age of 50 has been made to appear as a legitimate threshold by grouping the data for women under the age of 50 together and averaging and comparing them to women ages 50 and over grouped and averaged [50]. This makes continuously changing variables appear to change suddenly at the age of 50 when the reality is that this does not happen [49].

7. Analysts were misled thinking that an immediate reduction in deaths among women ages 50 and over meant that screening was more effective among older women. In fact, an immediate reduction in deaths from periodic screening is, virtually, impossible due to length bias [51]. The appearance of an immediate benefit was simply statistical fluctuation. A “delayed” benefit (5–7 years) is what is expected [52].

8. The incidence of breast cancer increases with age. An older woman has a higher risk than a younger woman, but this is true for women at any age. The arguments suggesting that screening is more effective in women age 50 or older are not supported. The actual number of women diagnosed with breast cancer is determined by the incidence at that age multiplied by the number of women at a given age at the time. The absolute number may, in fact be higher among younger women (as it was in 1995) if the numbers of younger women are larger than the comparison group of older women. Approximately 30–35,000 women are diagnosed with breast cancer each year while in their forties.

9. There is no scientific support for screening only high risk women. None of the randomized controlled trials (RCT) stratified by risk so that there is no RCT proof that screening only high risk women will save lives. Furthermore, only 25 % of women diagnosed with breast cancer each year are at elevated risk so that most women (75 %) with breast cancer would not have access to early detection [53] if only high risk women are screened.

10. Overdiagnosis by mammography has been grossly exaggerated by scientifically unsupported “guesses” and faulty estimates. There is little if any overdiagnosis of invasive breast cancers [31].

11. The use of the term “false positive” is misleading. Most of these are “recalls from screening”. The rate of recalls from mammography (10 %) is the same as the rate of recalls from cervical cancer screening (Pap testing). Most are resolved by a few extra pictures or an ultrasound. Only 1–2 % of screened women have an imaging guided needle biopsy using local anesthesia and 20–40 % of these reveal cancer. This is actually a much higher yield of cancer than when biopsies are done for palpable abnormalities [54], and cancers detected by mammography are more likely to be cured than those that are palpable. Despite what the 2009 USPSTF concluded, the anxiety, inconvenience, and the few imaging guided needle biopsies using local anesthesia associated with a recall from screening, are not equivalent to dying from breast cancer.

**Conclusion**

Randomized, controlled trials (RCT) are the only way to prove the efficacy of screening. These have shown a significant decline in deaths for women ages 40–74. We have no truly national statistics on mammography utilization in the US, but, based on the sudden increase in breast cancer incidence in the mid 1980s it is clear that mammography screening began at a national level at that time [22]. Supporting this conclusion is the fact that there was a sudden increase in the diagnosis of ductal carcinoma in situ (DCIS) at the same time [55]. Prior to 1985 ductal carcinoma in situ (DCIS) had made up only 3–5 % of cancers diagnosed each year. In the mid-1980s DCIS began to be diagnosed with increasing frequency [56]. Since DCIS is found almost exclusively by mammography, this made it fairly clear that screening was being utilized in sufficient numbers to affect national statistics. Since screening does not save lives immediately (length bias), it seems fairly certain that the similarly sudden decline in breast cancer deaths that began in 1990, was, in large part, due to screening. For the first time in 50 years the death rate from breast cancer began to decline. There are now 35 % fewer women who die each year than would have died had the death rate continued at the 1989 level.

Those seeking to limit access to screening claim that the major decline in breast cancer deaths since 1990 is due to advances in therapy. There have been improvements in therapy, but there is actually no direct evidence that therapy is the main reason for reduced deaths. The main support comes from computer modeling published in the New England Journal of Medicine [57]. The results of modeling are completely determined by the assumptions built into the models so that a model can arrive at whatever results you program in. This is clearly evident in the fact that the 7 models that were queried for the NEJM paper yielded 7 disparate results. One model calculated that screening was
responsible for only 28% of the decline in deaths while another model showed that it was responsible for 65% of the reduction, with the others in between. In fact, in actual studies of women, where all women have had access to modern therapies, it has been shown that women who participate in screening have a markedly lower death rate from breast cancer than those who do not participate despite the fact that all have access to the same therapies. This is true for women in their forties as well [17, 20–23]. In our own study in the Harvard teaching hospitals, more than 70% of the women who died from breast cancer were among the 20% not participating in screening [58].

Although not proof, males with breast cancer have access to the same therapy as females, but the death rate for males actually increased in the 1990s. Then in 2005 it returned to where it had been in 1990 with no parallel decline as has been seen in women. Men present with larger and later stage tumors. Certainly one of the major differences is that women have been participating in screening while men have not. Therapy has improved, but therapy saves lives when breast cancers are treated earlier.

The Panels reviewing screening guidelines need to carefully evaluate the data and understand the analyses so that women and their physicians can rely on their guidance. The data clearly support annual mammography beginning at the age of 40.

Conflict of interest Dr. Kopans has no conflict of interest to disclose.

References