Published in final edited form as:

Semin Oncol. 2010 June ; 37(3): 202–215. doi:10.1053/j.seminoncol.2010.05.006.

Principles of Cancer Screening: Lessons from History and Study Design Issues

Jennifer M. Croswell, M.D., M.P.H.,

Acting Director, Office of Medical Applications of Research, National Institutes of Health,
croswellj@od.nih.gov

David F. Ransohoff, M.D., and

Professor of Medicine, Clinical Professor of Epidemiology, Schools of Medicine and Public Health, University of North Carolina at Chapel Hill, ransohof@med.unc.edu

Barnett S. Kramer, M.D., M.P.H.

Associate Director for Disease Prevention, Office of Disease Prevention, National Institutes of Health, bk76p@nih.gov

Abstract

Early detection of cancer has held great promise and intuitive appeal in the medical community for well over a century. Its history developed in tandem with that of the periodic health examination, in which any deviations—subtle or glaring—from a clearly demarcated “normal” were to be rooted out, given the underlying hypothesis that diseases develop along progressive linear paths of increasing abnormalities. This model of disease development drove the logical deduction that early detection—by “breaking the chain” of cancer development—must be of benefit to affected individuals. In the latter half of the 20th century, researchers and guidelines organizations began to explicitly challenge the core assumptions underpinning many clinical practices. A move away from intuitive thinking began with the development of evidence-based medicine. One key method developed to explicitly quantify the overall risk-benefit profile of a given procedure was the analytic framework. The shift away from pure deductive reasoning and reliance on personal observation was driven, in part, by a rising awareness of critical biases in cancer screening that can mislead clinicians, including healthy volunteer bias, length-biased sampling, lead-time bias, and overdiagnosis. A new focus on the net balance of both benefits and harms when determining the overall worth of an intervention also arose: it was recognized that the potential downsides of early detection were frequently overlooked or discounted because screening is performed on basically healthy persons and initially involves relatively noninvasive methods. Although still inconsistently applied to early detection programs, policies, and belief systems in the United States, an evidence-based approach is essential to counteract the misleading—even potentially harmful—allure of intuition and individual observation.

Introduction: The ascendancy of intuition in cancer control

Calls from within the medical system for widespread screening for asymptomatic disease—with the attendant assumption that such early detection improves outcomes for patients—can be traced back more than a century. The evolution of programs for the early detection of

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

cancer can be linked, in part, to the development of the periodic health examination. Many authors pinpoint its inception to the British physician Horace Dobell.^{1–2} In his 1861 treatise, *Lectures on the Germs and Vestiges of Disease, and on the Prevention of the Invasion and Fatality of Disease by Periodical Examinations*, Dobell proposed the routine application of an exhaustive history, physical examination, and battery of laboratory tests to discover the “earliest evasive periods of defect in the physiological state;” he believed that periodic health exams were “the only means by which to reach the evil and to obtain the good.”³

A similar philosophy was emerging in the United States at the beginning of the 20th century. At the fifty-first annual meeting of the American Medical Association, physician George Gould extorted, “He is the greatest discoverer who finds the presymptom....There are a thousand undiscovered...advance scouts and forerunners, to be learned when the slight and unconscious departures from normality are studied by examinations of the supposedly well.”⁴ In statements like these, one can observe an emerging mindset in the medical community whereby there is a clearly demarcated physiological “normal,” as well as the hypothesis that diseases develop along linear paths of identifiable states of increasing abnormalities. This progressive course was intuited to be alterable if noticed early enough in its development.

These ideas began to be explicitly linked to cancer shortly thereafter. In 1907, Dr. Charles Childe published a book entitled *The Control of a Scourge, Or How Cancer Is Curable*. In it, he described the high mortality rates due to cancer as tragically unnecessary, and put “the finger on the flaw”:

Early cancer has no symptoms....The victims...are quite naturally lulled by the entire absence of symptoms into a sense of security....The ignorance of the importance of the early sign is so great, the temptation to make light of the apparently trivial symptom so natural....[But] cancer itself is not incurable...it is the delay that makes it so....If every case of cancer came under the notice of the [physician] at the earliest possible moment...it requires no stretch of the imagination...to say that the majority...would be cured.⁵

Although not yet explicitly codified into a formal theory, the concept of cancer as a series of advancing abnormal steps (i.e., the linear model of carcinogenesis) can already be seen to permeate medical discourse. This model of cancer development, in turn, can be seen to drive general assumptions about how best to tackle the disease. Because cancer was thought to advance along a series of progressive, increasingly dangerous steps, the logical deduction evolved that early detection of these aberrancies—by “breaking the chain” of disease development—must be of benefit to the affected individual. Indeed, Childe takes this assumption to its extreme, by suggesting that with due diligence, cancer could essentially be eradicated, simply by vigilant identification of “trivial” early symptoms.

The use of the periodic health exam for early disease detection became cemented into general medical practice in the 1920s, when the American Medical Association published an official endorsement and review of the method, and led campaigns to spread its use. The AMA deemed the universal institution of the practice an obvious step for practitioners: “Medical experience of the benefits of periodic examinations of presumably healthy persons is sufficiently widespread to make any detailed reference superfluous.”⁶ The intuitive appeal of this approach was so seductive that even anecdotal case reports and case series of its usefulness (which constituted the bulk of medical evidence during this period) were not demanded for its general approval and uptake into clinical practice.

The 30s, 40s, and 50s saw an upsurge of large-scale health campaigns directed at the public advocating early cancer detection, principally attributable to efforts of the American Society

for the Control of Cancer (ASCC, later to become the American Cancer Society). The ASCC organized a “Women’s Field Army” in the “fight” against cancer; the focus was on informing the public of the value of the Pap smear, clinical and self-breast exams, and the importance of being vigilant for early cancer warning signs. “Delay Kills!” was a frequent slogan; one poster released during WWII claimed that more people died every 2 weeks from a delayed cancer diagnosis than had died at Pearl Harbor.⁷

Criteria for screening: The limitations of intuition

In the late 1960s, the World Health Organization published a paper on the *Principles and Practice of Screening for Disease*. In it, Wilson and Jungner elucidated 10 fundamental principles that were intended to guide decision-making regarding institution of a given screening test:⁸

1. The condition sought should be an important health problem
2. There should be an accepted treatment for patients with recognized disease, and treatment should be better at an earlier stage
3. Facilities for diagnosis and treatment should be available
4. There should be a recognizable latent or early symptomatic stage
5. There should be a suitable test or examination
6. The test should be acceptable to the population
7. The natural history of the condition, including development from latent to declared disease, should be adequately understood
8. There should be an agreed-upon policy on whom to treat as patients
9. The cost of case-finding (including diagnosis and treatment of patients diagnosed) should be economically balanced in relation to possible expenditure on medical care as a whole
10. Case-finding should be a continuing process and not a “once and for all” project.

It is instructive to discuss a few of these points in detail in considering the limits of early detection strategies. By providing a clear set of criteria required to ensure the effectiveness of a screening test, these principles initiated a modicum of resistance against the previously unchecked swell of intuitive thought in cancer control. Applying the WHO criteria, the periodic health exam’s exhaustive search for any and all “advance scouts” of disease falls considerably short.

The first—that **the condition should be an important health problem**—speaks to the burden of disease in the population, and the overall risk-benefit ratio of utilizing mass screening for that group. One issue of consideration for rare diseases is whether energies might be better applied to refining treatment strategies, rather than population-based screening, because this would allow for more targeted use of finite resources. Furthermore, systematically testing the entire population for a rare condition might actually harm more individuals through the morbidity of the exam and diagnostic follow-up than could ever benefit from the test. The idea that an early detection strategy might have important associated harms that could outweigh any potential benefits was not commonly considered until this time.

Earlier treatment of a given disease must bestow a clinical benefit to a patient for early detection strategies to be worthwhile. If clinical outcomes are the same regardless of when in the course of the disease the person receives therapy, then there is no justification to be

made for diagnosing the person at an earlier point in time. Similarly, **there should be an accepted treatment for the disease**, because there is harm in diagnosing a disorder earlier when there is nothing to offer the patient in the way of mitigating the illness. In this situation, the person gains nothing from the early discovery, but any negative effects of learning about the presence of the disease—such as anxiety, depression, and economic burdens of care—can occur sooner.

There should be a recognizable latent or early symptomatic stage to the disease. This criterion points out that if there is no identifiable stage prior to the onset of symptoms, only diagnosis, and not early detection, is possible. The concept that **there should be a suitable early detection test or examination that is acceptable to the public** is closely related, since one aspect of suitability is the test's functionality. Furthermore, if the screening modality is highly invasive, inconvenient, or unpleasant, it may have a reduced chance of success because many individuals will simply refuse the test.

The natural history of the condition, including development from latent to declared disease, should be adequately understood. This criterion has several critical issues embedded within it. The first of these goes back to the need to establish that earlier diagnosis and treatment does, in fact, confer a clinical benefit to a patient. The natural history of a given disease impacts the probability that earlier intervention improves outcomes, because if one possible course of the disease is the spontaneous resolution/regression of a previously developing lesion, to avoid net harm to an individual, one should attempt to restrict detection and treatment to beyond that nodal point in the disease's development. An example of this situation would be cervical intraepithelial neoplasia (CIN), a non-obligate precursor of cervical cancer. The majority of CIN I lesions will, in fact, spontaneously regress within 1 year; recent evidence suggests approximately half of CIN II lesions will behave in a similar fashion.⁹ Cervical cryotherapy or electrosurgical loop excision may permanently alter the competency of the cervix and, as such, can have negative effects on a woman's fertility and pregnancies, so if performed on a lesion that would have been destined to regress, the intervention could result in pure harm for that person.¹⁰

A second point critical to consider when instituting new early detection programs is that screening alters the hitherto understood natural history of a given form of cancer. A useful analogy is picturing cancer as an iceberg of disease. Symptomatic lesions comprise the tip of iceberg--the part that is above the waterline--and, as such, can be easily observed. However, cancer includes a wide array of heterogeneous lesions that vary substantially in behavior. In the absence of screening, these asymptomatic lesions lie below the waterline of observation and little or nothing is known about the natural history of such tumors. Therefore, the information that has been gathered about the expected course and prognosis of a given cancer type is only based upon the behavior of symptomatic lesions. A new screening modality essentially dips beneath the water surface for the first time, and reveals previously hidden lesions. This complicates accurate predictions regarding the course of screen-detected disease, because the natural history of this early lesion has not previously been observed and followed. These asymptomatic lesions may be far less aggressive in nature or may follow a different disease course than those diagnosed on the basis of symptoms; one cannot simply assume that the natural history of screen-detected lesions would follow the same course as that of symptomatic tumors.¹¹

The 1960s and 70s also saw, for the first time, the medical profession seriously questioning and rigorously testing the effectiveness of the periodic health exam and select early detection modalities through randomized controlled trials. A new mode of thought was emerging in medicine, whereby clinical intuition alone, and a reliance upon pure deductive

reasoning based upon a scientific theory (that is, the linear model of carcinogenesis), were no longer considered sufficient proof of the merit of early detection strategies. The Kaiser Permanente Multiphasic Health Checkup Evaluation and the South-East London Screening Study Group were two such randomized trials that studied the impact of a comprehensive health examination on overall mortality.^{12–14} Neither demonstrated a positive result. Mammography was formally investigated in randomized trials, including the Health Insurance Plan of Greater New York (HIP) Study¹⁵, the Malmo study¹⁶, and the Swedish Two-County Trial.¹⁷ Although these trials had methodological limitations, all three showed evidence of reduced breast cancer deaths for screened women. The effect of screening with chest x-ray or sputum cytology on lung cancer mortality was also evaluated in the 1970s in three randomized trials sponsored by the National Cancer Institute at Mayo Clinic, Johns Hopkins, and Memorial Sloan-Kettering.^{18–20} Although these studies also had limitations, none demonstrated a reduction in lung cancer deaths; there was even a suggestion of increased harm (attributed to the morbidity of treatment) in the screened population.²¹

These non-intuitive findings from randomized controlled trials focused additional scrutiny on screening strategies. Some guidelines-formulating organizations noted serious deficiencies in the evidence base used to support routine screening recommendations, and outlined the implications of these evidence gaps in the provision of clinical care. Two of these groups were the Canadian Task Force on the Periodic Health Examination and the United States Preventive Services Task Force (USPSTF). The efforts of these organizations contributed to a conceptual shift in medicine, as the groups explicitly laid out and challenged core assumptions and intuitions comprising many clinical practices, particularly in the arena of early detection.

The analytic framework of the U.S. Preventive Services Task Force:

Avoiding heuristic thinking

The USPSTF in particular played a critical role in the move away from intuitive thinking and the emergence of evidence-based medicine through their development of a formal, quantitative, and explicit methodology for determining the overall risk-benefit profile of a given procedure for a population: the analytic framework.²² Figure 1 provides an example of a basic analytic framework for early detection strategies. The essence of the framework is its ability to delineate necessary steps in the chain of evidence to demonstrate that a given practice will be of more benefit than harm to an individual or population. This conceptual framework is exceptionally valuable because it forces the identification and verification of the assumptions underlying the use of a practice. It rejects intuition and heuristic thinking in favor of transparent and objective steps towards scientific proof. Walking through the analytic framework can be instructive in highlighting the critical elements necessary to demonstrate the worth of a clinical intervention.

Considering the target population and both benefits and harms of an intervention

First, the framework demands a clear definition of the kinds of people in the population under consideration. It asks, who is the target of screening? This is done because the overall risk-benefit profile of the early detection strategy may change depending upon the characteristics of the chosen population; in fact, it may not just shift quantitatively, but qualitatively (from a net benefit to net harm). Positive or negative results from one subset of people cannot simply be extrapolated to the entire population. Also of importance, it forces one to consider potential harms of both the screening test and that of treatment for the disease. As pointed out by the WHO criteria, the existence of a successful treatment strategy is critical for the success of a screening program; in a similar vein, the associated harms of treatment also carry weight in the overall risk-benefit balance for an early detection strategy.

The outcome of importance for screening: A reduction in cancer mortality

The framework indicates that the most direct way of establishing the benefit of a test is through a trial that demonstrates its use results in a reduction in deaths, or, at the very least, a reduction in important morbidity (conceptualized by the line and arrow marked with a 1). When direct evidence is lacking, the framework points out that one must consider each link in the evidence chain before drawing any conclusions about the utility of a screening test.

Avoiding dangers of using intuition: how cancer screening is like flying by instruments

Why is this done? Because intuition—particularly in the field of cancer screening—can easily lead physicians astray. This is because there are powerful biases associated with early detection strategies that can fool even the most careful observer. An analogy to flying can be made here—in common situations (such as when flying in cold front clouds, or where haze meets the water line), pilots can be led astray by their senses, and this is the reason for the existence of instrument flying. Pilots learn not to rely upon their own observations in certain scenarios, which can deceive them; instead, they put their trust in what the instruments are telling them. These tools have been developed to help pilots fly straight, true, and safe when their instincts are giving them faulty information. The pyramid of evidence, the analytic framework, and other evidence-based methodological tools work in a similar capacity; they provide assistance and assurance where common intuition fails.

The analytic framework attempts to look across studies at the sum total of the body of evidence to determine where evidence has replaced assumptions and where gaps in our knowledge still persist. However, biases also exist that can easily obfuscate our ability to interpret individual studies. This is particularly true for observational evidence, which by their design cannot control for all of the factors that experimental (e.g., randomized) trials can. Here are some of the common biases in cancer screening that can send faulty signals, feed into misleading intuitive beliefs, and deceive even the most astute clinician.

Biases in screening: Healthy volunteer bias

There may be fundamental differences between people that agree to participate in prevention and early detection programs and those that do not. This is such a common finding that epidemiologists have given a label to the phenomenon: the “healthy volunteer effect.” On average, individuals responding to invitations to participate in cancer screening or other preventive interventions tend to be wealthier, to have attained a higher educational degree, to be more attuned to health messages (and consequently, exercise more, and smoke and drink less), and to live longer, than persons who do not participate.

For example, the control populations from the Multiple Risk Factor Intervention Trial (MR FIT), the Colon Cancer Control Study, and the Physicians’ Health Study each was documented to have substantially lower-than-expected mortality rates as compared to the general population.^{23–25} Non-respondents in the Japanese Public Health Center Study Cohort had higher age-adjusted relative risks for all-cause mortality than did respondents.²⁶ In the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, standardized mortality ratios for participants (SMR—the ratio of observed deaths to expected deaths, where 100 means the rates are the same in both groups) were as low as 28 for diabetes, 37 for cardiovascular diseases, 43 for all-cause mortality, and 56 for non-PLCO cancers. In fact, even injuries and poisoning occurred almost half as frequently than expected in the general population: the SMR was 64!²⁷

The healthy volunteer effect can mislead a clinician into believing that early detection leads to better outcomes for his or her patients, because although the very decision to participate is

a powerful confounder, the practitioner's immediate observation is that screened individuals, on average, appear to do better than those patients who refuse to be screened.

Biases in screening: Lead-time bias

The intent of screening is to advance the date of diagnosis to an earlier point in time than it would otherwise have been made. However, it does not necessarily follow that an individual's lifespan will be prolonged as a result of this activity. Although a person diagnosed through screening may spend more time aware of the existence of his or her disease, the date of his or her death might well remain unaltered. Figure 2 depicts this distinction visually. Lead-time bias is the time that exists between screen-detected and symptom-detected diagnosis—its existence is a mathematical certainty for all screening tests (the magnitude will, of course, vary). In Figure 2, although the date of diagnosis has been moved up considerably through application of a screening test, the time of death for the individual remains identical.

The concept of lead-time bias helps explain why the use of 5-year survival rates to judge the value of cancer screening can be so misleading. Clinical trials have documented the counterintuitive potential of lead-time bias by comparing observed cause-specific mortality rates to 5-year survival rates in participants. The Mayo Lung Project, a randomized controlled trial of chest x-ray plus sputum cytology versus usual care for lung cancer screening, is one example. As Figure 3 shows, the 5-year survival rate after diagnosis appeared to strongly support the widespread implementation of screening: it was 36% among screened participants versus 19% for controls, or almost double. However, lung cancer mortality rates were not statistically significantly different between these two groups; in fact, they even trended towards an increase in deaths among those that were screened!^{*21} The efficacy of a screening test (when compared with usual care) cannot be established by examining survival rates: lead-time bias inserts an inescapable flaw into the measure. Lead-time bias can also contribute to the positive feedback loop of screening, because, once again, an individual practitioner is unable to directly observe this confounder. His or her direct observation would lead to the belief that screening prolongs life—even though all it may, in fact, prolong, is the period of time during which the patient must be cognizant of the presence of disease, with any attendant psychological impact.

Biases in screening: Length-biased sampling

Cancer screening tests are more effective at identifying slower-growing lesions than rapidly advancing ones. Every tumor has its own window of time during which it can be detected in its asymptomatic state by a given screening test. This time span is necessarily shorter in duration for rapidly growing lesions that might harm a person; the time span is longer in duration for slowly growing lesions that may never cause problems. For fast-growing lesions, the probability that a screening test will be applied during the critical window of opportunity is smaller. The most important consequence of this phenomenon is that screening tests, by their very nature, select for cancers that innately possess favorable prognoses. In fact, even after adjusting for tumor stage, the method of diagnosis (screening versus symptomatic) has been shown to be an independent prognostic factor for breast cancer.^{28–29} This does not mean that the screening test, in and of itself, impacted the course of that disease; it simply means that the test has “stacked the deck” with more indolent lesions.

*Although survival and mortality seem to be two sides of the same coin, survival is **not** 1-mortality. These two statistics have different denominators. 5-year survival is the number of individuals with the disease alive after 5 years divided by the number of individuals diagnosed with the disease. Mortality, on the other hand, is the number of individuals that have died from the disease divided by the total population at-risk for the disease during a given time interval. Unlike survival, mortality rates are not affected by lead-time bias.

Biases in screening: Overdiagnosis

Overdiagnosis represents an extreme form of length-biased sampling. It occurs when a lesion is diagnosed that is so indolent that it would never go on to cause problems for the individual. Overdiagnosis should be distinguished from misdiagnosis. Misdiagnosis represents variability or uncertainty among pathologists regarding whether a lesion is malignant or not. It occurs because of the inherent subjectivity in reading pathology slides. Overdiagnosis, on the other hand, represents a lesion that would generally be read as a malignant tumor; however, it does not behave like a “typical” malignancy. Overdiagnosis can occur in two situations: 1) cancers that are so benign that they have virtually no growth potential: in fact, they might even spontaneously regress; 2) cancers that grow so slowly that the person would die of another competing cause of death before the tumor generated symptoms.

Although this is an extremely counterintuitive proposition, it builds on the fundamental point discussed earlier—tumor biology is not homogenous, and the term “cancer” in all likelihood represents a multiplicity of disease processes. Although the traditional explanation of the obligate stepwise development of cancer through a series of increasingly abnormal stages is accurate in some situations, evidence is mounting that this model does not adequately explain the entire range of behaviors associated with cancer.

The classic indicator that overdiagnosis is occurring is a rise in incidence of early-stage tumors coupled with a minor or even nonexistent decline in incidence of late-stage disease. The operating principle of an effective early detection strategy is that one “pulls” advanced cancers out of the future and treats them at an earlier stage. One would therefore observe a strong relationship between a screening-induced increase in the number of early-stage cancers and a resulting decrease in late-stage disease in the population.

However, when U.S. trends for localized (confined to organ of origin) and distant (metastases beyond organ or origin or to lymph nodes) incidence rates are examined for several cancers with broadly disseminated early detection strategies, the expected correlations between rises in early-stage disease and declines in late-stage tumors are not seen. Figure 4 presents ecologic data from the SEER database for individuals 40 years of age and older for breast cancer, melanoma, and prostate cancer. For both breast cancer and melanoma, increases in the total cancer rate are being driven almost exclusively by changes in the proportion of early-stage disease, with almost no observed differences in the incidence rate of late-stage disease over a period of 30 years. For prostate cancer, although there is a decrease in the rate of late-stage disease observed, the absolute rate of decline makes up only a tiny fraction of the associated increase in early-stage disease. The inference is that there are more new cases of early-stage disease being identified than cases of late-stage disease being prevented: this represents overdiagnosis.

In fact, one recent study has estimated that since the introduction of PSA screening, more than 1 million additional men have been diagnosed and treated for prostate cancer than would have occurred in the absence of screening; however, the decline in the mortality rate for prostate cancer observed over this same time period does not equal—or indeed, even approach—this observed magnitude of change.³⁰ A recently completed randomized controlled trial of prostate cancer screening concluded that for every prostate cancer death averted using PSA, 48 additional men would need to be treated over 9 years.³¹

Overdiagnosis is a phenomenon that has been demonstrated for multiple types of cancer. One of the earliest examples to be shown in clinical studies was for neuroblastoma in infants and young children. In the 1980s, it was discovered that a urine test for vanillylmandelic acid (VMA) could detect the presence of neuroblastoma before the onset of symptoms. As

symptomatic neuroblastoma of early childhood is an aggressive disease, several countries, including Japan, Germany, and one province of Canada, launched population-based screening programs for the disease. The overall incidence rate of neuroblastoma climbed in these programs; however, there was minimal or no decline in the rates of late-stage, symptomatic disease diagnosed. Furthermore, mortality rates in these populations remained essentially unchanged.^{32–36} Careful examination of the screen-detected lesions demonstrated that these tumors generally appeared to have a biological make-up and behavior that was fundamentally different from that of symptom-detected neuroblastoma.^{37–39} Due to these disappointing findings, Japan abandoned the mass screening efforts in 2003, and there has been no noticeable change in neuroblastoma mortality rates since discontinuing the program.⁴⁰

Overdiagnosis has also been observed in cancers that are traditionally thought of as fundamentally more aggressive in nature. For example, the Mayo Lung Project, discussed previously, examined the impact of chest x-ray and sputum cytology screening on lung cancer mortality. There were more early-stage cancers detected in the chest x-ray arm (99) versus in the control arm (51); however, the number of advanced tumors diagnosed were nearly the same in each group (107 versus 109, respectively).⁴¹ Sixteen years after the original completion of the trial, the imbalance in the number of cancers detected in the two arms persisted. As the trial failed to show a mortality benefit in the screened arm, the investigators concluded that the extra tumors must have represented overdiagnosis: No advantage was gained by diagnosing and treating these additional lesions, and additional late-stage tumors (to balance out this discrepancy) were not subsequently discovered in the control arm.²¹

Even screening modalities that have evidence of efficacy from randomized controlled trials are not free from the impact of overdiagnosis. A Cochrane review of mammography trials has calculated the overdiagnosis rate at roughly one-third of screen-detected cases.⁴² Overdiagnosis has become such a concern that some researchers are now calling for a change in nomenclature for early-stage screen-detected lesions; instead of continuing to label these lesions “cancers,” several researchers have proposed the term “indolent lesions of epithelial origin (IDLE tumors).”⁴³

Potential harms of screening: An underappreciated aspect of cancer control practices

As previously highlighted, one of the important contributions of the USPSTF was the introduction of an analytic framework that explicitly considered all of the links in the chain of evidence required to establish the worth of a screening test. Just as critically, the framework demands equal attention be paid not just to the potential benefits of earlier detection and treatment, but to the possible harms as well. Although cancer screening appears to be an essentially innocuous activity, as it is often a quick and essentially painless blood sample or radiological exam, the downstream stream effects of the screening exam can be surprisingly substantial. Because screening exams are performed on basically healthy persons, and because they (at least initially) involve relatively noninvasive methods, intuitive thinking in cancer control frequently overlooks the critical question of potential drawbacks of screening interventions. A brief review of associated harms of early detection strategies is therefore instructive.

The most common (though not most serious) burden of screening is a false positive test. An estimate of the cumulative false positive rate for annual mammography demonstrated that within a decade, a woman had a 50% chance of receiving at least one false-positive exam.⁴⁴ A more recent analysis examined the likelihood of false-positive results within the context

of multiple modality cancer screening programs, and found that after 14 tests (3 years), a man's risk of obtaining at least one false positive test was 60%; a woman's, 50%.⁴⁵

False positive tests are of concern for three reasons. First, they have the potential to generate negative psychological consequences, such as anxiety, depression, and changes in the overall perception of one's health status. For example, one study of men with a false-positive prostate-specific antigen (PSA) test found that 6 weeks after receipt of a normal biopsy, 40% of these men--compared with 8% in the control group with normal PSA results--still worried "a lot" or "some of the time" about a diagnosis of prostate cancer.⁴⁶ In another study, men with a normal prostate biopsy after a false-positive PSA test were also more likely to visit a urologist (71% versus 13%), undergo a second PSA test (73% versus 42%), and have another biopsy (15% versus 1%) within 1 year than men with a negative PSA test.⁴⁷

Secondly, as the above finding hints, false-positive tests can trigger a cascade of more invasive diagnostic follow-up testing, which, in the case of a false positive test, represents pure harm, as the individual, not having cancer, cannot benefit from the testing process. However, he or she is subject to any of the complications associated with the confirmatory procedure. One study of the magnitude of this burden of unnecessary testing estimated that the probability that a man who has a false positive exam would undergo at least one invasive test as part of a multi-modality cancer screening program over 14 tests was 29%; a woman, 22%.⁴⁵ Depending on the cancer in question, the level of invasiveness can be quite high. For example, in the Pittsburgh Lung Screening Study (3,642 participants), the use of low-dose CT for lung cancer screening resulted in 28 subjects undergoing a major surgical procedure (thoracotomy or video-assisted thoracic surgery) for a benign final diagnosis, compared to 54 persons with a final diagnosis of lung cancer: a ratio of 1 surgery for benign disease to 2 for cancer.⁴⁸ In the Prostate, Lung, Colorectal, and Ovarian (PLCO) cancer screening trial, almost 40% of women with a false-positive transvaginal ultrasound for ovarian cancer underwent a laparotomy, some with associated hysterectomy and/or oophorectomy.⁴⁵

Third, frequent false-positive exams, and the diagnostic testing and follow-up cascades they create, can generate substantial economic burdens for patients and for the healthcare system. Lafata et al calculated expenditures triggered by false positive screening exams in the PLCO trial and found an extra \$1,024 and \$1,171 for an individual woman or man, respectively, were spent in the first year after the false-positive test.⁴⁹

We have already discussed in detail one of the major potential harms of screening: overdiagnosis. Overdiagnosis represents pure harm to an individual, because although by definition he or she cannot hope to benefit from the detection of the indolent lesion, the individual runs the risk of any attendant harms from the resulting diagnostic testing, the treatment, and even the diagnosis of cancer itself. Furthermore, even the act of labeling a person with an illness can have an important impact in and of itself on that individual's well-being and quality of life. For example, a study of hypertension in an industrial setting found that after screened individuals were diagnosed with high blood pressure, annual absenteeism from work increased by nearly 80% (approximately an extra week of work), regardless of the severity of the hypertension or whether pharmacotherapy was initiated.⁵⁰ In the United States, one 2009 study documented increased rates of unemployment associated with the diagnosis of cancer (34% of cancer survivors versus 15% of healthy controls).⁵¹ While one cause of this disparity is likely physical limitations associated with cancer, job discrimination⁵² and difficulties juggling work and chemotherapy treatment schedules⁵³ may also contribute to the difference.

Finally, harms associated with treatment must also be factored into the overall risk-benefit ratio of adopting a cancer screening strategy. For example, the harms of radical prostatectomy for prostate cancer can be considerable, ranging from a 16% rate of urinary incontinence to a 75% rate of erectile dysfunction.^{54–55} Another example would be an estimated thirty-day post-operative mortality of 4–6% for lobectomy and 11–16% for pneumonectomy in lung cancer patients.⁵⁶

Common arguments advanced to justify cancer screening interventions

The previous discussion sought to highlight the myriad ways in which even the most perceptive clinician can be led astray by logic and personal experience in the field of cancer screening. Obfuscating biases abound, and real potential exists to inadvertently cause harm; both of these points underscore the need for rigorous, experimental trials of screening modalities *before* their widespread application in the general population. The analytic framework, as envisioned by the USPSTF, should serve as a touchstone for those involved in policy setting for mass screening programs, as it eschews unreliable intuitive thinking in favor of a clearly defined, explicit blueprint by which a test can be proven to be of benefit to a population. A brief review of common arguments employed in the justification of cancer screening tests reveals that, frequently, the fundamental principles previously described appear to be misunderstood, de-prioritized, or considered not relevant.

A greater burden of disease

One of the most frequent arguments employed in the justification of the wide-scale application of a screening modality is that the magnitude of the disease and its attendant suffering in and of itself validates the utility of screening. A 1993 editorial in favor of PSA screening for prostate cancer demonstrated this reasoning:

The National Cancer Institute is conducting a prospective, randomized trial to determine whether or not screening reduces the prostate cancer mortality rate, but it will take sixteen years to complete this study. It is estimated that half a million men will die of prostate cancer before this study is completed, and it is unrealistic to expect clinicians to refrain from using PSA for cancer detection in the meantime.⁵⁷

An advocacy group that favors early detection strategies for lung cancer utilizes a similar argument in defense of the use of low-dose CT scans:

It is unconscionable for any agency, public or private, to block lung cancer screening for high risk populations on the basis of a flawed study which will not be completed until 2009 or beyond [i.e., the National Lung Screening Trial, a randomized controlled trial of chest x-ray and LDCT for lung cancer screening]. During that time, another one million people will die of lung cancer.⁵⁸

The American Cancer Society employed a rationale of “screen pending the evidence” back in the 1970s, when it advocated third-party payment for screening chest x-rays in part due to the perceived unacceptability of doing nothing in the face of continuing deaths while clinical trials were underway.⁵⁹

When one holds this argument up to the analytic framework, it is clear that while the burden of disease is obviously an important statistic that drives scientific investigations into both early detection and treatment strategies, it does not provide an answer to any links in the evidence chain either establishing the efficacy of the screening effort, or in terms of quantifying the potential harms that must be offset by benefits to be useful. It may be intuitively appealing to assume that because more people have a disease, or because the suffering generated by the disease is greater than that associated with other ailments, screening must therefore impart a greater impact. However, the assumption that the specific

screening intervention in question is actually effective drives this chain of logic, and without additional evidence, remains untested and unproven. Even an effective test may have a high enough associated burden of associated adverse events to generate a net harm. For example, in considering whether to screen for prostate cancer, efficacy, while essential, is not the only concern. As Michael Berry has aptly noted, "It is important to remember that the key question is not whether PSA screening is effective but whether it does more good than harm."⁶⁰

A higher risk of developing the cancer

Related to the issue of burden of disease is the concept that because a certain subpopulation is at higher risk of developing the disease or its complications, this represents sufficient cause to screen for the disease. The 2009 National Comprehensive Cancer Network guidelines on prostate cancer early detection highlight this approach to screening. The NCCN recommends targeting high-risk groups such as African-Americans and men with a family history of prostate cancer, and has reduced the age to start screening in these populations to 40. The stated rationale behind this change—despite the acknowledgement that published results from randomized controlled trials of PSA were limited to men ages 50–74—is that “young men who belong to a high risk group have a heightened chance of dying of prostate cancer and will thus benefit from early testing, [whereas] for older men, more judicious use...is warranted to prevent overdetection.” The guidelines go on to note, “The Surveillance, Epidemiology, and End Results (SEER) Database shows that prostate cancer deaths begin to appear in men in their 40s. Accordingly, to prevent these tragic, untimely deaths, screening for prostate cancer should begin earlier.”⁶¹

The American Cancer Society takes a similar approach to justifying the use of MRI for breast cancer screening as an adjunct to mammography. They recommend annual MRI screening for women that experienced chest radiation between ages 10 and 30 years, and for women with Li-Fraumeni, Cowden, and Bannayan-Riley-Ruvalcaba syndromes or that are first-degree relatives of others with the diseases, based solely on lifetime risk for breast cancer (and not evidence of effectiveness in either experimental or observational studies).⁶²

Although this type of rationale, when held up to the analytic framework, does get at the question of the population of interest (that is, who should be screened), it does not address the full chain of evidence and cannot by itself answer either questions of test efficacy or the balance of benefits and harms, even for the designated subpopulation. Intuition suggests that targeting screening efforts at those most likely to be affected by a disease will result in a greater net benefit while reducing adverse effects of screening (since more people tested will, in fact, have the disease in question, and fewer healthy persons will be subjected to the harms of the test); however, in this case, the underlying assumption that requires scrutiny is whether the test in question is, in fact, effective for *anyone*. Without evidence that the test reduces mortality, there can be no assurance that this intuition will hold true.

Finding more cancers: A higher test sensitivity

Another popular rationale used to advocate for new screening tests is that of acceptable operating characteristics—that is, “appropriate” sensitivity (and, less frequently, specificity). Closely related to this is the argument that a screening test’s ability to increase the incidence of early-stage disease—to find more cancers sooner—is in and of itself sufficient proof of effectiveness.

The ACS guidelines for breast cancer screening with MRI also employed this rationale. It concluded that the “efficacy of breast MRI has been established” on the basis of evidence that showed that MRI was more sensitive when compared with mammography for detecting

cancers (71–100% versus 16–40% in high-risk populations), rather than on evidence that it reduced disease-specific mortality as compared with mammography. Furthermore, although the guidelines acknowledged that MRI has “significantly lower” specificity than mammography—which would result in more false positive tests results, biopsies, and, potentially, unnecessary treatment—this did not impact the judgment that MRI had been proven effective and should be implemented. It was instead used as an additional reason why “screening should be recommended only to women who have a high prior probability of breast cancer,” with the untested assumption that the risk-benefit ratio would tip in the correct direction in this subgroup.⁶²

The difficulty with using sensitivity (or the ability to detect more cancers) alone as an endpoint for evaluating the efficacy of a screening test is that this measure cannot answer the question of whether finding these additional cancers actually results in a decrease in deaths or in more benefit than harm. Although it is tempting to assume that the ability to detect more early-stage lesions must be of benefit, there are two countering forces that can affect the overall effectiveness of the test.

The first of these is that one must consider any changes in the frequency and severity of harms that might result from the new screening test. If the test is more invasive or associated with a greater complication rate, or if it leads to more invasive testing or treatment, the accumulated harms could potentially outweigh gains.

Second, and most critically, it is important to consider what proportion of the increase in sensitivity represents aggressive disease (“good” sensitivity) versus overdiagnosis of essentially indolent disease (“bad” sensitivity). An illustrative example can be found in lung cancer screening. Low-dose computed tomography has been found to be a more sensitive screening exam for lung lesions than chest x-ray; however, there are ongoing concerns about the relative proportion of aggressive disease detected through this newer modality. Serious questions have been raised by the finding—replicated in several Japanese studies—that, when CT is utilized as a screening tool, smokers and non-smokers are diagnosed with the same rates of lung cancer.^{63–65} This observation is sharply at odds with current understanding of the epidemiology and etiology of lung cancer, and suggests the presence of overdiagnosis.

Proponents of colonoscopy for colorectal cancer screening have also promoted the argument that the ability to find more lesions is sufficient grounds to advocate for a screening modality. In 2000, the *New England Journal of Medicine* published 2 studies that showed that colonoscopy is able to detect right-sided neoplasms that flexible sigmoidoscopy does not.^{66–67} This was not an unexpected finding, given that a sigmoidoscopic exam only extends to the splenic flexure, whereas a colonoscopy extends to the cecum. However, the accompanying editorial (“Going the Distance—The Case for True Colorectal Cancer Screening”) used this information to advocate for mass colonoscopy as the primary screening strategy for Americans:

Although the studies...fall short of proving that life expectancy is increased by performing colonoscopic screening of persons 50 years of age or older who are at average risk for colorectal cancer, such an extrapolation of the data is virtually irresistible....These two new reports reinforce the growing suspicion among physicians that in recommending flexible sigmoidoscopy...we are promoting a suboptimal approach....The barrier to reducing the number of deaths from colorectal cancer is not a lack of scientific data but a lack of...commitment....I believe it is time for both government and private insurers to provide coverage for colonoscopic screening for all persons 50 years of age or older..⁶⁸

The *New York Times* picked up on the story and echoed the sentiment that “the test most commonly recommended to screen healthy adults for colorectal cancer misses too many precancerous growths and should be replaced by a more extensive procedure.”⁶⁹

Limiting the evaluation of the effectiveness of colorectal screening tests solely to sensitivity overlooked several key issues: 1) the relative harms of “the more extensive procedure”—including bowel perforations, intestinal bleeding resulting in hospitalizations, and adverse events from anesthesia—which are more commonly observed with colonoscopy than flexible sigmoidoscopy; and 2) the fact that one must consider the overall efficacy of a *program* of screening rather than a single application of the test. It is an important question to determine if a test applied every 10 years is, in fact, as effective in reducing deaths from developing cancers compared with tests performed every 5 years (flexible sigmoidoscopy) or annually (fecal occult blood testing).⁷⁰

Of interest, two population-based case-control studies have found that although colonoscopy was statistically significantly associated with a reduced risk of colorectal cancer death for left-sided tumors, no such association between the risk of death from right-sided colorectal cancer and the receipt of a complete (i.e., to the cecum) colonoscopy was observed.⁷¹⁻⁷² Such findings call into question the implicit assumption that simply identifying more cancers (or polyps) must result in better patient outcomes.

The focus on single-test sensitivity as the essential criterion for effectiveness continues to be emphasized in colorectal cancer screening guidelines even today. In 2008, the American Cancer Society, U.S. Multisociety Task Force on Colorectal Cancer (composed of the American Gastroenterology Association, the American Society for Gastrointestinal Endoscopy, and the American College of Gastroenterology), and the American College of Radiology issued joint guidelines on colorectal cancer screening, in which they set a rule that any test with at least 50% sensitivity at the time of a single application, would be considered recommendable.⁷³

Furthermore, by placing new emphasis on prevention of cancer rather than early detection as the primary goal of screening, the guidelines take the argument one step further to suggest that the efficacy of a given modality should now be based upon its ability to find not cancer, but rather, precancerous or suspicious lesions. Again, the challenge in utilizing these criteria as the benchmark for effectiveness is that, by themselves, they fail to consider other important factors in the risk-benefit equation, including a reduction in mortality, the issue of overdiagnosis and overtreatment, and the relative harms of screening and diagnostic follow-up associated with a particular test.

Conclusion: Protecting ourselves from our intuitions

The science of early detection has undergone important developments over the past 100 years. Intuitive reasoning in cancer control policy and practice reigned essentially unchecked for much of the late 19th and 20th centuries. In an atmosphere where deductive logic alone, rather than rigorous testing of theory, reigned supreme, the linear model of carcinogenesis drove key assumptions about the implicit worth of early detection strategies. In the latter half of the 20th century, recognition of the importance of explicit testing of heuristic assumptions underlying the promulgation of many screening tests began to grow, given increasing knowledge about important obfuscating biases that can lead clinicians astray. The development of an analytic framework as a tool to allow for consistent, clear, and objective evaluation of a given screening strategy independent of intuitive assumptions, served as a critical advance for the field. Although inconsistently applied to early detection programs, policies, and belief systems in the United States, an evidence-based approach is

essential to provide preventive care that maximizes benefits while minimizing the burdens of medical intervention. Just like pilots, scientists and clinicians (as well as their patients) benefit immensely from the utilization of formal instruments designed to counteract the misleading—even potentially harmful—allure of intuition and individual observation.

References

1. Han PK. Historical changes in the objectives of the periodic health examination. *Ann Intern Med* 1997;127(10):910–917. [PubMed: 9382370]
2. Boulware, LE.; Barnes, GJ.; Wilson, RF., et al. Value of the Periodic Health Evaluation. Evidence Report/Technology Assessment No. 136. (Prepared by The Johns Hopkins University Evidence-based Practice Center under Contract No. 290-02-0018). Rockville, MD: Agency for Healthcare Research and Quality; 2006 Apr. AHRQ Publication No. 06-E011
3. Dobell, H. Lectures on the Germs and Vestiges of Disease, and on the Prevention of the Invasion and Fatality of Disease by Periodical Examinations. London: Churchill; 1861.
4. Gould GM. A system of personal biologic examinations the condition of adequate medical and scientific conduct of life. *JAMA* 1900;284:134–138.
5. Childe, CP. The Control of a Scourge, or How Cancer is Curable. New York: E.P. Dutton & Company; 1907.
6. Emerson H. Periodic medical examinations of apparently healthy persons. *JAMA* 1923;80:1376–1381.
7. Lerner, BH. The Breast Cancer Wars. New York: Oxford University Press; 2001.
8. Wilson, JMG.; Jungner, G. Principles and Practice of Screening for Disease. Geneva: World Health Organization; 1968. Accessed online at: whqlibdoc.who.int/php/WHO_PHP_34.pdf on October 15, 2009
9. Castle PE, Schiffman M, Wheeler CM, Solomon D. Evidence for frequent regression of cervical intraepithelial neoplasia-grade 2. *Obstet Gynecol* 2009;113(1):18–25. [PubMed: 19104355]
10. Sadler L, Saftlas A, Wang W, et al. Treatment for cervical intraepithelial neoplasia and risk of preterm delivery. *JAMA* 2004;291(17):2100–2106. [PubMed: 15126438]
11. Kramer BS, Brawley OW. Cancer screening. *Hematol Oncol Clin North Am* 2000;14(4):831–848. [PubMed: 10949776]
12. Dales LG, Ramcharan S, Feldman R, et al. Multiphasic Health Checkup Evaluation Study. 3. Outpatient clinic utilization, hospitalization, and mortality experience after seven years. *Prev Med* 1973;2:221–235. [PubMed: 4723638]
13. Friedman GD, Collen MF, Fireman BH. Multiphasic Health Checkup Evaluation: A 16-year follow-up. *J Chronic Dis* 1986;39:453–463. [PubMed: 3711252]
14. The South-East London Screening Study Group. A controlled trial of multiphasic screening in middle-age: Results of the South-East London Screening Study. *Int J Epidemiol* 1977;6:357–363. [PubMed: 608798]
15. Habbema JD, van Oortmarssen GJ, van Putten DJ, et al. Age-specific reduction in breast cancer mortality by screening: an analysis of the results of the Health Insurance Plan of Greater New York study. *J Natl Cancer Inst* 1986;77(2):317–320. [PubMed: 3461193]
16. Andersson I, Aspegren K, Janzon L, et al. Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ* 1988;297(6654):943–948. [PubMed: 3142562]
17. Tabár 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(8433):829–832. [PubMed: 2858707]
18. Muhm JR, Miller WE, Fontana RS, Sanderson DR, Uhlenhopp MA. Lung cancer detected during a screening program using four month chest radiographs. *Radiology* 1983;148(3):609–615. [PubMed: 6308709]
19. Stitik FP, Tockman MS. Radiographic screening in the early detection of lung cancer. *Radiologic Clinics of North America* 1978;16(3):347–366. [PubMed: 746141]

20. Melamed MR, Flehinger BJ, Zaman MB, Heelan RT, Perchick WA, Martini N. Screening for early lung cancer. Results of the Memorial Sloan-Kettering study in New York. *Chest* 1984;86(1):44–53. [PubMed: 6734291]
21. Marcus PM, Bergstrahl EJ, Fagerstrom RM, Williams DE, Fontana RS, Taylor WF, et al. Lung cancer mortality in the Mayo Lung Project: Impact of extended follow-up. *Journal of the National Cancer Institute* 2000;92(16):1308–1316. [PubMed: 10944552]
22. Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, et al. Current methods of the U.S. Preventive Services Task Force: A review of the process. *Am J Prev Med* 2001;20(3S):21–35. [PubMed: 11306229]
23. Multiple Risk Factor Intervention Trial Research Group. Multiple Risk Factor Intervention Trial: risk factor changes and mortality results. *JAMA* 1982;248:1465–1477. [PubMed: 7050440]
24. Church TR, Ederer F, Mandel JS, et al. Estimating the duration of ongoing prevention trials. *Am J Epidemiol* 1993;137:797–810. [PubMed: 8484371]
25. Final report on the aspirin component of the ongoing Physicians' Health Study. Steering Committee of the Physicians' Health Study Research Group. *N Engl J Med* 1989;321:129–135. [PubMed: 2664509]
26. Hara M, Sasaki S, Sobue T, Yamamoto S, Tsugane S. Comparison of cause-specific mortality between respondents and nonrespondents in a population-based prospective study: ten-year follow-up of JPHC Study Cohort I. Japan Public Health Center. *J Clin Epidemiol* 2002;55(2):150–156. [PubMed: 11809353]
27. Pinsky PF, Miller A, Kramer BS, Church T, Reding D, et al. Evidence of a healthy volunteer effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Am J Epidemiol* 2007;165:874–881. [PubMed: 17244633]
28. Shen Y, Yang Y, Inoue LY, Munsell MF, Miller AB, Berry DA. Role of detection method in predicting breast cancer survival: Analysis of randomized screening trials. *J Natl Cancer Inst* 2005;97(16):1195–1203. [PubMed: 16106024]
29. Krzyzanowska M, I Tannock. Should screen-detected breast cancers be managed differently? *J Natl Cancer Inst* 2005;97(16):1170–1171. [PubMed: 16106016]
30. Welch HG, Albertsen PC. Prostate cancer diagnosis and treatment after the introduction of prostate-specific antigen screening: 1986–2005. *J Natl Cancer Inst* 2009;101(19):1325–1329. [PubMed: 19720969]
31. Schröder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* 2009;360(13):1320–1328. [PubMed: 19297566]
32. Yamamoto K, Hayashi Y, Hanada R, et al. Mass screening and age-specific incidence of neuroblastoma in Saitama Prefecture, Japan. *J Clin Oncol* 1995;13(8):2033–2038. [PubMed: 7636545]
33. Woods WG, Tuchman M, Robison LL, et al. A population-based study of the usefulness of screening for neuroblastoma. *Lancet* 1996;348:1682–1687. [PubMed: 8973429]
34. Woods WG, Gao RN, Shuster JJ, et al. Screening of infants and mortality due to neuroblastoma. *N Engl J Med* 2002;346(14):1041–1046. [PubMed: 11932470]
35. Schilling FH, Spix C, Berthold F, et al. Neuroblastoma screening at one year of age. *N Engl J Med* 2002;346(14):1047–1053. [PubMed: 11932471]
36. Yamamoto K, Ohta S, Ito E, et al. Marginal decrease in mortality and marked increase in incidence as a result of neuroblastoma screening at 6 months of age: Cohort study in seven prefectures in Japan. *J Clin Oncol* 2002;5(1):1209–1214. [PubMed: 11870162]
37. Nickerson HJ, Matthay KK, Seeger RC, et al. Favorable biology and outcome of stage IV-S neuroblastoma with supportive care or minimal therapy: A Children's Cancer Group study. *J Clin Oncol* 2000;18(3):477–486. [PubMed: 10653863]
38. Brodeur GM, Look AT, Shimada H, et al. Biological aspects of neuroblastomas identified by mass screening in Quebec. *Med Pediatr Oncol* 2001;36:157–159. [PubMed: 11464873]
39. Kitanaka C, Kato K, Ijiri R, et al. Increased Ras expression and caspase-independent neuroblastoma cell death: Possible mechanism of spontaneous neuroblastoma regression. *J Natl Cancer Inst* 2002;94(5):358–368. [PubMed: 11880474]

40. Katanoda K, Hayashi K, Yamamoto K, Sobue T. Secular trends in neuroblastoma mortality before and after the cessation of national mass screening in Japan. *J Epidemiol* 2009;19(5):266–270. [PubMed: 19652443]
41. Wolpaw DR. Early detection in lung cancer. Case finding and screening. *Med Clin North Am* 1996;80(1):63–82. [PubMed: 8569301]
42. Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews*. 2009 4. Art. No.: CD001877. DOI: 10.1002/14651858.CD001877.pub3.
43. Esserman L, Shieh Y, Thompson I. Rethinking screening for breast cancer and prostate cancer. *JAMA* 2009;302(15):1685–1692. [PubMed: 19843904]
44. Elmore JG, Barton MB, Moceri VM, Polk S, Arena PJ, Fletcher SW. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338(16):1089–1096. [PubMed: 9545356]
45. Croswell JM, Kramer BS, Kreimer AR, et al. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Ann Fam Med* 2009;7(3):212–222. [PubMed: 19433838]
46. McNaughton-Collins M, Fowler FJ, Caubet JF, et al. Psychological effects of a suspicious prostate cancer screening test followed by a benign biopsy result. *Am J Med* 2004;117(10):719–725. [PubMed: 15541320]
47. Fowler FJ Jr, Barry MJ, Walker-Corkery B, et al. The impact of a suspicious prostate biopsy on patients' psychological, socio-behavioral, and medical care outcomes. *J Gen Intern Med* 2006;21(7):715–721. [PubMed: 16808772]
48. Wilson DO, Weissfeld JL, Fuhrman CR, et al. The Pittsburgh Lung Screening Study (PLuSS): outcomes within 3 years of a first computed tomography scan. *Am J Respir Crit Care Med* 2008;178(9):956–961. [PubMed: 18635890]
49. Lafata JE, Simpkins J, Lamerato L, et al. The economic impact of false-positive cancer screens. *Cancer Epidemiol Biomarkers Prev* 2004;13(12):2126–2132. [PubMed: 15598770]
50. Haynes RB, Sackett DL, Taylor DW, Gibson ES, Johnson AL. Increased absenteeism from work after detection and labeling of hypertensive patients. *N Engl J Med* 1978;299(14):741–744. [PubMed: 692548]
51. de Boer AG, Taskila T, Ojajärvi A, van Dijk FJ, Verbeek JH. Cancer survivors and unemployment: A meta-analysis and meta-regression. *JAMA* 2009;301(7):753–762. [PubMed: 19224752]
52. Main DS, Nowels CT, Cavender TA, Etschmaier M, Steiner JF. A qualitative study of work and work return in cancer survivors. *Psychooncology* 2005;14(11):992–1004. [PubMed: 15744780]
53. Nachreiner NM, Dagher RK, McGovern PM, Baker BA, Alexander BH, Gerberich SG. Successful return to work for cancer survivors. *AAOHN J* 2007;55(7):290–295. [PubMed: 17665826]
54. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcomes among prostate-cancer survivors. *N Engl J Med* 2008;358(12):1250–1261. [PubMed: 18354103]
55. Potosky AL, David WW, Hoffman RM, et al. Five-year outcomes after prostatectomy or radiotherapy for prostate cancer: the prostate cancer outcomes study. *J Natl Cancer Inst* 2004;96(18):1358–1367. [PubMed: 15367568]
56. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital Volume and Surgical Mortality in the United States. *N Engl J Med* 2002;346(15):1128–1137. [PubMed: 11948273]
57. Catalona WJ. Screening for prostate cancer: Enthusiasm. *Urology* 1993;42(2):113–115. [PubMed: 7690167]
58. Lung Cancer Alliance. LCA Position on Lung Cancer Screening. Web page. October 26, 2009. Accessed at: http://www.lungcanceralliance.org/eirc/screen/screening_positions.html on November 1, 2009.
59. McNaughton Collins M, Barry MJ. Controversies in prostate cancer screening: Analogies to the early lung cancer screening debate. *JAMA* 1996;276:196–199. [PubMed: 8667559]
60. Barry MJ. Screening for prostate cancer--the controversy that refuses to die. *N Engl J Med* 2009;360(13):1351–1354. [PubMed: 19297564]
61. Kawachi MH, Babaian RJ, Bahnson RR, et al. National Comprehensive Cancer Network Practice Guidelines in Oncology, v.2.2010. Prostate Cancer Early Detection. 2009 August 7; Accessed at: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site on October 15, 2009.

62. Saslow D, Boetes C, Burke W, et al. American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer J Clin* 2007;57:75–89. [PubMed: 17392385]
63. Sone S, Takashima S, Li F, et al. Mass screening for lung cancer with mobile spiral computer tomography scanner. *Lancet* 1998;351:1242–1245. [PubMed: 9643744]
64. Sone S, Li F, Yang ZG, et al. Characteristics of small lung cancers invisible on conventional chest radiography and detected by population based screening using spiral CT. *Br J Radiol* 2000;73:137–145. [PubMed: 10884725]
65. Sone S, Li F, Yang ZG, et al. Results of 3-year mass screening programme for lung cancer using mobile low-dose spiral computed tomography scanner. *Br J Cancer* 2001;84:25–32. [PubMed: 11139308]
66. Lieberman DA, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Chejfec G. Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380. *N Engl J Med* 2000;343(3):162–168. [PubMed: 10900274]
67. Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings. *N Engl J Med* 2000;343(3):169–174. [PubMed: 10900275]
68. Podolsky DK. Going the distance—The case for true colorectal cancer screening. *N Engl J Med* 2000;343(3):207–208. [PubMed: 10900282]
69. Grady D. More extensive test needed for colon cancer, studies say. *The New York Times*, Section A, p.1, column 5. 2000 July 20;
70. Pignone M, Saha S, Hoerger T, Mandelblatt J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. *Ann Intern Med* 2002;137(2):96–104. [PubMed: 12118964]
71. Baxter NN, Goldwasser MA, Paszat LF, et al. Association of colonoscopy and death from colorectal cancer. *Ann Intern Med* 2009;150(1):1–8. [PubMed: 19075198]
72. Brenner H, Hoffmeister M, Arndt V, Stegmaier C, Altenhofen L, Haug U. Protection From Right- and Left-Sided Colorectal Neoplasms After Colonoscopy: Population-Based Study. *J Natl Cancer Inst* 2010;102:89–95. [PubMed: 20042716]
73. Levin B, Lieberman DA, McFarland B, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: A joint guidelines from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* 2008;58:130–160. [PubMed: 18322143]

Figure 1. The Analytic Framework of the U.S. Preventive Services Task Force

Generic analytic framework for evaluating a screening test, adapted from the U.S. Preventive Services Task Force. From Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, et al. 2001. Current methods of the U.S. Preventive Services Task Force: A review of the process. *Am. J. Prev. Med.* 20(3S): 21–35.

Figure 2. Lead-time Bias

Figure 3. Survival versus Mortality: The Mayo Lung Project

Figure 3 demonstrates that although 5-year survival is improved in the screening arm compared with the control arm, deaths due to lung cancer were actually greater in the screening arm. This illustrates the problem with using survival rather than mortality calculations when attempting to evaluate the worth of a screening modality—lead-time bias gives a misleading appearance of benefit in survival rates even when there is increased disease-specific mortality in the screened population. From reference ²¹.

Breast cancer

Prostate cancer

Melanoma

Figure 4. U.S. Trends in Localized versus Distant Cancer Incidence for Breast Cancer, Prostate Cancer, and Melanoma

Figure 4 depicts total, localized, and distant cancer incidence rates for breast, prostate, and skin cancer (melanoma) per 100,000 persons ages 40 and older in the U.S. “Localized” and “distant” are defined as by SEER historic stage A: “Localized” is an invasive cancer confined to the organ of origin; “distant” is an extension of metastasis to organs not adjacent to that of origin or to distal lymph nodes. From: Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 9 Regs Limited-Use, Nov 2008 Sub (1973–2006) <Katrina/Rita Population Adjustment> - Linked To County Attributes - Total U.S., 1969–2006 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2009, based on the November 2008 submission.