Beyond the mammography debate: a moderate perspective

C. Kaniklidis, Research Director, No Surrender Breast Cancer Foundation*

ABSTRACT

After some decades of contention, one can almost despair and conclude that (paraphrasing) “the mammography debate you will have with you always.” Against that sentiment, in this review I argue, after reflecting on some of the major themes of this long-standing debate, that we must begin to move beyond the narrow borders of claim and counterclaim to seek consensus on what the balance of methodologically sound and critically appraised evidence demonstrates, and also to find overlooked underlying convergences; after acknowledging the reality of some residual and non-trivial harms from mammography, to promote effective strategies for harm mitigation; and to encourage deployment of new screening modalities that will render many of the issues and concerns in the debate obsolete.

To these ends, I provide a sketch of what this looking forward and beyond the current debate might look like, leveraging advantages from abbreviated breast magnetic resonance imaging technologies (such as the ultrafast and twist protocols) and from digital breast tomosynthesis—also known as three-dimensional mammography. I also locate the debate within the broader context of mammography in the real world as it plays out not for the disputants, but for the stakeholders themselves: the screening-eligible patients and the physicians in the front lines who are charged with enabling both the acts of screening and the facts of screening at their maximally objective and patient-accessible levels to facilitate informed decisions.

Key Words Mammography, screening, breast cancer, controversy, debate, benefits, harms

CURR ONCOL. 2015 June;22(3):220-229

REFLECTIONS ON SOME CENTRAL THEMES

All screening programmes do harm; some do good as well, and, of these, some do more good than harm at reasonable cost.

— Sir Muir Grey, former director of the U.K. National Screening Committee

Some Issues Deserving More Attention

Several core issues require closer examination:

1. The potential benefits of mammographic screening independent of whether ultimate survival is affected, including diagnosis at earlier stages, typically with smaller tumours and node negativity; reduced likelihood of aggressive treatments and morbidities; detection of high-risk lesions (most diagnosed in the screened group) allowing for chemoprevention or magnetic resonance imaging (MRI) surveillance against occult malignancies, or both; and avoidance of compromised quality of life after diagnosis of advanced disease.

   It is not clear, as often assumed, that survival is the best (or only) measure for judging mammographic screening. And it remains open whether claimed harms truly are disproportionate to benefits when using, for instance, a more nuanced definition of over-detection than simply breast cancer (BCa) diagnosed but not BCa expired (distinguishing, for example, longer recurrence-free survival from simple overall survival benefit).

2. The benefit of screening in the context of modern adjuvant therapy beyond the many screening trials begun 30–40 years ago.

3. The issue of mammography being available outside the screening program.

As to “extra-program” mammography, there are many trials, such as the Malmö Mammographic Screening Trial in which mammography was available...
outside the screening program, and the Pan-Canada Study\(^6\), with screening unlikely to have terminated at trial conclusion. For protocol integrity, we would typically require that neither the control nor the study group be screened post-trial, otherwise overdiagnosis (strictly speaking, “overdetection,” although I use these terms interchangeably for convenience) could be overestimated, given the absence of a fair opportunity for the compensatory drop in bca incidence expected after the end of screening\(^7\).  

4. Variable length of follow-up

As to this factor, I note that it is not implausible that more aggressive bcas can lead to significant bca-specific mortality during the first 10 years without early detection and surgical removal, while some more growth-indolent cancers could incur mortality after 10–20 years of follow-up in the absence of screening, given potentially significant intra-tumour heterogeneity, with early screening detection preventing many small or well- or moderately-differentiated tumours from developing into larger, more poorly differentiated tumours, against the common claim that mammography screening primarily detects mostly indolent cancers, recognizing some propensity for dedifferentiation and worsening of tumour malignancy grade as disease progresses\(^8\)–\(^13\).

**Individualized Patient Data**

It further follows that there may be significant methodologic limitations in studies without rich discriminatory power founded on individualized patient data:

1. Tumour detection modes\(^14\) differentiating (a) tumour detection at screening, (b) inter-screening interval detection, and (c) detection within the subgroup invited to but not attendant at screening, compared with (d) detection in non-invited non-attendant women

2. Ability to ascertain bca-specific mortality compared with other-cause mortality

3. Benefits and harms relativized to specific bca molecular subgroups or phenotypes, because it is highly unlikely that tumours detected would behave substantially the same regardless of whether they are more indolent endocrine-positive, versus Her2-positive, versus more aggressive triple-negative disease (and which might themselves have to be differentiated as to subtype)

No study failing to find a mortality benefit from early detection by screening had full access to these discriminatory data, without which reliable assessment of screening impact is significantly compromised, acknowledged by many\(^15\), even some of screening’s strongest critics\(^16\)–\(^18\), allowing for continued indeterminacies and controversies.

**Screening Attendance Versus Invitation**

Woman non-attendant at mammographic screening cannot be expected to receive any benefit from screening, and so ideally the subgroup of screening-invited but non-attendant women should not be included in any study group measuring the impact and survival value of mammographic screening. This failure infects several trials and reviews that rely solely on invitation to screening (U.S. Preventive Services Task Force (USPSTF) systematic review\(^19\)–\(^22\), Nordic Cochrane Review/Meta-analysis\(^23\) and others\(^24\)). In contrast, studies that relied solely on women actually attendant at regular screening\(^1\)–\(^3\),\(^25\),\(^26\) concluded in favour of a significant mortality reduction from screening mammography. Note, however, that the methodologies of several of the Swedish trials have been the subject of challenge (including by Dr. Narod\(^27\)), with some justice. The devil is, again, in the methodology.

**Trial Consistency**

Recently Peter Jüni and Marcel Zwahlen analyzed mammography trials\(^28\) as to trial consistency with respect to the reduction in bca versus non-bca deaths, using that index as a marker of proper trial design, because mammography cannot be expected to incidentally detect treatable non-breast causes, only to reduce bca deaths. A claimed screening benefit on non-bca deaths would tend to suggest baseline imbalances (selection bias), random occurrence, or performance bias (differences in care favouring the screening group\(^29\)). Alternatively, a significant increase in non-bca deaths suggests either baseline imbalances favouring the control group or detection bias from differential misclassification of deaths (screening group favoured for bca deaths, control group for non-bca deaths\(^28\),\(^29\)). Eleven screening trials were meta-analytically assessed for trial consistency, with consistent trials [HIP/New York Trial\(^30\)–\(^32\), Malmö Mammographic Screening Program Trial\(^33\), Östergötland County Trial\(^33\),\(^34\), Canadian National Breast Screening Study (CNBSS) \(^13\) , CNBSS-2\(^36\), U.K. Age Trial\(^37\)–\(^41\)] evidencing a 15% reduction of bca deaths and no reduction in non-bca deaths.

**Randomization**

Although it is widely claimed that the published patient-specific data comparing the two groups in the cnbss trials is suggestive of some patients with palpable findings being preferentially assigned to the screening arm, such assignment is disputed by the Canadian trialists. Thus, in this issue of Current Oncology, Dr. Yaffe\(^42\) regards, plausibly, as possible evidence of randomization protocol compromise, first, that a 1.09 hazard ratio (HR) was found in the screened group compared with the control group, but only a 0.9 HR when adjusted to exclude deaths from prevalence screen-detected cancers, a 19% drop; and second, that the trial found a HR for bca mortality in the mammography arm of 1.46 for prevalence screen-detected cancers, suggesting that women randomized into the mammography arm of the trial were at surprisingly elevated risk of death from bca. Dr. Narod counters with a rival interpretation to not implausibly fit the same facts, concluding that the odds ratio for deaths among cases diagnosed during the prevalence screens should in fact be expected to be higher. He reasons that 106 of the 142 cancers in the initial excess should be ascribed to overdiagnosis and 36 to early diagnosis, deaths from those 36 cancers being counted among the prevalence screens in years 2–4, causing the odds ratio for death among the prevalence screen-detected group to be higher than that in the incident screen-detected group, so that the HR computed for the combined years 1–4 is what we must properly rely on, and not the component phases individually.
Two distinctly divergent narratives thus arise from one and the same set of underlying facts. Nonetheless, I would argue that we should be able to agree on the more fundamental fact that, on good methodologic practice, symptomatic patients should ideally be excluded from any screening trial, such methodologically robust screening trials being exclusively intended for asymptomatic patients.

**Overdiagnosis (Overdetection): Heart of Darkness**

If our attention is restricted to only those studies that explicitly account and control for cancer incidence during screening and for lead time (length of time between detection by screening and when a bca diagnosis would have been made absent screening), given that length-time bias (less severe cases diagnosed given disease heterogeneity) can engender overdetection if there is unnecessary treatment of detected tumours that are indolent or slow-growing\(^43\), then those studies collectively show an overdetection rate range at only 1%–10%\(^44-54\). This is true however—a large proviso—only if we accept that, for reliable estimation, the length of follow-up required to assure no significant overestimation of overdiagnosis and to adjust for the potential bias from residual detection lead-time effects is 25 years or more of follow-up, as elegantly demonstrated by Stephen Duffy and Dharmishta Parmar\(^55\) using a well-motivated exponential sojourn time model. Such a model helps to account for the wide variation in estimates of overdiagnosis\(^56-61\), but such follow-up is no easy requirement to meet.

For women 50 years of age or older, the first comprehensive **Euroscreen/EUNICE** review of breast screening programs (2 million women in 18 countries)\(^62\) concluded that the chance of saving a woman’s life by population-based screening is greater than that of overdiagnosis, with the combined estimate of overdiagnosis or overdetection solely from studies correctly adjusted for lead-time and underlying trend being 6.5%: for every 1000 women screened biennially, 7–9 lives are saved and 4 women are overdiagnosed. Thus, for every 1 nca overdiagnosed or overdetected, 2 lives would be saved\(^62\). Although those estimates are more favourable than many others cited, their strength is reliance solely on women attendant at, not just invited to, screening, and on eligible studies being restricted to those explicitly accounting for lead-time and underlying trends of increasing bca incidence, leading Robert Smith of the American Cancer Society to conclude, “The strong evidence of benefit associated with exposure to modern mammography screening suggests that it is time to move beyond the randomized controlled trial estimates of benefit and consider policy decisions on the basis of benefits and harms estimated from the evaluation of current screening programs”\(^63\).

And we must also acknowledge in this context the real harms of underdiagnosis, not least of which is a progressive decline in survival for each omitted annual mammography screening: thus, a recent study\(^64\) found that women who had missed any of their previous 5 annual screenings incurred more than a doubling of risk (specifically, an increase by a factor of \(2.3\)) for all-cause mortality compared with subjects having no missed screenings, the \(\text{HR}\) becoming statistically significant at even just 2 annual missed exams, arguing against a biennial schedule.

**REAL-WORLD MAMMOGRAPHY**

The “Women Vote”

In terms of women affected in the controversial 40s age group, 89% want yearly mammograms in their 40s\(^65\), and only 38% believe that false-positive results should be considered in mammography decisions\(^66\), being perceived by more than 90% as an acceptable consequence of screening\(^67\). Women expressed an overwhelming preference to err on the side of caution in preferring the risk of overtreatment to the risk of undertreatment\(^68\). Also note that public acceptance of false-positives remains high, 63% judging as reasonable 500 false-positives to save 1 life, and 37% even judging 10,000 or more to be acceptable\(^69\). There is a disconnect suggesting that in “real-world mammography” the issue of the harms of overdiagnosis and false-positives are seen very differently by researchers and clinicians than by the screening-eligible women themselves. Moreover women’s responses to the issue of overdiagnosis when reasonably understood was in part dependent on estimates of magnitude: only the 50% estimate caused substantial concern; in contrast, the 1%–10% and 30% estimates were seen as more acceptable levels of risk\(^69\). As to the professionals, it has been found that a large proportion of primary care professionals “have neither the capacity nor the training to make a major contribution to supporting informed choice about cancer screening for their patients”\(^70\).

Indeed, there is a well-documented public enthusiasm about cancer screening, especially but not only in the United States. So, although 38% of respondents in one survey had experienced at least one false-positive screening test perceived as “very scary,” 98% were glad of screening, indicating a clear preference for knowing about the presence of a cancer regardless of its implications, and 56% had a testing preference even for highly indolent malignancy\(^71\), perhaps as commentator Lisa Rosenbaum has suggested\(^72\) because of a preference to avoid regret, rather than anxiety—and hence, I would add, undertreatment rather than overtreatment. Even as of 2011, a majority of screening-eligible women did not intend to comply with uspstr guidelines, seeing screening as obligatory\(^73\). This intent in turn intersects with the issue of screening persistence, the consistency of screening attendance. Thus, one study\(^74\) determined that, despite the fact that receiving a false positive mammography screen generated significant worry among 60% of subjects, 70% nonetheless maintained that mammography screening was necessary despite any worry incurred, cross-confirmed in another study\(^75\) evaluating the association between screening persistence and stage at bca diagnosis among elderly women, which found that, compared with women non-persistent with mammography screening, women who were persistent (measured as having at least 3 screening mammograms in the 5 years before bca diagnosis) were significantly more likely to be diagnosed at earlier stages of bca. But that same “screening enthusiasm” could place the public at risk of overtreatment. One example would be screening mammography in women with a limited life expectancy: among women with a life expectancy of less than 7 years, more than one third to one half received mammographic screening, as did 22.2% of women with estimated life expectancy of less than 5 years.
expectancy of only less than 4 years\textsuperscript{76}. These findings suggest overuse, aggravated by many authorities, such as the American Cancer Society\textsuperscript{77}, failing to incorporate projected lifespan into screening guidelines.

**Communicating the Harms**

Mammography guidelines (uspstr) yielded less understanding (6.2\%) and more confusion (30.0\%), greatest among women in their 40s, with considerable difficulty appreciating overdetection\textsuperscript{78} and its difference from false-positives and overtreatment\textsuperscript{9,86}, suggesting the need to develop and test more effective communication strategies\textsuperscript{81}, a formidable challenge aggravated by many seeking information from the “unwashed” Internet. Readability of online patient education materials on mammographic screening was poor, with high scores on “gobbledygook” measures (readability indices)\textsuperscript{82,83}. In an arena confusing even to many professionals, we are clearly failing our audience in often achieving patient–professional discussions and online patient education materials of only minimal comprehensibility. Fortunately, patient decision aids based on validated measures of the primary outcome (informed choice) can improve informed values-based choices, patient–practitioner communication, and realistic perception of outcomes while reducing decisional conflict\textsuperscript{84–86}. One such is the online Breast Cancer Screening Decision Aid (from the Public Health Agency of Canada)\textsuperscript{87–89}, with its quantitative estimates of major risks and benefits of screening; another is the BreastScreen Australia decision aid\textsuperscript{90}, the first mammography screening decision aid validated in a randomized controlled trial\textsuperscript{91}.

**Guideline Adherence and Quality**

After the 2009 uspstr guidelines (screening initiation at age 50 or older and biennial screening to age 74), gynecologists and internal or family medicine professionals still heavily recommend screening beginning at age 40 and annual screening\textsuperscript{92,93}. Nor did the uspstr guidelines usher in a significant reduction in mammography rates, which even increased slightly among women in their 40s\textsuperscript{93}. Screening behavior among younger women post-uspstr has been largely insensitive to the new guidelines\textsuperscript{94,95}, and in Canada, two thirds of female physicians less than 45 years of age had undergone mammography, suggesting low compliance with the Canadian Task Force on Preventive Health Care guidelines\textsuperscript{96}. In addition, guideline quality can be highly variable. Of 11 guidelines for mammography screening (ages 40–49) undergoing critical appraisal by the AGREE methodology, with underlying evidence review assessed by AMSTAR, 5 evidence reviews were rated as poor quality, and only 3 of good quality. Of the guidelines themselves, only 2 were strongly recommended\textsuperscript{97}, risking potential confusion among professionals, the public, and policymakers\textsuperscript{98}.

**Informed Choice?**

As per the decisions survey\textsuperscript{99} and other studies\textsuperscript{100,101}, cancer screening decisions by patients in consultation with health care providers consistently failed to meet criteria for being informed, with patients overestimating both the risks for being diagnosed with and dying from a specific cancer and with more than 90\% of conversations addressing the pros but only 19\% the cons of screening, aggravated by patients’ low numeracy in applying risk reduction information\textsuperscript{102–104}. Furthermore, scientific articles tend to emphasize the major benefits of screening over the harms\textsuperscript{95–107}, with results for overdiagnosis not quoted in 87\% of reviewed articles. Nor was harm often quantified even in randomized screening trials\textsuperscript{107}, with false-positives quantified in just 4\% and overdiagnosis in 7\%, suggesting that patients and even health care providers are unlikely to be making well-informed choices about cancer screening. It is worthwhile to remember that patients and professionals alike do not have intrinsically fine numeracy on these issues: Is screening 1000 or 2000 patients to save 1 death from bc acceptable, or a devil’s bargain? What of 700? What of 200? Individual thresholds show wide variance in such cases of fuzzy judgments, leading to indeterminate disagreements. And as noted in my editorial in this issue, is 1904 as the number needed to invite one person to be screened and 1904 as the number needed to save one life from breast cancer, two thirds of female physicians giving up the struggle of making a decision based on the evidence and in discriminating genuine from only presentation-level disagreements. In certain non-trivial cases, there may be less controversy and more concordance in the debate than the disputes as they play out suggest.

**The Influence of Author Specialties**

A review of author specialties in screening guideline development confirmed the influence of the intellectual and professional interests of the authors\textsuperscript{108,109}: no guideline not recommending routine screening had a radiologist member, suggesting that, compared with generalists, specialists deriving income from screening (radiologists) or treatment (medical oncologists) might tend to recommend routine screening and harbour more strongly pro-screening predilections, although deliberate and conscious bias cannot be concluded from these data. The finding does argue for a broader array of physician specialties and for inclusion of nonphysician health care providers and nonclinical scientists to secure greater balance and objectivity.

**BEYOND THE MAMMOGRAPHY DEBATE**

**New Screening Modalities: Abbreviated MRI Technologies**

We must finally move past the cnbss trial (and the uspstr guidelines) as pivotal to all arguments in the debate [it is not, as sometimes extravagantly claimed, the “worst clinical trial ever done” (nor is it the best—there are better, and there are many worse)]—after all, here as elsewhere we must be concerned with the weight or balance of what the systematically reviewed, methodologically assessed and critically appraised evidence aggregated to date determines, and not with the results of one or a few studies—
and focus on more constructive efforts to improve bca screening. So, despite some diminishing limitations as to the detectability of ductal carcinoma in situ (dcis), presenting as calcifications (but with 98% detectability for high grade dcis), and lobular carcinoma in situ, which exhibits limited neovascularity, inflammation, and the molecular features of the tumour, many of these correlating with proliferative—possibly metastatic—potential, and with a sensitivity or negative predictive value of 98.9%–100% using the streamlined ultrafast (first postcontrast subtracted T1-weighted image) protocol. Based more on tumour biology and functionality, that approach captures invasive cancers (associated with neovascularity) and dcis beyond the old paradigm of anatomy-based screening. But despite residual issues (such as price and availability), the new MRI technologies might yet emerge as the best breast screening tests deployable in the longer term, subject to confirmation in prospective clinical trials.

Although MRI is traditionally deployed as a second-line imaging method for resolution of equivocal mammography diagnostics and in the setting of elevated familial risk, comparative trials have decisively shown significantly higher diagnostic sensitivity and cancer yield for the technique, with reduced interval cancer rates compared with those for screening ultrasonography and mammography. Interest has in part been dampened by cost considerations (including those from false-positive diagnoses) largely stemming from intensive, lengthy acquisition and reading protocols; however, costs have been dramatically reduced by abbreviated breast MRI (ABMRI) without compromise to diagnostic accuracy and cancer yield. Such a streamlined protocol could potentially effect substantial MRI cost reductions and allow for batch MRI screening, with emergence as the standard for breast screening obviating many of the residual debating points surrounding mammography. Concerns about discarding all potentially valuable dynamic information are addressed by protocols such as twist (time-resolved angiography with stochastic trajectories), which preserve accuracy while sustaining short image acquisition times (first-ever 1.02 s). Thus these streamlined MRI protocols are well poised to overcome the key limitations of MRI for screening.

New Screening Modalities: Digital Breast Tomosynthesis

In the interim, the three-dimensional reconstruction technology of digital breast tomosynthesis was approved by Health Canada (in 2009, but is limited in availability) and by the U.S. Food and Drug Administration [but only as an adjunct to two-dimensional (2D) mammographic imaging]. One appeal is the reduction of false-positives (approximately 15%) and increased cancer detection rates (approximately 28%; for invasive cancers, 40% or more, together with a one-third reduction in recall rates and consequent false-positive biopsies, especially in younger women and women with dense breasts in whom mammography is limited, and with no negative effect on sensitivity, which could potentially tip the risk:benefit ratio toward tomosynthesis, and although compared with mammography alone, tomosynthesis roughly doubles total radiation exposure, this can be obviated using the synthesized 2D images created from the three-dimensional tomosynthesis dataset, dramatically reducing the radiation (43% or more) to that of a standard mammogram (eliminating separate 2D exposures), with superior lesion visualization for microcalcification cluster detectability, and having the potential to render conventional 2D mammography obsolete.

A comparison of ABMRI with tomosynthesis shows higher but diminishing costs of ABMRI; an absence, with ABMRI, of radiation exposure and its associated potential harms (secondary cancers, cardiotoxicity); and far more aggressive reduction in recall rates under tomosynthesis, with MRI-based approaches still sustaining an approximately 7%–14% recall rate.

Taming Costs

As of 2014, costs for 30-minute MRI sessions run widely between $277 and $965, but average approximately $500 ($423) and closer to $400 under U.S. Medicare, and are as low as approximately $200 in highly competitive markets (New York City). In contrast, a digital mammogram currently averages $115–$135 in the United States; however, ABMRI technologies might effectively save some $300 per woman scanned, bringing costs down to a competitive $200 or less with further economies. In contrast, tomosynthesis already runs relatively affordably at approximately $192 for combination digital mammography and tomosynthesis under new 2015 U.S. Medicare Rules.

BEYOND DEBATE

Collectively, the considerations set out here—of critically appraised central themes, of mammography in the real world, and of new screening modalities that could, in the nearer future, obviate many of the concerns and conflicts in the current mammography debate—argue for a transformation of our views of and reactions to the dispute to a more nuanced, constructive, and forward-looking perspective that will better serve the interests of the ultimate stakeholders: screening-eligible women and their health professional advisors. And be forewarned that we will need such a transformation, given that the revised draft guidelines from the uspsrt, just released 21 April 2015, reaffirm the original controversial 2009 position of recommended commencement of mammographic screening for asymptomatic average-risk women at age 50, under a biennial schedule through to age 75, but not thereafter (despite the American College of Radiology and the American Cancer Society setting no upper age limit for screening). And although open to public comment until 18 May 2015, it is not anticipated,
given clear signals from the panel, that the final guidelines will change appreciably. Unsurprisingly, the revisions have already reignited intense and passionate debate, almost instantly countered and criticized by Drs. Barbara Monsées (American College of Radiology), Paula Gordon (University of British Columbia), Daniel Kopans (Massachusetts General Hospital), and Richard Wender (American Cancer Society), among numerous others, mirroring two slightly earlier spirited debates, including some of the same principals, one at the American Roentgen Ray Society annual meeting in April 2015, with Drs. Anthony Miller and Cornelia Baines (both University of Toronto) defending the cNMS guidelines in essential agreement with those of the uSPSr, the other being the veritable firestorm reported at the annual European Society of Radiology/European Congress of Radiology meeting in March 2015.

But we are, as I hope this modest contribution suggests, becoming wiser in this debate, achieving insights and dissecting arguments at deeper levels, and uncovering greater confluence that represents and enables true progress. As the eminent criminologist Freda Schaffer Adler famously noted:145

It is not only by the questions we have answered that progress may be measured, but also by those we are still asking. The passionate controversies of one era are viewed as sterile preoccupations by another, for knowledge alters what we seek as well as what we find.

CONFLICT OF INTEREST DISCLOSURES
I have read and understood Current Oncology’s policy on disclosing conflicts of interest, and I declare that I have none.

AUTHOR AFFILIATIONS
* No Surrender Breast Cancer Foundation, Locust Valley, NY, U.S.A.

REFERENCES


BEYOND THE MAMMOGRAPHY DEBATE: A MODERATE PERSPECTIVE, Kaniklidis

108. Saitz R. “We do not see the lens through which we look”: screening mammography evidence and non-financial conflicts of interest. Evid Based Med 2013;18:81–2.