



# HHS Public Access

Author manuscript

*Surg Clin North Am.* Author manuscript; available in PMC 2016 October 01.

Published in final edited form as:

*Surg Clin North Am.* 2015 October ; 95(5): 953–966. doi:10.1016/j.suc.2015.05.009.

## Principles of Cancer Screening

**Paul F. Pinsky**

9609 Medical Center Drive, Room 5E108, Bethesda, MD, 20910, USA, pp4f@nih.gov

### Synopsis

Cancer screening has long been an important component of the struggle to reduce the burden of morbidity and mortality from cancer. Notwithstanding this history, many aspects of cancer screening remain poorly understood. This article presents a summary of basic principles of cancer screening that are relevant for researchers, clinicians and public health officials alike.

### Keywords

cancer screening; overdiagnosis; lead time; targeted screening; number needed to screen; performance characteristics

### Introduction

For over a half century, cancer screening has been an important component of the struggle to reduce the burden of morbidity and mortality from cancer. In certain cases, such as with cervical cancer, the effects have been dramatic, with mortality decreasing over 80% in the U.S. after implementation of wide-spread screening with Pap smears<sup>1</sup>. For most other cancers, however, the effects of screening have been substantially less pronounced. Benefits of screening have generally been on the modest side, and there has been increasing recognition of screening-related harms. The promise of cancer screening still beckons, though, and many new technologies continue to be evaluated for their potential to generate new screening modalities.

A standard introduction for a scientific paper concerning cancer screening proceeds as follows. The five year survival rate for cancer X is very low. Among stage I cases, though, five year survival is much higher; however, few cases are diagnosed in this stage. Therefore, if cancer X could only be detected earlier, the prognosis for subjects diagnosed with cancer X can be much improved.

In a nutshell, this is the basic, and very intuitive, rationale behind cancer screening. While intuitively appealing, though, some caveats are in order. First, the fact that cancers diagnosed in early stages have a (relatively) good prognosis does not necessarily mean that

---

**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.

Disclosures: None

if those cancers currently diagnosed in late stage were detected earlier they would also have the same favorable prognosis. It is possible, for some cancer types at least, that inherent properties of late stage tumors, such as their potential for early metastasis, and not the time of initial treatment, determine their eventual clinical outcome. Second, as a preventive intervention, screening tests are applied to asymptomatic and apparently healthy populations, where, due to the relatively low prevalence of any given cancer type, the vast majority of those being screened cannot benefit from the screening but can be harmed, due to either the screening test itself or to downstream consequences of it.

On the more favorable side for screening, there is the concept of detection through screening of pre-cancerous lesions. While standard screening programs are an example of secondary prevention, whereby cancer incidence is not reduced but mortality from the cancer is, screening modalities, such as those for colorectal and cervical cancer, that detect early cancers as well as pre-cancerous lesions provide both primary prevention (i.e., incidence reduction) and secondary prevention. In addition to incidence reduction being a substantial benefit in itself, both in terms of patient well-being and societal costs, screening modalities that reduce cancer incidence have greater magnitude reductions in cancer mortality than those that only provide secondary prevention.

In this manuscript, we will provide an overview of some basic concepts and principles in cancer screening. In Section 2, we discuss performance characteristics of screening tests (related to test accuracy). In Section 3 we discuss measures of screening benefit, while in section 4 some potential biases associated with evaluating screening benefits are reviewed. Section 5 deals with harms of screening. Section 6 discusses the concept of cost-effectiveness of screening and the related concept of targeting screening to high risk groups. Finally Section 7 summarizes current recommendations for cancer screening in North America.

## 2. Performance Characteristics of Screening Tests

The performance characteristics of a screening test refer to its ability to accurately predict disease state. Table 1 shows some common test performance characteristics. Sensitivity and specificity, and more generally the receiver-operating characteristic (ROC) curve of sensitivity at varying levels of specificity for continuous or ordinal valued tests, are critical in the research setting for evaluating the potential of new screening modalities. Positive predictive value (PPV) is more relevant in the clinical setting, in that it assesses the probability that a patient with a positive test has the cancer of interest. Importantly, PPV depends not only on sensitivity and specificity, but also critically on the prevalence of the cancer being screened for (technically, this is the prevalence of underlying, undiagnosed cancer). With fixed sensitivity and specificity, PPV decreases as prevalence decreases. Since cancer prevalence in a screened population will be low, even high specificity values can lead to very low values of PPV, regardless of sensitivity. For example, for a prevalence of 0.6% (e.g., breast cancer in women recommended for mammography screening) and sensitivity and specificity of 90%, PPV is only 5%; PPV increases to 10% and 21%, respectively, as specificity increases to 95% and 98%. This highlights the general requirement in cancer screening that specificity should be quite high. Negative predictive value (NPV) can be

useful in a diagnostic context for a rule-out; however, in a screening context it is generally not informative because the pre-test probability that the disease is absent is already so high.

The term “false positive rate” in a screening context usually refers to 1-Specificity (or 100-Specificity when specificity is expressed as a percent). However, it may be used also to refer to 100-PPV, so use of the term may be confusing. Since specificity is generally well above 50% in a cancer screening context and PPV below 50%, the value of the “false positive rate” will usually give a hint as to which definition is being used (a false positive rate over 50% usually refers to 100-PPV and under 50% to 100- Specificity).

A cancer screening test may produce a continuous measure (such as a blood concentration, e.g. prostate-specific antigen or PSA) or a binary (Yes/No) result of whether a suspicious lesion is present. With some imaging modalities, e.g., low-dose CT for lung cancer screening, the size of the largest nodule may essentially provide a continuous-valued result among positive screens. With continuous or ordinal-valued tests, a standardized cutoff generally determines whether the test is considered positive (e.g., >4ng/ml for PSA). Additionally, for continuous or ordinal-valued tests, PPV can be reported per positive test category or range as well as overall. For example, in the National Lung Screening Trial (NLST), overall PPV at baseline for low-dose CT was 3.8%, but PPV ranged from 0.5% for positive screens with nodule size of 4–6mm to 41% for positive screens with nodule size >30mm<sup>2</sup>.

Note that a common error is to attempt to estimate PPV directly from a 2 by 2 table from a case-control study. Since the case to control ratio is arbitrary in such a design, it is erroneous to compute PPV as the number of cases with a positive test divided by the number of all study subjects with a positive test. In a prospective (cohort) study such a method is legitimate, but in a case-control design an outside estimate of prevalence is needed in order to compute PPV.

A standard assumption is that a cancer screening test must have a high level of sensitivity in order for it to be effective. However, high sensitivity alone for a screening modality does not necessarily imply that the modality will have any mortality benefit, and a sensitivity level sufficient for one screening modality and cancer type may not be sufficient for another. For example, a recent large study demonstrated a sensitivity for film screen mammography, a screening modality with a demonstrated mortality benefit, of 66%<sup>3</sup>. In the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial, the combined modality of CA125 and trans-vaginal ultrasound for ovarian cancer screening had a similar 66% sensitivity; however, there was no observed mortality benefit of screening (the screened arm actually had slightly higher mortality from ovarian cancer)<sup>4</sup>.

With screening tests where pre-cancerous lesions may be detected, with a proven resultant decrease in cancer incidence, such as those for colorectal cancer, it is important to also consider sensitivity for such lesions. The fecal immunochemical test (FIT) has shown relatively high sensitivity for colorectal cancer, 61–91%, as compared to 90–95% for colonoscopy<sup>5</sup>. However, for the important cancer precursor of advanced adenoma, the relative sensitivity of FIT compared to colonoscopy is much lower.

This example of FIT and colonoscopy brings up the concept of program sensitivity. Program sensitivity measures the ability of a periodic screening regimen (e.g., annual screening) to detect cancer early, or to detect cancer precursors. Since FIT is recommended for annual use, while colonoscopy is recommended only every 10 years, FIT could have several opportunities to detect a pre-clinical cancer, or an advanced adenoma, whereas colonoscopy would (generally) only have one. Therefore, examined as an overall screening program, the sensitivity of FIT for cancer or advanced adenoma is closer to that of colonoscopy than the values at a single time point might imply.

Analogous to program sensitivity is the concept of the cumulative false positive rate. With periodic screening, the chance of experiencing at least one false positive test obviously increases over time. The choice of a specific screening program, in terms of testing frequency, does not affect the false positive rate at any given screen but does influence the cumulative false positive rate. For example, biennial instead of annual screening over a fixed age range would clearly lead to a reduced cumulative false positive rate over subjects' lifetimes; note it could also reduce program sensitivity.

### 3. Measures of Screening Benefit

In a randomized controlled trial (RCT) of cancer screening, the primary outcome is typically cancer-specific mortality, defined as the rate of death from the cancer of interest<sup>6</sup>. A metric of screening efficacy is the cancer-specific mortality rate ratio (RR), or the ratio of cancer specific death rates in the screening versus control arm. Overall mortality is not used as the primary endpoint in cancer screening trials because, since deaths from the cancer of interest will be a small fraction of all deaths, there is too much "noise" from non-relevant deaths and the trial would require enormous sample size to be adequately statistically powered<sup>6</sup>.

Meta-analyses of randomized trials of mammography showed pooled RR estimates for breast cancer mortality of 0.85 for women aged 40–49, 0.86 for women 50–59 and 0.68 for women 60–69<sup>7</sup>. In the National Lung Screening Trial (NLST) comparing low-dose CT to chest radiograph, the updated lung cancer mortality RR was 0.84<sup>8</sup>. For PSA-based prostate cancer screening, the European Randomised Study of Screening for Prostate Cancer (ERSPC) showed a mortality RR of 0.79; however, the U.S. Prostate, Lung, Colorectal and Ovarian (PLCO) trial did not show any mortality benefit (RR=1.09)<sup>9,10</sup>. Four randomized trials of flexible sigmoidoscopy screening showed a mean (range) of RR for colorectal cancer mortality of 0.75 (0.69–0.80)<sup>11–14</sup>. Several trials are ongoing for colonoscopy screening but none have reported primary outcomes to date. Since colonoscopy covers the entire colorectum and flexible sigmoidoscopy only the distal colorectum, the RR would be expected to be lower for colonoscopy. Mortality RRs for distal colorectal cancer with flexible sigmoidoscopy were lower than those for overall colorectal cancer (mean 0.67, range 0.50–0.87)<sup>11–14</sup>.

For screening modalities that also prevent cancer, such as flexible sigmoidoscopy and colonoscopy, the cancer incidence rate ratio from a RCT is a measure of efficacy for primary cancer prevention. RRs for colorectal cancer incidence in the above four flexible sigmoidoscopy trials averaged 0.81 (range 0.77–0.87)<sup>11–14</sup>.

Another metric commonly used is the number needed to screen (NNS), which is calculated as the reciprocal of the difference in cancer-specific mortality rates between arms<sup>15</sup>. For example, in NLST, which enrolled high-risk ever smokers, the NNS was 320, meaning 320 (NLST-eligible) subjects would have to be screened with low-dose CT according to the NLST protocol (a total 3 of annual screens) to prevent one lung cancer death<sup>8</sup>. The NNS measure is more appropriate from a public health and cost-effectiveness perspective because it measures how many deaths will be prevented per number of subjects undergoing screening. The NNS takes into account the mortality RR as well as the background death rate of the cancer of interest. For the same RR, NNS increases (meaning screening becomes less efficient) as the background death rate decreases. For example, comparing low-dose CT in high-risk smokers and mammography, the mortality RRs are roughly the same but the death rate from lung cancer in high risk ever smokers is approximately 5 times the death rate from breast cancer in the population recommended for mammography. Therefore, the NNS is roughly five times higher for mammography in its recommended population than for low-dose CT in its recommended population of high-risk ever smokers.

### Population Measures

A screening modality only has true benefit when it disseminates outside of research settings to the general population. The ultimate metric of benefit of a cancer screening test is the overall population reduction in deaths from the cancer of interest that is attributable to screening. This will be a function of the degree of dissemination of screening as well as the effectiveness of screening in reducing cancer-specific mortality. For screening modalities that prevent cancer incidence (e.g., screening for colorectal and cervical cancer), reduction in population incidence is also a critical metric.

The reduction in mortality that is attributable to screening is often difficult to assess unless there is a dramatic effect, such as with cervical cancer, where the screening modality reduced incidence and very sharply reduced mortality. Mortality rates from cervical cancer in the U.S. were 12–13 (per 100,000 women) in the 1950's before the common use of Pap smears, versus around 2 currently, and cervical cancer incidence decreased by about 60% from before to after the introduction of Pap smears<sup>1</sup>. Colorectal cancer screening also reduces cancer incidence, although the effects to date have not been as dramatic as for cervical cancer screening. Colorectal cancer incidence rates (age adjusted) in the U.S. were about 65 per 100,000 from 1975–1985 (pre-screening), with corresponding mortality rates of around 28 per 100,000<sup>16</sup>. In 2011, incidence was 39 per 100,000 and mortality 15 per 100,000<sup>16</sup>.

For those screening modalities that do not prevent cancer, analyzing secular trends in mortality is a starting point to evaluating the effect of screening. However, because the mortality benefits of such screening are generally modest at best, it is often difficult to tease apart the effect of screening from that of improvement in treatment or other factors. For example, a high-profile modeling analysis of mammography by seven independent groups (incorporating breast cancer natural history, mammography sensitivity, RCT results, population incidence and mortality trends and mammography usage rates) estimated the

effect of the introduction of mammography on breast cancer mortality. The estimates across the 7 models ranged from 7.5% to 22.7% for the mortality decrease due to mammography<sup>17</sup>.

#### 4. Common Biases in Assessing the Benefits of Screening

The RCT is the gold standard for assessing the effectiveness of a cancer screening modality. However, screening may be implemented in the absence of a RCT, or before results from a RCT are reported. For example, prostate cancer screening with PSA began to be implemented in the U.S. in the early 1990s, before an RCT was initiated<sup>18</sup>. No RCTs reported until 2009, when the European trial (ERSPC) reported a positive result and the U.S.PLCO trial reported a negative result, making the question of PSA benefit still uncertain<sup>9–10</sup>. In addition, with changes in screening technology or improvements in treatment, prior RCT results may have become out of date and more recent data may be needed to evaluate screening as currently practiced.

Researchers thus must often rely on observational or population-level studies to help assess screening benefit. This is problematic, however, due to several important potential biases that tend to affect studies of screening.

Two common related biases in non-randomized studies of screening are lead time bias and overdiagnosis bias<sup>6</sup>. Early detection through screening implies an advancement in the time of diagnosis of the cancer from what would have otherwise occurred in the absence of screening. The concept of "lead time" refers to the length of this period of time advancement. Since diagnosis is advanced and before any symptoms, it is possible that, in the absence of screening, clinical diagnosis would never have occurred, either due to the inherent indolence of the cancer or due to competing causes of death. This phenomenon of screening detecting a cancer that never would have otherwise become clinically apparent is known as overdiagnosis. Both overdiagnosis and lead-time are theoretical concepts, in that they can generally not be observed in a given individual but can be estimated statistically in populations. Figure 1 shows hypothetical examples of individual time courses of screened subjects, illustrating lead time and overdiagnosis, as well as a true benefit of screening. Lead time and overdiagnosis can lead to scenarios in observational studies where screening appears to be beneficial even though the modality may actually have no effectiveness in reducing mortality from the cancer.

A common approach in observational studies of screening is to compare cancer survival among screened and non-screened populations. Note that in the cancer screening context, "survival" refers to death rates following cancer diagnosis whereas "mortality" refers to the death rate in an entire population cohort, whether diagnosed with cancer or not. Survival is a notoriously misleading statistic for evaluating screening effectiveness due to the effects of lead-time and overdiagnosis bias, which tend to make survival rates more favorable in screened populations. By definition, there are no deaths from the cancer of interest among overdiagnosed cases or during screen detected cases' lead time (except for deaths caused by the cancer treatment itself). Therefore, even in the absence of any mortality benefit of screening, survival metrics can be greatly improved upon implementation of screening. For prostate cancer in the U.S., 5-year relative survival increased from around 70% in the late



1970s to 99.3% by 2005<sup>16</sup>. In terms of death rates (i.e., 100-survival), this corresponds to a 97% decrease (30% to 0.7%). Yet mortality rates decreased much more modestly, from 31 per 100,000 in the late 1970s to 24 per 100,000 in 2005 (only a 23% decrease). Whether (or how much of) this 23% decrease was due to screening, as compared to improvements in treatment, is not totally clear; however, even with no mortality benefit a large survival improvement would still have been observed, due to lead time and overdiagnosis bias. When examining survival specifically in subjects with screen-detected cancers, as opposed to analyzing survival in population registries, the potential for bias is even greater since it is not diluted by cancers in subjects not undergoing screening.

For comparison, for Hodgkins disease, a cancer with no screening but improvements in treatment, five-year survival improved from 1975 to 2006 from 71.4% to 88.3%, corresponding to a change in 5- year case-fatality rates (i.e., 100-survival rate) from 28.6% to 11.7%, a 59% decrease. In that same period, death rates from the disease decreased by a similar 66%<sup>16</sup>.

Utilizing survival statistics to make a case for screening benefit is thus extremely problematic, though it is still commonly employed. The case of prostate cancer also highlights how misleading improvements in 5-year survival rates over time can be when there has been a corresponding increase in the utilization of screening.

Another bias in evaluating the effect of screening is selection bias<sup>6</sup>. This issue arises when one examines the (cancer-specific) mortality rate among a group undergoing screening to that in a group not undergoing screening or to population-wide statistics. Since those who choose to be screened may be different with respect to the incidence of and survival from the cancer of interest, these underlying factors, and not the screening itself, may be contributing to any observed differences in mortality rates between the screened and non-screened (or population-wide) group. Other studies examine the same geographic population over different time periods or different geographic populations at the same time with different exposures to screening; again, differences other than the screening per se may be operating across the populations.

## 5. Harms of Screening

In general, there are relatively few medical harms of the actual cancer screening tests themselves. Modalities that rely on X-ray radiation (low-dose CT, mammography) do convey some added cancer risk, however, that excess risk is generally acknowledged to be quite small<sup>19–20</sup>. Colonoscopy carries some risks, including for perforation, which is estimated in the range of 2–4 per 10,000<sup>21,22</sup>. There is minor discomfort with some screening tests, as well as some reported short-term anxiety associated with false positive test results<sup>23</sup>.

However, the primary medical harms of cancer screening come from downstream events, from the diagnostic work-up of false positive screens and from diagnosis and treatment of cancers that may never have become clinically apparent without screening (overdiagnosed and overtreated cases). With respect to harms from the diagnostic work-up of false positives, the overall harm will be a function of both the false positive rate (defined as 100-Specificity)

and the nature of the diagnostic work-up. For example, in the NLST, the false positive rate was quite high (average of 24% over all screening rounds); however, only a small fraction of false positives, 2.6%, were followed up by invasive diagnostic procedures (thoracotomy, thoracoscopy, bronchoscopy or needle biopsy)<sup>24</sup>. Of these, 2.4% were associated with a major complication (9.6% with any complication). Thus, as a proportion of all low-dose CT screens (without a cancer diagnosis), the major complication rate was very low, 1.5 per 10,000 (6.0 per 10,000 for any complication). For another example, with screening for prostate cancer with PSA and digital rectal exam (DRE), the PLCO trial found a post-biopsy complication rate of 2.0%<sup>25</sup>. Based on the false positive rate (12–14% across screening rounds), the frequency of follow-up biopsy and this complication rate, the frequency of complications from biopsy was about 7 per 10,000 screens. While these rates per screen are low, for many screening programs persons may undergo 20–30 screens over their lifetime, so the lifetime risk of experiencing a complication would be substantially greater than the one-time risk. Also, with mass screening of healthy populations, the absolute number of people with complications as a result of screening-induced diagnostic work-up is considerable.

### Overdiagnosis

Overdiagnosis and overtreatment are harms of screening that are being increasingly recognized. The harms of overdiagnosis depend not only on the magnitude of overdiagnosis, but also on the costs and harms, and likelihood, of treatment. Overdiagnosis is the critical issue in PSA-based screening for prostate cancer since there is both an elevated level of overdiagnosis and a high rate of serious quality of life harms (e.g., urinary incontinence, impotence) associated with the standard curative treatments of radiation and radical prostatectomy. Overdiagnosis is also a serious issue in mammography screening for breast cancer, and has even been recognized in low-dose CT lung cancer screening.

Quantitative definitions of overdiagnosis rates vary considerably; therefore, care must be taken when comparing estimates from different sources or across cancer types. For example, the “overdiagnosis rate” may be defined as the percentage of screen detected cancers that are overdiagnosed, the percentage of all cancers that are overdiagnosed in a population attending screening or in a population invited to attend screening, and finally, as the percentage of overdiagnosed cases among an entire population, only a fraction of which is actually undergoing screening.

Perhaps the most straightforward method for estimating the overdiagnosis rate is from a randomized screening trial of screening versus no screening; here the excess of screened arm (versus control arm) cancers following a long enough observation period without further screening, in order to allow the control arm to “catch up” to the screened arm in diagnosed cases, can be used to estimate overdiagnosis rates. In the ERSPC, prostate cancer incidence was 63% higher in the screened than the control arm; assuming no further control arm catch-up, this leads to an estimate of about 50% of the screen detected prostate cancers in ERSPC being overdiagnosed<sup>9</sup>. However, due to variations in data sources and methods, as well as definitions, estimates of overdiagnosis rates vary greatly. A meta-analysis showed a range of



estimated overdiagnosis rates for invasive breast cancer associated with mammography of 1–54%<sup>26</sup>.

A method of indirectly assessing overdiagnosis is to examine the change in incidence rates after the introduction of screening, assuming there has been no observed change in any known major risk factors for the cancer, as is true for prostate cancer in the U.S. Annual prostate cancer incidence rates in the U.S. ranged from 105 to 115 (per 100,000 men) in the period from 1980–1985, before the start of the PSA era<sup>16</sup>. In the middle of the PSA era, 1995–2005, the average prostate cancer incidence rate was 173, almost a 60% increase.

## 6. Cost-effectiveness and Targeted Screening

Even if a cancer screening modality has been shown in RCTs, or with other solid evidence, to reduce mortality from the cancer of interest, and if the benefits of the screening clearly outweigh the (medical) harms, there is still the question of cost-effectiveness. Cost-effectiveness analysis (CEA) is an approach to assessing the benefits and harms of a medical intervention that also takes into account resource utilization and/or cost issues.

CEAs often employ the metric of cost per quality adjusted life year (QALY) gained from screening, where the QALY attempts to incorporate quality of life effects as well as longevity. In a cancer screening context, the harms of screening contribute to lowered QALYs. For example, with PSA screening, the side effects of treatment decrease the QALYs in those treated cases that are estimated to have been overdiagnosed. Note that if the case were not overdiagnosed, then presumably the side effects and QALY decrement would have occurred even without screening, so such side effects would not impact the CEA. The costs of screening include those of the screening test itself, the diagnostic follow-up and treatment of any complications thereof, and costs of overtreated cases. It may also include factors such as time lost from work for the screening visit, and other associated costs.

A standard benchmark for interventions that are cost-effective is \$100,000 per QALY. Several analyses of colorectal screening have shown that at least some modalities (including annual screening with fecal immunochemical test (FIT) and flexible sigmoidoscopy) can actually reduce health care costs while increasing QALYs (i.e., their cost per QALY is actually negative)<sup>27</sup>. This is because CRC screening can reduce the incidence of CRC, CRC is a relatively common cancer and the cost of treating CRC can be very high. For other modalities where the screening does not reduce cancer incidence, the estimated cost per QALY has always been positive, meaning that the introduction of these modalities into the health care system does not reduce health care costs. Although some costs associated with treatment of late stage cancers may be reduced, this effect is overwhelmed by the cost of screening and diagnostic follow-up.

Cost effectiveness and cost per QALY estimates for the same screening modality and scenarios often vary widely, due to not only the underlying data sources and/or models that quantify the various harms and benefits, but also due to variability in numerical cost values, differences in what types of costs are included (e.g., lost productivity), and the choice of perspective (provider/patient versus societal). Assessing quality of life effects, in order to estimate QALYs, is also inherently subjective and variable. A study using five breast cancer

natural history models with costs and some other factors standardized across models showed a cost per QALY mean (range) for biennial screening with digital mammography for women 50–74 of \$36,000 (\$23,000–\$72,000)<sup>28</sup>.

### Targeted Screening

A general principle of cancer screening is that the net benefit (benefits minus harms) of screening generally increases as the incidence rate of the cancer of interest in the screened population increases. This follows because only those who have the cancer of interest can benefit from screening, whereas the harms of screening and diagnostic follow-up apply to all screened persons. Likewise, the relative cost-effectiveness of screening also increases with incidence, since costs as well as harms accrue from all screened persons.

Therefore, a strategy to maximize the net benefit of screening, as well as its cost-effectiveness, is to target screening to an identifiable subset of the population that is at above average risk for the cancer of interest; this is sometimes denoted as risk stratification or targeted screening.

For most cancer screening programs in the U.S., the primary criteria for who is recommended for screening are age and sex. The minimum age requirement is utilized because cancer incidence generally increases with age and screening in a setting where the incidence of the cancer of interest is too low may lead to an unfavorable benefits to harms tradeoff and cost-effectiveness ratio. Some screening recommendations also incorporate family history. For example, with colonoscopy screening for colorectal cancer (CRC), an earlier start age and more frequent colonoscopy is recommended for those with a first-degree relative diagnosed with CRC under age 60, or 2+ first-degree relatives diagnosed with CRC<sup>29</sup>.

The primary exception to only using age and sex, and family history for choosing who should be screened is lung cancer screening with low-dose CT, where most guidelines recommend screening only for those at high risk due to heavy smoking history (30+ pack years and current smoker or recent quitter)<sup>30–31</sup>. Figure 2 illustrates the power of targeted screening in this setting to increase the efficiency of screening by selecting for higher risk subjects. Although screening is more efficient in the targeted population, the reach of screening in the overall population will be diminished with targeting, as will the overall proportional reduction in cancer mortality achievable.

Several studies examined targeting for bladder cancer screening based on age, sex and smoking history. Since smoking is less strong a risk factor for bladder than for lung cancer, the degree of risk concentration achievable with bladder cancer is less than that for lung cancer<sup>32–33</sup>.

## 7. Current Recommended Cancer Screening Tests in North America

Table 2 displays all of the recommendations for cancer screening modalities from the United States Preventive Services Task Force (USPSTF)<sup>34</sup>. There are only two cancers for which screening received the top A recommendation; not coincidentally, these are the two cancers

where screening reduces incidence as well as mortality, namely cervical and colorectal. An additional two cancers received a B recommendation for screening, breast cancer with mammography and lung cancer with low-dose CT. Screening for several cancers received an I rating, meaning there is insufficient evidence to judge whether screening is beneficial. The recommendation for PSA-based screening for prostate cancer was changed to a D (from a C) in 2012. Screening for ovarian, pancreatic and testicular cancer also received a D recommendation.

In Canada, the Canadian Task Force on Preventive Health Care issues recommendations for cancer screening<sup>35</sup>. Cervical cancer screening with Pap smears is recommended for age 30–69 (strong recommendation) and 25–29 (weak recommendation). Mammography is recommended every 2–3 years for women 50–74 (weak recommendation). For colorectal cancer, there is a recommendation from 2001 for fecal occult blood testing (strong) and flexible sigmoidoscopy (weak); new guidelines are expected soon that will cover colonoscopy. For prostate cancer, there is a recommendation not to undergo screening. Lung cancer screening guidelines with low-dose CT are also forthcoming.

## References

1. Habbema D, de Kok I, Brown ML. Cervical cancer screening in the United States and the Netherlands: A tale of two countries. *Milbank Quarterly*. 2012; 90:5–37. [PubMed: 22428690]
2. National Lung Screening Trial Research Team. Results of initial low-dose computed tomographic screening for lung cancer. *New Engl J Med*. 2013; 368:1980–1991. [PubMed: 23697514]
3. Pisano ED, Gatsonis G, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *New Engl J Med*. 2005; 353:1773–1783. [PubMed: 16169887]
4. Buys SS, Partridge E, Black A, et al. Effect of screening on ovarian cancer mortality: The Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *J Amer Med Assoc*. 2011; 305:2295–2302.
5. Whitlock EP, Lin JS, Liles E, et al. Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. *Ann Int Med*. 2008; 149:638–658. [PubMed: 18838718]
6. Croswell JM, Ransohoff DF, Kramer BS. Principles of cancer screening: lessons from history and study design issues. *Semin Oncol*. 2010; 37:202–215. [PubMed: 20709205]
7. U.S. Preventive Services Task Force. Screening for breast cancer: U. S. Preventive Services Task Force Recommendation Statement. *Annals Int Med*. 2009; 151:1716–1726.
8. Pinsky PF, Church T, Izmirlian G, Kramer BS. The National Lung Screening Trial: Results stratified by demographics, smoking history and lung cancer histology. *Cancer*. 2013; 119:3976–3983. [PubMed: 24037918]
9. Schroder FH, Hugosson H, Roobol MJ, et al. Screening and prostate cancer mortality: results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. *Lancet*. 2014; 384:2027–2035. [PubMed: 25108889]
10. Andriole GL, Crawford ED, Grubb R, et al. Prostate cancer screening in the randomized Prostate, Lung, Colorectal and Ovarian Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst*. 2012; 104:125–132. [PubMed: 22228146]
11. Schoen RE, Pinsky PF, Weissfeld JL, et al. Colorectal-cancer incidence and mortality with screening flexible sigmoidoscopy. *New England Journal of Medicine*. 2011; 366(25):2345–2357. [PubMed: 22612596]
12. Atkin WS, Edwards R, Kralj-Hans I, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet*. 2010; 375:1624–1633. [PubMed: 20430429]

13. Segnan N, Armaroli P, Bonelli L, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial – SCORE. *J Natl Cancer Inst.* 2011; 103:1310–1322. [PubMed: 21852264]
14. Holme O, Loberg M, Kalager M, et al. Effect of flexible sigmoidoscopy screening on colorectal cancer incidence and mortality: a randomized trial. *J Amer Med Assoc.* 2014; 312:606–615.
15. Rembold CM. Number needed to screen: development of a statistic for disease screening. *British Medical Journal.* 1998; 317:307–312. [PubMed: 9685274]
16. SEER – National Cancer Institute. [Accessed Jan 15th, 2015] SEER Statistical Summaries. [www.seer.cancer.gov](http://www.seer.cancer.gov).
17. Berry DA, Cronin KA, Plevritis SK, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. *New Engl J Med.* 2005; 353:1784–1792. [PubMed: 16251534]
18. Etzioni R, Legler JM, Feuer EJ, et al. Cancer surveillance series: interpreting trends in prostate cancer – part III: Quantifying the link between population prostate-specific antigen testing and recent declines in prostate cancer mortality. *J Natl Cancer Inst.* 1999; 16:1033–1039. [PubMed: 10379966]
19. De Gelder R, Draisma G, Heijnsdijk E, et al. Population-based mammography screening below age 50: balancing radiation-induced vs prevented breast cancer deaths. *Br J Cancer.* 2011; 104:1214–1220. [PubMed: 21364575]
20. Brenner DJ. Radiation risks potentially associated with low-dose CT screening of adult smokers for lung cancer. *Radiology.* 2004; 231:440–445. [PubMed: 15128988]
21. Bielawska B, Day AG, Lieberman DA, Hookey LC. Risk factors for early colonoscopic perforation include non-gastroenterologist endoscopists: a multivariable analysis. *Clin Gastro Hepatol.* 2014; 12:85–92.
22. Rutter CM, Johnson E, Miglioretti DL, et al. Adverse events after screening and follow-up colonoscopy. *Cancer Causes Control.* 2012; 23:289–296. [PubMed: 22105578]
23. Byrne MM, Weissfeld J, Roberts MS. Anxiety, fear of cancer, and perceived risk of cancer following lung cancer screening. *Med Decis Making.* 2008; 28:917–925. [PubMed: 18725404]
24. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New Engl J Med.* 2011; 365:395–409. [PubMed: 21714641]
25. Pinsky PF, Parnes HL, Andriole G. 1. Mortality and Complications Following Prostate Biopsy in the PLCO Cancer Screening. *Trial.Br J Urol Int.* 2014; 113:254–259.
26. Biesheuvel C, Barratt A, Howard K, et al. Effects of study methods and biases on estimates of invasive breast cancer over-detection with mammography screening. *Lancet Oncol.* 2007; 8:1129–1138. [PubMed: 18054882]
27. Sharaf RN, Ladabaum U. Comparative effectiveness and cost-effectiveness of screening colonoscopy vs. sigmoidoscopy and alternative strategies. *Am J Gastroenterol.* 2013; 108:120–132. [PubMed: 23247579]
28. Stout NK, Lee SJ, Schechter CB, et al. Benefits, harms and costs for breast cancer screening after US implementation of digital mammography. *J National Cancer Inst.* 2014; 106
29. Rex DK, Johnson DA, Anderson JC, Schoenfeld PS, Burke CA, Inadomi JM. American College of Gastroenterology guidelines for colorectal cancer screening 2008. *Am J Gastro.* 2009; 104:739–750.
30. U.S. Preventive Services Task Force. Screening for lung cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals Int Med.* 2014; 160:330–338.
31. Pinsky PF, Berg C. Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered. *J Med Screen.* 2012; 19:154–156. [PubMed: 23060474]
32. Krabbe LM, Svatek RS, Shariat SF, et al. Bladder cancer risk: Use of the PLCO and NLST to identify a suitable screening cohort.
33. Mir MC, Stephenson AJ, Grubb R, et al. Predicting risk of bladder cancer using clinical and demographic information from prostate, lung colorectal and ovarian cancer screening trial participants. *Cancer Epidemiology Biomarkers Prevention.* 2013; 22:2241–2249.
34. United States Preventive Services Task Force. Recommendations for primary care practice. <http://www.uspreventiveservicestaskforce.org>.

35. Canadian Task Force on Preventive Health Care. Canadian Task Force on Preventive Health Care guidelines. <http://canadiantaskforce.ca/ctfphc-guidelines>.

Author Manuscript

Author Manuscript

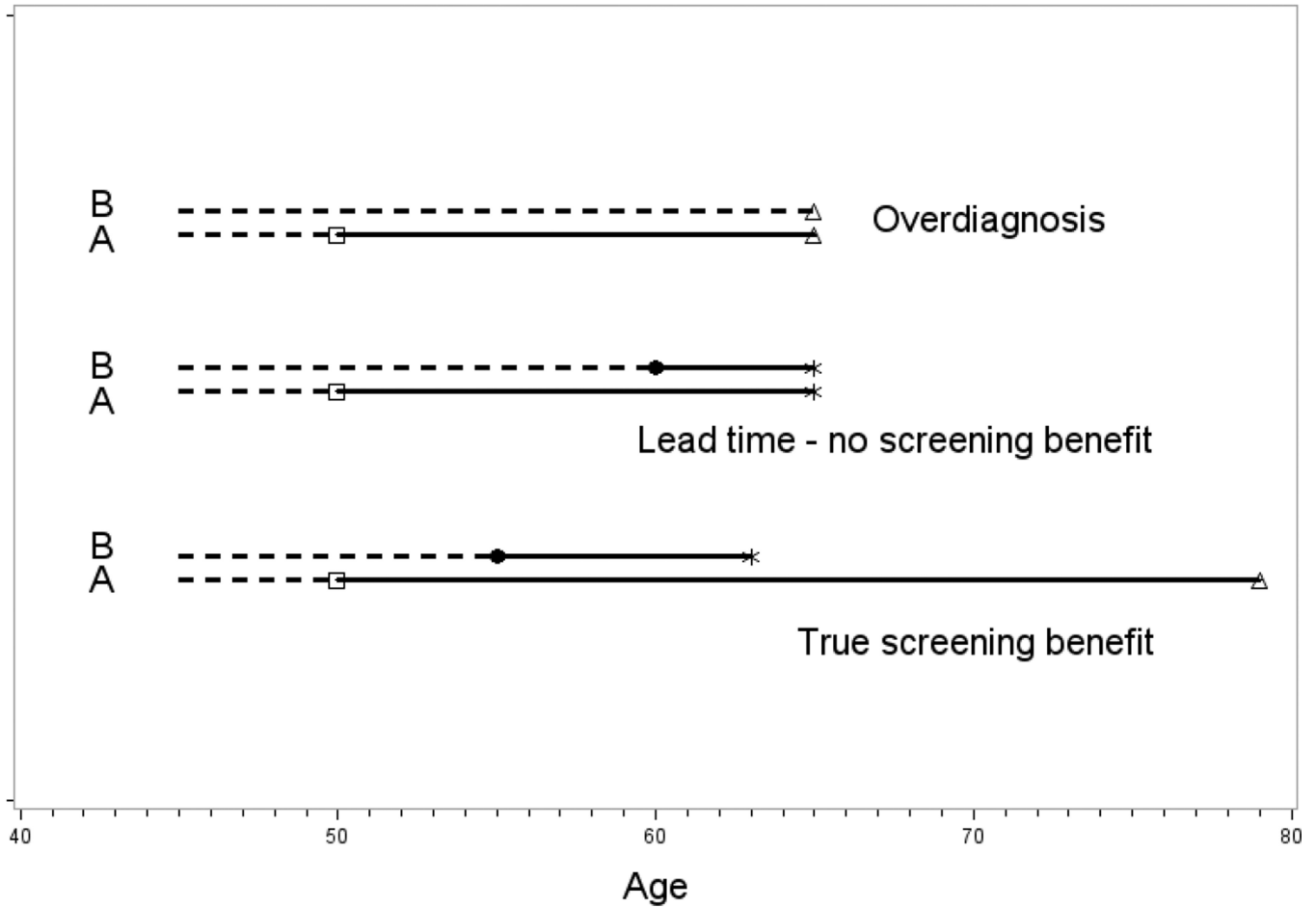
Author Manuscript

Author Manuscript

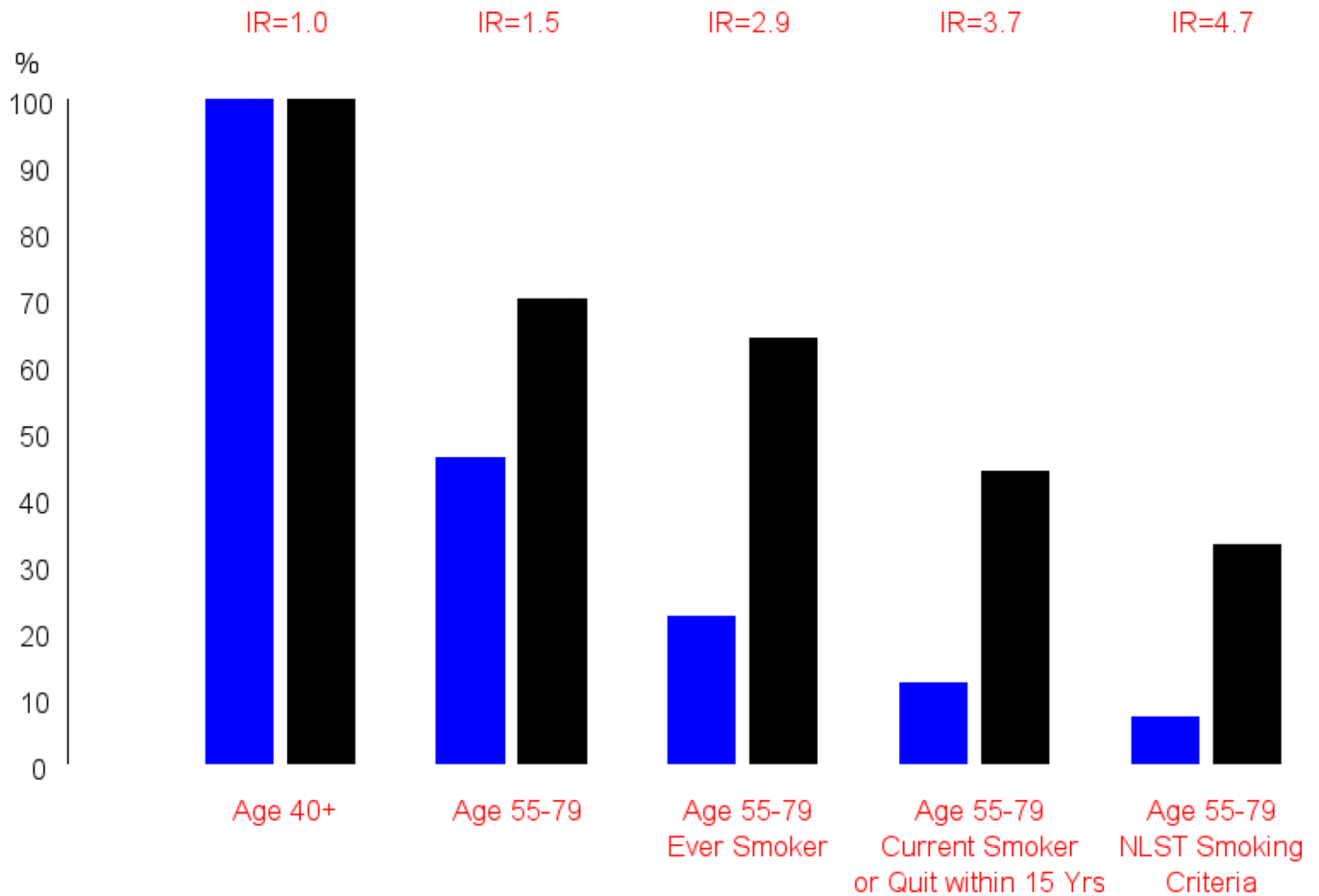
### Key Points

- Early detection of cancer through screening can reduce cancer mortality; detection of precancerous lesions, achievable currently with colorectal and cervical cancer screening, reduces cancer incidence as well
- Sensitivity and specificity are critical metrics for researchers assessing the predictive ability of a screening modality; positive predictive value (probability of cancer given a positive test) is more relevant for clinicians
- The gold standard for evaluating cancer screening tests is the randomized controlled trial (RCT). Caution must be taken when using observational data, and especially survival statistics, to assess cancer screening.
- Harms from screening include false positive tests and their downstream sequelae, including invasive diagnostic tests and complications thereof, as well as overdiagnosed and overtreated cancers
- Targeting screening to high risk subjects is a strategy to make screening more efficient, in terms of optimizing the benefits to harms tradeoff and the cost-effectiveness of screening





**Figure 1.** Schematic of Screening Scenarios. In each panel, A represents the scenario with screening at age 50 and B represents the hypothetical scenario if there had been no screening. Square and circle represent screen-detected and clinical diagnosis, respectively, dotted and solid line represent preclinical cancer and post-diagnosis cancer, respectively. Star and triangle denote death from the cancer of interest and death from other causes, respectively. In the lead time - no screening benefit scenario, “survival” increases from 5 to 15 years upon screening, but time and cause of death are unchanged.



**Figure 2.**

Targeted screening through risk stratification for low-dose CT lung cancer screening. Blue bars represent overall population (over age 40), black bars represent incident lung cancers. For the scenario of screening everyone over age 40 (left-most column), 100% of the population and 100% of lung cancers (in the 40+ population) are covered. With increasing age and/or smoking history requirements, the percentage of incident lung cancers covered decreases, but the percentage of the population screened decreases more sharply. The incidence ratio (IR) represents the ratio of incidence rates in the restricted population over that in the overall population; it is also the ratio of the height of the black versus blue bars. NLST smoking criteria is 30+ pack years and current smoking or quit within 15 years. (Adapted from Pinsky PF, Berg C. Applying the National Lung Screening Trial eligibility criteria to the US population: what percent of the population and of incident lung cancers would be covered. *J Med Screen* 2012; 19:154–156; with permission.)

**Table 1**

Common performance characteristics of screening tests

<b>Performance Characteristic</b>	<b>Definition</b>
Sensitivity	Proportion of subjects with cancer who test positive
Specificity	Proportion of subjects without cancer who test negative
Receiver-Operating Characteristic (ROC) Curve	Curve of Sensitivity at varying levels of 1-Specificity
Area under the ROC Curve (AUC)	Area below the ROC curve; 1=perfect prediction, 0.5=no predictive ability
Positive Predictive Value (PPV)	Proportion of subjects who test positive that have cancer
Negative Predictive Value (NPV)	Proportion of subjects who test negative that do not have cancer

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

## USPSTF Cancer Screening Guidelines

Cancer	Screening Modality	Updated	Population	Recommendation
Colorectal	Colonoscopy/Fecal Occult Blood Test/Flexible Sigmoidoscopy	2008	Age 50–75	A
			Age 75–85	C
			Age > 85	D
Breast	Mammography	2009	Women age 40–49	C
			Women age 50–74	B
			Women age 75+	I
Prostate	Prostate-specific antigen (PSA)	2012	Men any Age	D
Lung	Low-dose CT	2013	Age 55–80, 30+ Pack years current smoker or quit within 15 years	B
Cervical	Pap smear, HPV testing	2012	Women age 21–65	A
			Women under 21 or over 65 (w/ prior screens)	D
Ovarian	CA125, Trans-vaginal ultrasound	2012	Any Age	D
Skin	Whole-body skin examination	2009	Any Age	I
Oral cavity	Clinical examination	2013	Any Age	I
Bladder	Urinalysis for hematuria; urine cytology	2011	Any Age	I
Pancreatic	Ultrasound, abdominal palpation, serology	2004	Any Age	D
Testicular	Self or clinical examination	2011	Adolescent/Adult Men	D

Recommendation Code: A (Recommends the service – High Certainty of substantial net benefit); B(Recommends the service– High certainty of moderate benefit/ moderate certainty of moderate-substantial net benefit); C (Recommends selective offering or providing service based on professional judgment/patient preference); D(Recommends against the service); I(Insufficient evidence; balance of benefits & harms cannot be determined). (From U.S. Preventive Services Task Force. Available at: <http://www.uspreventiveservicestaskforce.org/>; with permission.)