

10:20-10:50

Establishing Guidelines for Breast Cancer Screening Programs

Director of Cancer Screening, American Cancer Society, Atlanta, GA 30303, U.S.A. (robert.smith@cancer.org)

Robert A. Smith

Introduction

Before the availability of breast imaging, the control of breast cancer relied entirely on treating symptomatic breast cancer. Today, breast cancer control is significantly influenced by the opportunity to diagnose breast cancer at a more favorable stage, due both to increased mammography utilization as well as increased awareness among women of the importance of reporting new symptoms promptly to a clinician (1, 2). Although continued progress in our understanding of the epidemiology of breast cancer suggests the potential for risk reduction through lifestyle modification and chemoprevention (3, 4), there still is no dependable preventive strategy, apart from prophylactic surgery in very high risk women (5), that can measurably reduce the absolute risk of breast cancer or eliminate the need for regular screening. For the foreseeable future, early detection and appropriate treatment will remain the cornerstone of the disease control strategy for breast cancer in average and high risk women. Early detection is increasingly being regarded as representing a continuum, from the detection of occult breast cancer with mammography to earlier diagnosis of symptomatic breast cancer when mammography is not available, or when mammography has failed to detect breast cancer while still asymptomatic. The Breast Health Global Initiative also has advanced the principle that downsizing symptomatic breast cancer is possible in countries that do not have the resources or level of disease burden to justify mammography screening, and this goal is achievable by increasing awareness about the impor-

tance of reporting the earliest symptoms to a health care professional (6, 7). Advancing our knowledge about the best strategies for implementing breast cancer early detection programs is an urgent international priority because global breast cancer incidence rates are increasing.

Designing a Breast Cancer Screening Program

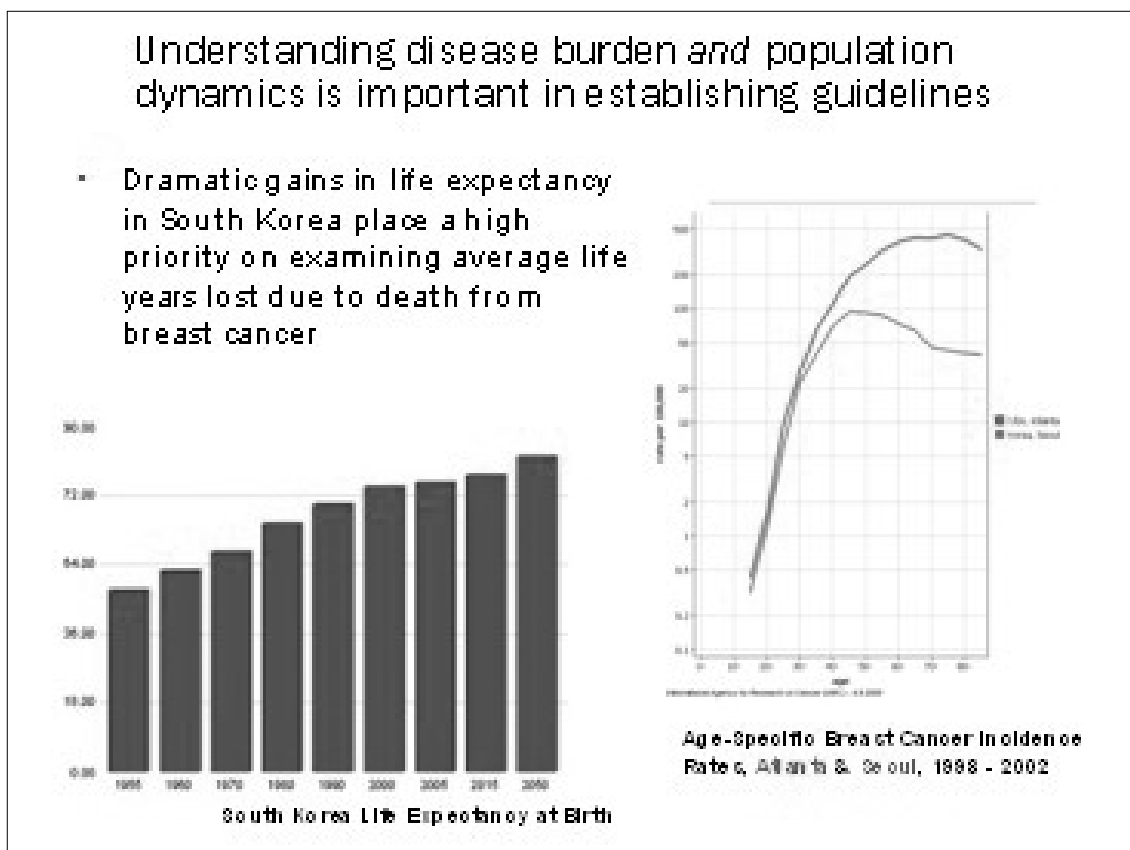
Decisions about implementing breast cancer screening, and the age groups to invite to breast cancer screening, can be made by following the W.H.O. principles and practices for screening for disease developed by Wilson and Junger in 1968 (8). Indeed, the application of Wilson and Junger's principles overall, and for specific age subgroups, oblige us to consider the entirety of the evidence that is relevant to screening policy decisions. The first consideration are measures of disease burden, i.e., is breast cancer an important health problem in the population overall, and if so, in which age groups is it most prevalent? Second, what is known about the period when the disease is detectable, i.e., the *soujourn time*, and specifically, does it differ in different age groups? Third, treatment at an earlier stage should offer better outcomes than if breast cancer were treated at a later stage - this is well established. Fourth, the tests that will be used for breast cancer screening should achieve acceptable levels of sensitivity and specificity, i.e., the test should have good accuracy to detect breast cancer when it is present, and also should correctly identify as normal the large majority of women without breast cancer. We should also ask, is the performance of these tests

similar or dissimilar in different age groups? If so, are the differences so great that the value of screening in some age groups is questionable? The importance of accuracy is self-evident, but the comparative costs and consequence of a false-positive or false-negative test result, may also be a consideration when determining thresholds of acceptable performance. Moreover, the complements of sensitivity and specificity (i.e., false-negatives and false-positives) can provide insights not only into what is being achieved in a screening program but also what potentially could be achieved in the screening program with improvements in accuracy or program design. The importance of on-going evaluation of all program elements is essential. Finally, the test(s) must be affordable, and acceptable to targeted individuals. While each of these criterion is important, and all are important collectively, there are no *established* thresholds for these criteria. Some are fixed, such as the age-specific disease burden, while others are amenable to improvements in cost-effectiveness, such as a the modifying the screening interval and improving interpretative performance.

In a country making a decision about which groups to invite to screening, the relevant decision factors include: available resources and capacity (mammography units, personnel, diagnostic and treatment services, etc.); mammography performance data (with sensitivity analyses); the pattern of age-specific incidence and mortality; and the age-specific average longevity. These data provide planners with the ability to estimate the expected cancer detection rate, the rate of false negatives and false positives, the marginal cost per year of life saved, and the expected reduction in the rate of premature mortality. Thus, planners might invite a certain age-range to screening, and expand that age range over time, or also allow another age range to self-select or be referred to screening. Initially, inviting a higher risk group to screening, such as women with a family history, is also a reasonable method for launching a population-based screening program, but planners should not be deceived in believing that this strategy will identify most women with breast cancer. Apart from women who carry mutations of known penetrance on breast cancer susceptibility genes, most risk factors for breast cancer account for only a small percentage of the attributable risk (9). For population-based screening, Shapiro and colleagues showed great insight in examining inci-

dence-based mortality and years of potential life lost (YPLL) in setting the study population for the HIP study (10), and countries planning the implementation of mammography screening programs should consider including this information in their planning (see below). Planners also should step aside from the cacophany of extreme opinions in the mammography debate and focus on the modern evidence related to age-specific accuracy and benefits of screening as a foundation for local program design. Further, each country should examine their own data carefully, and while they can derive some benefit from previous evidence reviews, guidelines, and the experience of established breast cancer programs, local conditions may warrant different strategies for the control of breast cancer.

As an example, consider the age-specific incidence of breast cancer in Atlanta, Georgia, in the U.S., and Seoul, South Korea (see below). Age-specific guidelines for when to begin mammography screening in the U.S. mostly endorse the age of 40, although a few groups endorse beginning screening at age 50 based on the lower observed benefit from the RCTs, and concerns about the marginal cost per year of life saved (MCYLS). In a comparative effectiveness analysis, women ages 50 years and older in the U.S. have a much lower MCYLS due to the higher incidence rate, higher positive predictive value, and good expected longevity. This more favorable MCYLS continues until some point when expected longevity begins to decline and the MCYLS to increase. In Western countries, which have progressively higher incidence rates from the premenopausal period through the post menopausal period, the MCYLS is comparatively higher in younger women when incidence rates are lower, and also in older women when incidence rates are high, but there are fewer years of life left to save. In Asian and Latin American countries, for example, the age-incidence curve is similar to Western countries in the premenopausal period, but does not show the continued steep increase in the post menopausal period. In the presence of similar incidence and similar performance of screening, the MCYLS will be more similar, if not more favorable, in younger compared with older women due to the greater number of years of life to save. As shown below, the age-incidence curve in South Korea reveals that the incidence rate is more similar than different in women ages 40–70,



only ranging between 50 and 100 cases per 100,000 women per year. In fact, the breast cancer incidence rate at age 40 and 60 are quite similar. Given the dramatic gains in life expectancy in South Korea since the mid-20th century, the potential to reduce the rate of premature mortality in younger women diagnosed with breast cancer is similar in Seoul, South Korea when compared with breast cancer rates in Atlanta, GA, which are similar to other U.S. cities.

Age to Begin Screening-The Anatomy of a Long Standing Debate

While it would seem relatively straightforward to determine an age to begin screening, the age to start screening has been controversial since the earliest evaluations of the RCT's. The HIP study was the first RCT of breast cancer screening and it included women ages 40-64. Most other RCTs simply followed their design. In the HIP trial, Shapiro and colleagues chose this age range because they observed that it included about 75% of the years of potential life lost (YPLL) due to a death from breast cancer before age 80. A diagnosis between ages 40-49 accounted for

34% of the YPLL before age 80, and a diagnosis between ages 50-64 accounted for 38% (10). In the U.S., the proportion of deaths in any given year attributable to a diagnosis between ages 40-49 and 50-59 is similar (15-20%). Thus, in the U.S. and Europe, from a disease burden standpoint, including women in their forties in breast cancer screening programs is sensible. However, early evidence related to the efficacy of screening in women under age 50 was not persuasive.

Although the early trials were not designed to have sufficient statistical power to address the question of age-specific benefits, mortality reductions in the group randomized in their forties were smaller and did not achieve statistical significance compared with women ages 50 and older at the time of randomization (11). While 1st generation trials (HIP and 2-County) did not show a statistically significant age-specific benefit for women aged 40-49 (12, 13), and although the results of the Canadian Trial were also negative, they were not widely accepted due to the appearance of randomization anomalies (14, 15). Opponents of screening women under age 50 were inclined to accept the RCT results, and to argue that breast cancer was, in fact, more aggressive, grew faster, and that the

higher prevalence of breast density in women younger than age 50 all contributed to poorer performance of mammography in this age group (16).

In these early years, proponents of screening favored accepting the RCT results for the broad age range of women included in the trial, rather than age-specific outcomes. They argued that there was no biological basis to regard breast cancer in women under age 50 as inherently different compared with breast cancer after age 50 (17). They also noted that the non-significant mortality reductions in the RCTs were still respectable, and with adequate statistical power would have achieved statistical significance (18). It was argued that in addition to low statistical power, the early RCT's showed poorer results in younger women due to a tendency towards single view mammography and 2+ yearly screening intervals (19).

This debate continued while policy makers waited for additional years of follow-up, and for the results of newer RCTs. Further, over time a new explanation for the smaller observed benefit in younger women gained favor. Since mammography was shown to be beneficial to women ages 50+, and since it took considerably longer for differences in mortality rates to be observed in women randomized in their forties, then it was likely that any benefit observed among women randomized in their forties was due only to those women randomized in their forties who were diagnosed with breast cancer after age 50 (20). Because the analysis of age-specific mortality differences in trials is based on age at randomization (not age at diagnosis), at the conclusion of the invitations to screening, the group randomized in their 40s will include a proportion of women diagnosed with breast cancer *after* they have had a 50th birthday. According to this hypothesis, which has been referred to as “age migration” or “age creep,” if a benefit from screening truly begins at age 50, then any observation of benefit for women in their 40s in the RCTs is due to the benefit of diagnosing breast cancer after age 50 among those women randomized in their 40s.

How do we reconcile these different interpretations of the evidence? First, it is important that none of these arguments used to explain the poorer performance in women under age 50 was based on direct observation. Second, it is true that the early RCTs were not designed for age-specific analysis. In a comparison of breast cancer deaths in women under vs.

over age 50, there is considerably lower statistical power in the younger age group due to both lower incidence in the 40–49 group, fewer cases in a 10 year age-span vs. one that spans 20–25 years, and thus, smaller numbers of breast cancer deaths. Third, the evaluation of age-specific benefits in a prospective screening trial results in lead-time bias due to the fact that age at diagnosis is influenced by the intervention (ie, mammography). While the question of age-specific benefits is important, it is not answered easily with age-specific retrospective analysis, but rather with a very large RCT in which a very small age-range is randomized, and within which the invited group can not reach the next age decade of interest during the years of invitation (21). This question can also be directly examined in the evaluation of service screening, which we'll describe later (22). While the original conjecture about age-migration was based on modeling (20), this hypothesis has been examined by evaluating actual RCT data in Sweden by age at randomization and age at diagnosis. Not only was there no empirical support for the hypothesis of age-migration, the benefit observed by age at diagnosis was greater in the women diagnosed with breast cancer in their forties than among those diagnosed after age 50 (23). No support for this hypothesis was observed in a similar analysis of Gothenburg trial data (24). Despite lack of empirical support for the age-migration hypothesis, empirical findings that refute the hypothesis, and the fact that there are clear methodological reasons why this sort of analysis of RCT data introduces bias, the age-migration hypothesis continues to be a popular and intuitive basis for conjecture as a likely explanation of the observed benefit in younger women with increasing years of follow-up (25, 26).

Continued analysis of RCT data pointed to other accounts for the smaller benefit in younger women in the early trials. Early evidence of significantly elevated interval cancer rates in the younger age groups pointed to the shorter tumor sojourn times in premenopausal women and the need for a shorter (12–18 months) screening interval (27–29). Further, the delay in benefit in women under age 50 vs. women ages 50+ was shown to be a function the interplay between the screening interval and tumor progression in younger women. Since the mean breast cancer sojourn time (i.e., potential lead time) is shorter in women younger than age 50 compared with women older than age 50,

and since the majority of the world's trials screened women aged 40 to 49 at an interval of 24 months, faster tumor progression in women in this age group meant that they were less likely to benefit from mammography than would women aged 50 and older. The consequence of screening all women at the same 24-month interval becomes further evident when cumulative mortality trends are examined by age and tumor histology (30, 31). When cumulative mortality trends are compared for ductal grade 3 tumors, a benefit begins to appear after 5 years among women aged 50 to 74 but is not evident in the follow-up period among women aged 40 to 49. These higher malignancy grade tumors are more aggressive, have a worse prognosis, and, compared with ductal grade 2, medullary, and invasive lobular cancers, account for a higher proportion of breast cancer deaths that occur sooner after diagnosis. In contrast, when comparing trends in the invited versus control group in the trial for ductal, grade 2, lobular, and medullary carcinoma, women aged 40 to 49 and 50 to 74 showed similar trends in cumulative mortality. Furthermore, for both women younger and older than age 50 who were diagnosed with these less aggressive cancers, the mortality benefit does not appear until after 7 to 8 years of follow-up, consistent with the observation that survival is better for these tumor types and the mortality that occurs will occur later. When examining the overall trend in cumulative mortality by age, these same trends are evident; in other words, for the groups invited to screening, a mortality benefit begins to emerge after approximately 5 years for women aged 50 to 74 and after 8 years for women aged 40 to 49.

Support for this observation was also evident in surveillance data (32). When tumors are grouped by size, grade, and nodal involvement, there is little difference in observed survival by age. However, Tabar and colleagues have also observed that the likelihood that a tumor will have progressed to grade 3 at the time of diagnosis, and in particular for tumors > 10–15 mm, is higher in women younger than age 50 than in women aged 50 and older, indicating that one important difference between breast cancer in premenopausal women compared with postmenopausal women is the greater likelihood of tumor dedifferentiation to less favorable histology (33). As a general rule, it is important to diagnose tumors at the earliest opportunity, and it is apparent that this goal is especially important

in premenopausal women. Indeed, the 2nd generation trials (Malmo and Gothenberg) adopted the shorter screening interval based on these early observations of higher interval cancer rates, and observed statistically significant mortality reductions in women randomized in their forties, ranging from 36% to 44% (24, 34), results that were similar to those observed in women ages 50+. The timing of the benefit also was similar to women aged 50+. A third RCT, the UK Age Trial, showed a 24% reduction in deaths after adjustment for non-compliance, that although not statistically significant (RR = 0.76, 95% CI 0.51–1.01) was judged to be consistent with findings from other RCTs (26).

Individual Trial Results vs. Meta-Analyses

At the time the analysis of Swedish RCT data had accounted for the smaller benefit in women under age 50 and the Gothenberg and Malmo RCTs had shown favorable benefits mostly due to shorter screening intervals in younger women, there also had been sufficient long-term follow-up for meta-analyses to demonstrate a statistically significant mortality reduction in women aged 40–49 (35). Since that time, meta-analysis has been the preferred method for estimating age-specific benefits from the RCTs rather than identifying the best RCTs to estimate age-specific benefit. This is understandable, but unfortunate because the overall estimates are influenced not only by trials that showed poor and good results, but by the greater statistical power of the Canadian trial, which did not show a benefit. It should be understood that the underlying purpose of meta-analysis is to reconcile the lack of statistical power in multiple small studies with similar results, and not to reconcile widely disparate results from dissimilar trials. The most recent meta-analyses were conducted by Humphrey et al as part of the evidence report accompanying the most recent update of the U.S. Preventive Services Task Force breast cancer screening guidelines. With follow-up ranging from 11–19 years and excluding the Edinburgh trial, they observed a summary relative risk of 0.85 (95% CI, 0.73–0.99) for women ages 40–49, and for women ages 50+, the summary relative risk was 0.78 (95% CI, 0.70–0.87) years. When only the five Swedish trials are included in a meta-analysis, which is a more homogeneous group of RCTs, there

was a 29% mortality reduction for women ages 40–49 at randomization (RR=0.71, 95% CI 0.57–0.89) and a 26% mortality reduction for women ages 50–74 (RR=0.74, 95% CI 0.64–0.86) (36). In summary, there are five important observations about these meta-analyses to evaluate age-specific results:

- Meta-analyses commonly compared women 40–49 with women 50–69. Age 50 provides a somewhat arbitrary benchmark for the age of menopause, although there is no evidence that age 50 is an important benchmark. An age comparison of 40–54 vs. 55–70 would be more informative, as would comparisons in 5 or 10 year age groups;
- As noted above, as was the case with individual trials, benefits (lower mortality) for women aged 50 and older appear early in the follow-up period, whereas these benefits occur later for women aged 40 to 49. This pattern has been explained due to the interplay between tumor biology in younger women and screening intervals in the early trials that were too long (2+ years) (31). This different pattern is not evident in later trials or modern service screening;
- With accumulating years of follow-up in the 40 to 49 group, the relative risk of breast cancer mortality steadily improves in the group invited to screening;
- Depending on inclusion and exclusion criteria, and the duration of the follow-up period, meta-analysis results have shown considerable variation. As expected, the magnitude of the benefit is less for all studies combined than for some individual studies.
- The primary purpose of meta-analysis is to overcome the limitations of small sample sizes, and this was the initial purpose for combining the RCT data for women randomized in their forties, especially with longer follow-up. However, this early approach to reconcile the limitations of the trials for age-specific analyses has led to a tendency to presume that the true benefit of mammography is best estimated by combining all the RCT results, even when the statistical power of some more recent individual trials was adequate.

While we can accept meta-analysis results as a good measure of the efficacy of breast cancer screening,

Relative Risk of Breast Cancer Mortality and of Incidence of Node Positive Disease in the 8 Mammography Randomized Controlled Trials

RCT	RR breast cancer mortality	RR node positive diagnosis
HIP	0.78	0.85
Malmo	0.78	0.83
Two-county	0.68	0.73
Edinburgh	0.78	0.81
Stockholm	0.90	0.82
NBSS1	0.97	1.20
NBSS2	1.02	1.09
Gothenburg	0.79	0.80

should we accept them as a good measure of the effectiveness of screening? The answer is no, since they represent a weighted average of the results of older and newer technology and older and newer understanding of the interplay between that technology and the natural history of the disease. Further, individual RCT results and the results of meta-analyses represent the effectiveness of an invitation to screening, and not screening *per se* due to intention-to-treat methodology. This point is especially important when RCT data are evaluated to examine the relative risk (RR) of being diagnosed with an advanced breast cancer, which has been shown to be a very good surrogate for the RR of dying from breast cancer (see below). Indeed, this is the fundamental purpose of breast cancer screening, i.e., detecting breast cancer earlier in its natural history when treatment is more effective. Overall, and within age-specific categories, the RCTs show close parallels between the relative risk of being diagnosed with an advanced breast cancer and the relative risk of dying from breast cancer (30, 37). Those RCTs that succeeded in reducing the incidence rate of advanced disease in women in their 40's and women 50+ also showed significant mortality reductions. Thus, while it is not possible to entirely reconcile why the different RCTs showed a range of outcomes specific to age at randomization, it is possible to directly examine the link between success at reducing the incidence rate of advanced disease and the resulting reduction in breast cancer mortality.

The Evaluation of Service Screening

If the RCTs are limited in their ability to answer key,

modern questions about the effectiveness of mammography, what about observational data? Since the efficacy of breast cancer screening has been proven, there is little true potential for further prospective RCTs in which one group of women are invited to screening and another group is offered usual care. Such a study would be unethical, and also likely would have few willing participants. At this time, it is important to recognize that careful evaluation of observational data can be quite informative. Of course, prospective randomized studies to evaluate comparative screening protocols is possible provided that each group is invited to screening, and matched-pair designs also provide an opportunity to avoid ethical challenges. In settings where a particular protocol has been in place for some time, the evaluation of service screening can focus on both mortality reductions in the entire population, which somewhat simulates an intention-to-treat analysis, but also mortality reductions in women who actually attend community screening programs, which is the more relevant estimate of benefit and age-specific benefit, especially since the measurement of benefit can be adjusted for known biases based on what was learned in the RCTs. The evaluation of service screening also can be applied to estimate differences in mortality over time due to screening compared with improvements in therapy and increased awareness (38), although establishing the relative contribution of screening and non-screening factors is complex and can be only indirectly estimated. Nevertheless, the evaluation of service screening programs has demonstrated that mortality reductions observed in the RCTs are often exceeded in programs with high quality and high compliance. For example, an analysis of service screening in the two counties where the Swedish Two County Trial took place (Dalarna and Ostergotland) compared the mortality rates from breast cancer in the 20 years following the introduction of screening with the corresponding rates in the 20 years preceding the introduction (i.e., pre-screening and screening epochs) (39). The investigators avoided misclassification of exposure to screening by limiting the pre- and post comparisons to deaths occurring only from tumors diagnosed within the pre- and post-screening epochs. Methodological adjustments for self-selection bias and changes in the incidence of breast cancer over the period of study were applied. Among women aged 40–69 exposed to

screening, there was a statistically significant 44% (RR=0.56; 95% CI 0.49–0.64) reduction in mortality in the screening epoch compared to the pre-screening epoch, and a 16% (RR=0.84, 95% CI, 0.71–0.99) mortality reduction that was independent of screening (i.e., in women not exposed to screening). Among women aged 40–49 at the time of diagnosis, the investigators observed a statistically significant 48% (RR=0.52, 95% CI 0.4–0.67) breast cancer mortality reduction among women exposed to screening, compared with a 44% mortality reduction in women 50+. The similarity in the mortality reduction highlights the importance of policy that supports a shorter screening interval for premenopausal women compared with post-menopausal women. More extensive evaluations of service screening data have been conducted by the Swedish Organized Service Screening Evaluation Group (SOSSEG), which is a team of researchers representing local Swedish investigators and officials, but also researchers from the U.S., U.K., and Taiwan that have evaluated service screening in 13 population areas representing 10,177,113 person years of observation and 23,092 breast cancers (40–42). After adjustment for selection bias, there was 43% mortality reduction (RR, 0.57; 95% CI, 0.53–0.62) among the women exposed to screening compared with all women in the prescreening epoch. In this same series, among women aged 40–49 exposed to screening, there was a significant 45% reduction in the risk of being diagnosed with a tumor greater than 2 cm in size compared with tumor sizes in the prescreening epoch (RR = 0.55; 95% CI, 0.46–0.66), and among women 50–69, a 33% reduction in being diagnosed with a tumor greater than 2 cm in size in the population exposed to screening compared with the population in the prescreening epoch (RR=0.67; 95% CI 0.62–0.72) (42). This observation is similar to that observed in the RCTs, i.e., the trend in the incidence rate of advanced disease in all women and in age-specific subgroups is a leading and accurate predictor of the mortality reduction that eventually will be seen.

While these are long term evaluations, shorter periods of study are not only practical, but easier to carry out and potentially very influential for changing behaviors and improving quality. One example of a short-term evaluations is failure analysis, which is a common methodology across many endeavors that seeks to explain the cause of undesirable results for

the purpose of preventing their occurrence. If an early detection program is in place, and there is a persistently high rate of presentation with advanced disease, what can we learn about why these cancers are not being detected earlier. Failure analysis can be carried out in almost any setting, large or small, since the simple goal is to understand what could reasonably be described as “avoidable” events. Failure analysis requires identifying consecutive cases of breast cancer in an institution and through interviews and medical chart audits, identifying if there were missed opportunities to detect her cancer earlier, or conversely, whether and why programs were resulting in an improvement in stage at diagnosis at one institution compared with another. Given the solid evidence and technology supporting early breast cancer detection, clinicians and institutions should be interested in “explaining” why a woman was diagnosed with advanced breast cancer, and whether there were opportunities that were missed that might have resulted in a more favorable stage at diagnosis.

Efficacy vs. Effectiveness

Once efficacy was established by individual trials and meta-analyses, and now persuasively confirmed by the evaluation of service screening, the debate evolved to be more focused on both comparative efficacy, where the benefits of mammography are judged to be less in younger women, and various measures of cost-effectiveness, i.e., the marginal per year of life saved, the number needed to screen to save one life, mammography performance (comparing false positive and false negative rates), and other measures (43). Both proponents and opponents can cite manuscripts in reputable journals to bolster their position related to age-specific benefits, and commonly will neglect good evidence that conflicts with their opinion. Indeed, Kopans has shown that journals themselves appear to have an established a “position” as measured by the proportion of supporting or non-supporting manuscripts published (44). However, it is important to consider the totality of the evidence and the historical accumulation of evidence. As Andy Rooney commented, “Ignoring the facts does not change the facts.”

One of the most important facts is that breast cancer screening gradually improves in effectiveness with

increasing age. Comparing women ages 40–49 with women ages 50+ obscures this transition. The table below from the U.S. National Cancer Institute’s Breast Cancer Surveillance Consortium is instructive. As can be seen from the evaluation of over 3 million mammography examinations, the improvement in the effectiveness of screening improves with increasing age. There are no large homogenous age groups, with sensitivity and specificity being quite respectable, but improving incrementally in each age-subgroup with increasing age. Thus, the data do not support the conclusion that mammography in women under age 50 is error prone but highly accurate after age 50 (45). In fact, it is clear that the improvement in effectiveness in mammography is incremental with increasing age. The argument that women under age 50 have levels of breast density that make screening extremely difficult is not supported by these performance data... while density levels are higher in younger women and affect accuracy, differences in sensitivity, specificity, and PPV between women in five and 10 year age groups are small.

Performance Measures for 3,603,832 Screening Mammography Examinations-1 from 1996 to 2006 by Age

	Sensitivity ^a	Specificity ^a	PPV ^a	Recall ^a
Total	80.2%	91.4%	4.3%	8.9%
Age 40-49	70.8%	89.8%	1.5%	10.3%
Age 45-49	74.3%	89.8%	2.3%	10.3%
Age 50-54	78.4%	90.9%	3.3%	9.2%
Age 55-59	81.6%	91.5%	4.6%	8.8%
Age 60-64	80.0%	91.9%	5.4%	8.4%
Age 65-69	82.5%	92.4%	6.3%	8.0%
Age 70-74	82.9%	93.1%	7.9%	7.3%
Age 75-89	84.5%	93.6%	9.8%	6.9%

Conclusion

Establishing guidelines for a breast cancer screening program should be a strictly evidence-based process. However, it is important to follow the principles of Wilson and Junger so that the recommendations and program are designed to meet the needs of women in the community where the program will be established.

Too often planners dogmatically embrace what has been designed in another country without careful consideration of the historical and local factors that led to those decisions, which may also have not been carefully considered in the beginning or over time. Guidelines should not be static, but rather should be periodically reconsidered in light of new evidence on the epidemiology of the disease, detection technology, and advances in therapy. The experts on a guidelines panel should be free of outside influences and conflicts of interest, and they should have the expertise in their field. Finally, on-going evaluation of breast cancer screening programs is essential to insure the guidelines are being followed, and that the guidance was correct. Following a careful, evidence-based process, combined with regular evaluation and feedback on program performance, will insure the most cost-effective delivery of services, and the confidence of the target population and health care workers.

Acknowledgements: For many of these ideas, collaborations, and for many years of friendship: Professors Stephen Duffy, Tony Chen, Amy Yen, Won Chul Lee, and Laszlo Tabar

References

1. Stockton D, Davies T, Day N, McCann J. Retrospective study of reasons for improved survival in patients with breast cancer in east Anglia: earlier diagnosis or better treatment [published erratum appears in *BMJ* 1997 Mar 8;314(7082):721] [see comments]. *Bmj* 1997;314:472-5.
2. Chu KC, Tarone RE, Kessler LG, et al. Recent trends in U.S. breast cancer incidence, survival, and mortality rates [see comments]. *J Natl Cancer Inst* 1996;88:1571-9.
3. Willett LR, August DA, Carson JL. Breast cancer prevention. New options for patients and clinicians. *N J Med* 2002;99:27-34.
4. Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res* 2004;6:213-8.
5. Fatouros M, Baltoyianis G, Roukos DH. The predominant role of surgery in the prevention and new trends in the surgical treatment of women with BRCA1/2 mutations. *Ann Surg Oncol* 2008;15:21-33.
6. Anderson BO, Yip CH, Smith RA, et al. Guideline implementation for breast healthcare in low-income and middle-income countries. *Cancer* 2008;113:2221-43.
7. Yip CH, Smith RA, Anderson BA, et al. Guideline implementation for breast healthcare in low- and middle-income countries. *Cancer* 2008;113:2244-56.
8. Wilson JMG, Junger G. Principles and practice of screening for disease. Geneva, Switzerland: World Health Organization; 1968.
9. Smith RA. Risk-based screening for breast cancer: Is there a practical strategy? *Seminars in Breast Disease* 1999;2:280-91.
10. Shapiro S, Venet W, Strax P, Venet L. Periodic Screening for Breast Cancer: The Health Insurance Plan Project and its Sequelae. Baltimore: Johns Hopkins Press; 1988.
11. Fletcher SW, Black W, Harris R, Rimer BK, Shapiro S. Report of the International Workshop on Screening for Breast Cancer [see comments]. *J Natl Cancer Inst* 1993;85:1644-56.
12. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* 1982;69:349-55.
13. Tabar L, Fagerberg CJ, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet* 1985;1:829-32.
14. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years [published erratum appears in *Can Med Assoc J* 1993 Mar 1;148(5):718] [see comments]. *Cmaj* 1992;147:1459-76.
15. Boyd NF. The review of randomization in the Canadian National Breast Screening Study. Is the debate over? *Cmaj* 1997;156:207-9.
16. National Institutes of Health Consensus Development Panel. National Institutes of Health Consensus Development Conference Statement: Breast Cancer Screening for Women Ages 40-49, January 21-23, 1997. National Institutes of Health Consensus Development Panel. *J Natl Cancer Inst* 1997;89:1015-26.
17. Kopans DB. Breast screening in women under 50. *Lancet* 1991;338:447.
18. Feig SA. Estimation of currently attainable benefit from mammographic screening of women aged 40-49 years. *Cancer* 1995;75:2412-9.
19. Tabar L, Duffy SW, Burhenne LW. New Swedish breast cancer detection results for women aged 40-49. *Cancer* 1993;72:1437-48.
20. de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ. Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials [see comments]. *J Natl Cancer Inst* 1995;87:1217-23.
21. Moss S. A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. Trial Steering Group. *J Med Screen* 1999;6:144-8.
22. Gabe R, Duffy SW. Evaluation of service screening mammography in practice: the impact on breast cancer mortality. *Ann Oncol* 2005;16 Suppl 2:ii153-62.
23. Duffy SW, Day NE, Tabar L, Chen HH, Smith TC. Markov models of breast tumor progression: some age-specific results. *J Natl Cancer Inst Monogr* 1997;22:93-7.
24. Bjurstam N, Bjorneld L, Duffy SW, et al. The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39-49 years at randomization [see comments]. *Cancer* 1997;80:2091-9.
25. Sox H. Screening mammography for younger women: back to basics. *Ann Intern Med* 2002;137:361-2.
26. Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L. Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet* 2006;368:2053-60.
27. Tabar L, Fagerberg G, Day NE, Holmberg L. What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. *Br J Cancer* 1987;55:547-51.
28. Chen HH, Duffy Sw, Tabar L, Day NE. Markov chain models for progression of breast cancer, part 1: tumour attributes and the pre-clinical screen-detectable phase. *J Epidemiol Biostat* 1997;2:9-23.

29. Chen HH, Duffy Sw, Tabar L, Day NE. Markov chain models for progression of breast cancer, part 2: prediction of outcomes for different screening regimes. *J Epidemiol Biostat* 1997;2:25-35.
30. Organizing Committee and Collaborators. Breast cancer screening with mammography in women aged 40-49 Years. Report of the organizing oommittee and collaborators, Falun meeting, Falun, Sweden (21and 22 March, 1996). In *J Cancer* 1996;68:693-9.
31. Tabar L, Chen HH, Fagerberg G, Duffy SW, Smith TC. Recent results from the Swedish Two-County Trial: the effects of age, histologic type, and mode of detection on the efficacy of breast cancer screening. *J Natl Cancer Inst Monogr* 1997;22:43-7.
32. Ries L, Henson D, Harras A. Survival from breast cancer according to tumor size and nodal status. *Surgical Oncology Clinics of North America* 1994;3:35-50.
33. Tabar L, Fagerberg G, Chen HH, Duffy SW, Gad A. Tumour development, histology and grade of breast cancers: prognosis and progression. *Int J Cancer* 1996;66:413-9.
34. Andersson I, Janzon L. Reduced breast cancer mortality in women under age 50: updated results from the Malmo Mammographic Screening Program. *J Natl Cancer Inst Monogr* 1997;22:63-7.
35. Smart CR, Hendrick RE, Rutledge JH, 3rd, Smith RA. Benefit of mammography screening in women ages 40 to 49 years. Current evidence from randomized controlled trials [published erratum appears in *Cancer* 1995 Jun 1;75(11):2788]. *Cancer* 1995;75:1619-26.
36. Hendrick RE, Smith RA, Rutledge JH, 3rd, Smart CR. Benefit of screening mammography in women aged 40-49: a new meta-analysis of randomized controlled trials. *J Natl Cancer Inst Monogr* 1997;22:87-92.
37. Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH. The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* 2004;42:793-806, v.
38. Tabar L, Vitak B, Tony HH, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001;91:1724-31.
39. Tabar L, Yen MF, Vitak B, Chen HH, Smith RA, Duffy SW. Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. *Lancet* 2003;361:1405-10.
40. Swedish Organized Service Screening Evaluation Group. Reduction in breast cancer mortality from the organised service screening with mammography: 2. Validation with alternative analytic methods. *Cancer Epidemiol Biomarkers Prev* 2006;15:52-6.
41. Swedish Organized Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev* 2006;15:45-51.
42. Swedish Organized Service Screening Evaluation Group. Effect of mammographic service screening on stage at presentation of breast cancers in Sweden. *Cancer* 2007;109:2205-12.
43. Qaseem A, Snow V, Sherif K, Aronson M, Weiss KB, Owens DK. Screening mammography for women 40 to 49 years of age: a clinical practice guideline from the American College of Physicians. *Ann Intern Med* 2007;146:511-5.
44. Kopans DB. Bias in the medical journals: a commentary. *AJR Am J Roentgenol* 2005;185:176-7.
45. NCI-funded Breast Cancer Surveillance Consortium co-operative agreement (U01CA63740 UC, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, U01CA70040), . Performance Measures for 3,603,832 Screening Mammography Examinations1 from 1996 to 2006 by Age. In: National Cancer Institute; 2008.