Screening for Breast and Prostate Cancers: Moving Toward Transparency

David H. Newman

Manuscript received February 23, 2010; revised April 27, 2010; accepted April 28, 2010.

Correspondence to: David H. Newman, MD, Department of Emergency Medicine, Mount Sinai School of Medicine, 3 East 101st St, New York City, NY 10029 (e-mail: davidnewman44@mac.com).

Despite mortality reductions found in early trials, recent population-based data suggest that breast and prostate cancer screening have not yielded expected benefits. Whereas evidence-based appraisals generally mistrust disease-specific mortality as a primary outcome measure, cancer screening trials have consistently used this endpoint, largely because of the impracticality of studies with enough statistical power to detect all-cause mortality reductions, which would require millions of subjects. The acceptance of disease-specific mortality as a practical surrogate for all-cause mortality may explain the discrepancy between expected and actual impact. Screening may reduce deaths from the target cancer but may increase deaths from other causes, most likely because of overdiagnosis, an increasingly recognized risk of cancer screening. Recognition of the discrepancy between the expected and the actual impact of screening and recognition of overdiagnosis as a source of harm may be critical for understanding and projecting the potential impact of cancer screening programs.

J Natl Cancer Inst 2010;102:1008–1011

In November 2009, the US Preventive Services Task Force changed their routine screening mammography recommendations to exclude women younger than age 50 years (1). One month earlier, a special report in the Journal of the American Medical Association concluded that breast and prostate cancer screening in the United States has led to an increase in diagnoses without a commensurate reduction in invasive disease or cancer mortality (2). Both publications mark an evolution in perceptions of cancer screening, and both focus on disease-specific cancer mortality reductions, an outcome measure that has driven three decades of policy and practice.

Despite being widely accepted as an outcome measure in many screening trials, disease-specific mortality is essentially a surrogate marker for all-cause mortality, a universally accepted endpoint. However, those newly encountering the evidence are often surprised to discover that no trial data have demonstrated a survival benefit with prostate-specific antigen (PSA) or mammography screening.

In 1963, the world’s first mega trial of mammography (>60,000 subjects) began enrollment. At all stages of data analysis, the study found that breast cancer mortality was lower in the mammography group than in the nonscreening group (3). Four more large trials confirmed this finding (4–7) and helped to spearhead rapid growth in screening mammography and advocacy for early detection.

Four of the five large trials that found breast cancer mortality reductions, however, found no overall mortality reduction (8). In addition, three subsequent trials of 140,000 women, which were all considered more rigorous in their methods than the earlier studies (primarily because of more balanced randomization), detected no impact on either all-cause or breast cancer mortality (9–11). In 2001, the Cochrane Collaboration (>470,000 subjects), in the most comprehensive review to date (12), found no detectable all-cause mortality benefit to screening mammography. Similarly, in the early 1990s, enrollment began in two mega trials of PSA screening, one in a European population (13) and the other in the United States (14). In 2009, initial results from those trials demonstrated that prostate cancer mortality was reduced in the European population (13) but not in the United States (14). All-cause mortality, however, was not affected in either trial.

Medical practice, education, and research methods have evolved since the first screening trials began, and the medical community’s approach to scientific literature has moved in step. With the rise of evidence-based medicine, new emphasis has been placed on critical appraisal of research. In particular, patient-oriented research and outcomes have become a point of emphasis (15,16). Whereas reductions in disease-specific mortality were once considered to be robust outcomes, such findings are now felt to be incomplete. Conversely, all-cause mortality is considered unambiguous and patient oriented. In the current evidence climate, a medicine that reduces deaths from stroke but leads to an equal or greater number of coronary or other deaths would not be considered beneficial, even if the drug’s primary mechanism were aimed at stroke prevention.

There has been debate, however, on the issue of disease-specific vs all-cause mortality in the literature on cancer screening trials. Defenders have noted that disease-specific mortality is and always will be only a small fraction of all-cause mortality in cancer screening trials. With such small ratios of disease-specific to all-cause mortality, overall reductions may be unreliable or difficult to detect (17,18). It is argued that large trials, and even systematic
reviews, have had inadequate statistical power to detect an all-cause mortality reduction, a finding that would require millions of subjects (19). The impracticality of studies and numbers this large is a salient point, but the argument highlights two critical facts. First, screening for breast and prostate cancers remains unproven as a method to reduce overall mortality. Whereas disease-specific mortality may be a plausible surrogate for all-cause mortality, reductions in all-cause mortality have not been proven but have only been conjectured on the basis of reductions in disease-specific mortality. Second, inadequate statistical power to detect all-cause mortality reductions in nearly half a million women for mammography and nearly a quarter million men for PSA indicates that if there are all-cause mortality benefits from these modalities, they are extremely small, which belies widespread perceptions of breast and prostate cancer screening.

There is consensus that all-cause mortality is a more valid outcome than disease-specific mortality, in part because it may demonstrate effects that are not anticipated (20–24). All-cause mortality measures have been pivotal in discovering the ineffectiveness and, in some cases, the dangerousness of once-common therapies, such as fibrates for cholesterol reduction (20), antiarrhythmics following myocardial infarction (21,22), and liberal red cell transfusion in the critically ill (23). Measurement of all-cause mortality has also demonstrated the benefit of treatments previously believed harmful, such as beta-blockade for patients with chronic congestive heart failure (24). In each of these cases, the use of plausible surrogate markers led to widespread acceptance of a therapy, but the ultimate findings and their mechanisms were not anticipated. Subsequent trials with adequate statistical power reversed years of conventional wisdom.

All-encompassing endpoints are therefore essential to understanding the impacts of screening, particularly in the context of interventions with potentially complex and multifaceted effects. Thus, when the effect on a surrogate endpoint does not demonstrate parallel impact with the broader endpoint for which it is a surrogate, this should raise a flag. In a 2002 examination of cancer screening trials, investigators noted that all-cause and disease-specific mortality exhibited opposing directions of effect in five of the 12 studies reporting both. In two of the seven trials in which endpoints did coincide, their magnitudes of effect were inconsistent. The authors conclude that “a reduction in disease-specific mortality should not be cited as strong evidence of efficacy when all-cause mortality is the same or higher” (25). This conclusion has direct relevance to prostate and breast cancer screening. All-cause mortality was unchanged or higher for mammography in eight of the nine trials in the Cochrane review (12) and unchanged with PSA screening in both recent trials (13,14).

This discrepancy in mortality effects is broadly ignored in current discussions of cancer screening data. Following the recent publication of two PSA screening trials, articles in the lay press (26), the studies themselves (13,14), and a peer-reviewed editorial (27) all failed to discuss the lack of overall mortality benefit. Instead, discussions were confined to the implications of findings on disease-specific mortality.

There are many factors contributing to this selective reporting, including two of particular note. First, expert opinion may often lead to guidelines or recommendations that reflect a level of certainty disproportionate to supporting evidence (28,29). Prominent scientific groups appear to have accepted disease-specific mortality as an endpoint and therefore strongly support mammographic screening, including the US Preventive Services Task Force (30), the American Medical Association (31), the American College of Radiology (32), and the American Cancer Society (33). Such broad endorsements lend a formidable air of expertise and consensus and may play a role in discouraging discussion on the question of overall efficacy. Second, the absence of all-cause mortality benefits from screening is poorly understood and suggests the possibility of deaths attributable to mammography. But such effects have been difficult to identify in large studies designed to track one cause of death. Disease-specific mortality reductions remain measurable and supported by a mechanism (early detection), whereas the lack of all-cause mortality benefit has eluded explanation. This discrepancy may lead to ready acceptance of screening because of a “mechanism bias,” by which findings that reflect established mechanisms are more easily accepted than those that do not (34).

Both the Journal of the American Medical Association report (2) and findings in the US Preventive Services Task Force report (1), however, hint at an increasingly plausible theory to explain the discrepancy between disease-specific and all-cause mortality, that is, overdiagnosis. Recent data suggest that a substantial fraction of detected cancers are of no threat. Observational and trial data have documented remission of untreated neoplasms, including cervical (35), colon (36,37), renal cell (38), melanoma (39), and breast (40) cancers. These data suggest either spontaneous regression or eradication by host mechanisms. In addition, data from large interventional screening trials for breast cancer (41), prostate cancer (42), and neuroblastoma (43) have uncovered surprisingly high rates of overdiagnosis (44). In perhaps the best documented example, urine screening of infants for neuroblastoma led to the detection of two to three times as many cancers as occurred in the unscreened group, suggesting that fully one-half to two-thirds of detected cancers would never have become clinically apparent (43,45).

For breast cancer, overdiagnosis is common. Ductal carcinoma in situ, which is disproportionately detected by mammography when compared with other forms of breast cancer, poses no threat in most cases, but treatment is often undertaken based on risk of subsequent disease (46). Recent analysis also suggests that one in three “invasive” cancers detected by mammography are nonthreatening, although all are treated (44). For prostate cancer, the slow growth and unpredictable metastases manifested in the disease continue to stir debate about treatment approaches (47), and one large study of PSA screening demonstrated a 50% overdiagnosis rate (42).

Modern treatments for breast and prostate cancers typically include some combination of surgery, radiation, and chemotherapy. Mortality rates from these therapies are low but real (48,49). If up to one-half of treated cancers are not life-threatening, it is plausible that during trials of screening modalities, small lifesaving benefits may be eclipsed by deaths due to treatment. In one neuroblastoma screening study (43), investigators found that the three fatalities in the screened group were attributable to complications of treatment (two from surgery and one from chemotherapy), all in children with early-stage cancer and no clinical manifestations.
Despite high rates of overdiagnosis, however, there has been no consistent increase in mortality because of screening. Mammography and PSA may therefore be effective when partnered with an accurate prediction of threat. Unfortunately, reliably forecasting the action of a complex neoplastic process inside of a complex organism has proven elusive.

Many common interventions have been accepted into practice on the basis of surrogate endpoints. Now that surveillance and public health data have suggested disappointing results from screening (2), and with an increasingly plausible explanation (overdiagnosis), our wide acceptance of disease-specific mortality as a surrogate endpoint should be revisited. The search for effective methods of cancer prevention and treatment must continue, but to continue to progress and to engender trust moving forward, we may have to acknowledge that for decades now, the most important outcomes in cancer screening trials have been ignored.

References


**Funding**

There were no funding sources for this article.

**Notes**

No potential conflicts of interest exist in the authorship of this manuscript.

**Affiliation of author:** Department of Emergency Medicine, Mount Sinai School of Medicine, New York City, NY.