Additional Considerations

9

The topics in this chapter are relevant to assessment of cancer screening data but did not have an obvious home in the earlier chapters of this primer. As you will see, they are quite varied in scope. Each falls in one of three categories: data interpretation, methodology, and policy.

9.1 Topics Regarding Data Interpretation

9.1.1 Number Needed to Screen

Number needed to screen, or NNS, indicates how many individuals need to be screened so that one fewer individual dies of the cancer of interest. NNS is only relevant if cancer screening reduces mortality. NNS estimates for cancer screening tests tend to be in the hundreds to thousands of individuals. For example, the NNS for lung cancer screening with low dose computed tomography (LDCT) calculated from the National Lung Screening Trial (NLST) data was 320 [1].

The first step in calculating NNS is to subtract the cause-specific mortality rate in the presence of cancer screening from the cause-specific mortality rate in the absence of cancer screening. That quantity, which is a rate, is called the absolute risk reduction, and is an indication of extent of death prevented by

P. M. Marcus, *Assessment of Cancer Screening*,
https://doi.org/10.1007/978-3-030-94577-0_9

cancer screening. NNS equals the reciprocal of the absolute risk reduction. A fictional example is presented in Table 9.1. The absolute risk reduction in that table is 20 per 1000 person-years. The NNS is 1000/20, or 50.

NNS is calculated assuming that the only factor that contributes to the difference in mortality is cancer screening. It is best to use data from randomized controlled trials (RCTs), as data that come from other sources could reflect confounders of the screening/cause-specific mortality relationship.

9.1.2 Generalizability of Results

Generalizability refers to the applicability of results from a study, experimental or observational, to groups other than the study participants. Issues of generalizability are what drive the need to assess effectiveness. A cancer screening test may be efficacious in an RCT, but its ability to be effective in a community setting is not guaranteed by that finding.

Most cancer screening guidelines are based on findings of RCTs. Because cancer screening RCTs are long, large, and expensive undertakings, few are done. Not surprisingly, the urge to take the results of an RCT conducted in one population and apply them to another population is strong. The populations at hand could be dissimilar regions of one country, two countries in the same part of the world with different health care systems, or two countries far away from one another with dramatically different cultural norms.

Table 9.1 Calculating number needed to screen (NNS)

	Person-years (PY)	Number who die of the cancer of interest	Cause-specific mortality rate
Screened	10,000	100	10 per 1000 PY
Unscreened	15,000	450	30 per 1000 PY
NNS calculations: 30 per 1000 PY - 10 per 1000 PY = 20 per 1000 PY; 1000/20 = 50 NNS is 50			

Data are fictional

It should not be assumed that a beneficial effect of cancer screening seen in one population will be replicated in another population if the two populations have different risk factor profiles. An example is lung cancer screening: the cancer screening process may not confer the same magnitude of benefit in asbestos workers, say, as it does in cigarette smokers. It is not wise to extrapolate results from one population to another if the two populations have different clinical practices, clinical resources, and access to health care. Low and middle income countries have begun to establish cancer screening programs based on experience in high income countries, yet differences in medical resources, access to transportation, and rurality may not allow easy, frequent, or productive visits to cancer screening or treatment centers. Cultural norms also may impact cancer screening uptake and cancer treatment choices.

The assumption that a null effect of cancer screening is generalizable from one population to another also can be unwise. A region with a preponderance of late-stage, untreatable cancers may benefit from cancer screening, whereas the same cancer screening practice may have little to no impact in a region where most patients have earlier stage disease for which treatment is available.

Studies done in regions assumed to be similar enough to produce comparable findings can and have produced conflicting results. The phenomenon has been observed in breast cancer screening, but the best example comes from prostate cancer screening. There are two notable RCTs of prostate cancer screening: the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial (PLCO) [2] and the European Randomized Study of Screening for Prostate Cancer (ERSPC) [3]. PLCO, an RCT done in the US, found no reduction in prostate cancer mortality, while ERSPC, an RCT done in many countries in Europe, did. The two studies employed different cancer screening protocols, which may explain, at least in part, the discordant findings. Nevertheless, discussions regarding the conflicting results have focused on contamination in PLCO's control arm and likely inferior prostate cancer treatment in ERSPC's control arm. Random variation or a

systematic difference (that is, the contamination and treatment issues) may very well be responsible, but it also is necessary to consider the possibility that prostate cancer screening may be of benefit in one region but not the other.

9.1.3 Concurrent Changes in Treatment

Cancer screening does not operate in a vacuum. While cancer screening tests are under investigation, disseminating, or their use reaches a steady state, changes in clinical practice are occurring as well. Advances have led to a better understanding of tumor composition, which in turn have led to new and highly effective therapies for some tumors. Cures are possible today that were not possible 20 years ago. This situation begs this question: if cancer treatment has improved, especially at regional and distant stages, is screen detection at an early stage still necessary?

In the presence of concurrent changes in treatment, an RCT can still evaluate whether cancer screening is of benefit as long as individuals in both arms have access to the same treatments. Concurrent changes do present a problem in time trend studies; it is impossible to know whether reductions in cancer mortality are due to uptake of a new cancer screening regimen or availability of a new treatment.

An RCT to determine whether a cancer screening test affects a benefit cannot be established each time a shift in clinical practice occurs. Creative use of available data can shed some light, however. The ecologic study of Autier et al. [4], mentioned in Chap. 7, examined the issue of concurrent changes in breast cancer screening uptake and treatment by examining time trends for three pairs of regions in Europe. Each region in a pair had similar access to breast cancer treatment yet a different date of widespread mammography adoption. While not without limitations, that analysis suggests that recent reductions in breast cancer mortality are not overwhelmingly due to cancer screening.

9.2 Topics Regarding Methodology

9.2.1 Microsimulation Modeling

Microsimulation modeling of cancer screening is a technique in which computer-generated (fictional) life histories are manipulated by applying assumptions about factors that affect cancer screening outcomes. Models produce outcomes, such as cause-specific mortality, for a variety of assumptions and cancer screening scenarios, providing insight into benefits and harms of cancer screening. The National Cancer Institute's (NCI's) Cancer Intervention and Surveillance Modeling Network (CISNET) initiative has taken the lead in microsimulation modeling for cancer screening [5].

Microsimulation modeling is possible given unprecedented improvements in computational power in recent years. The use of microsimulation modeling in lieu of establishing RCTs has been suggested, because RCTs cannot address every proposed cancer screening strategy. Microsimulation modeling is arguably most valuable when done in conjunction with data from population-level databases, completed RCTs, or large, well-conducted prospective cohort studies, as certain assumptions needed to generate life histories can be based on real-life experience.

No microsimulation model will perfectly replicate reality. However, these models have become a popular and useful tool to investigate "what if" situations. Results from CISNET models, in conjunction with RCT and cohort data, are now used by the United States Preventive Services Task Force [6] when developing cancer screening recommendations .

9.2.2 Magnitude of Overdiagnosis

The excess incidence method was presented in Chap. 6 as a way to calculate the degree of overdiagnosis in an RCT, but it is not the only method available. Some methods employ assumptions about the distribution of lead time [7], while others compare changes in

incidence that have occurred over time, generally in conjunction with other factors [8, 9]. Statistical modeling, including microsimulation modeling, has been utilized in the effort to determine the magnitude of overdiagnosis or a range of plausible magnitudes.

There has been heated discussion as to which method will produce the correct answer. That assumes, of course, that there is one correct answer. But overdiagnosis only exists in the context of cancer screening, and therefore, the magnitude of overdiagnosis is a function of aspects of the cancer screening regimen, including test, screening interval, compliance, and those who are screened. Magnitude also is a function of the intensity of diagnostic evaluation that follows a positive test. There is no one correct answer; there are many correct answers, with each dependent on many factors.

The desire to quantify the magnitude of overdiagnosis is related to the desire to weigh the benefits and harms of cancer screening, something that is most easily done when a single number can be attached to each. In lieu of a single number, a range of plausible measures of overdiagnosis can be used in sensitivity analyses.

9.2.3 Incidence and Prevalence Screens

When discussing burden of disease, the terms prevalence and incidence refer to disease that is existing and new, respectively. The terms prevalence and incidence are sometimes used in cancer screening to describe the initial and later screens, respectively, performed as part of a cancer screening program or an RCT. The initial screen is expected to lead primarily to detection of cancers that have stalled in Phase B, while incidence screens are expected to lead primarily to detection of cancers that have moved into Phase B since the last cancer screening test. All other things being equal, the yield on prevalence screens is expected to be higher than the yield on incidence screens. Also, the prognosis for cancers detected on the prevalence screen is expected to be more favorable than for those detected on incidence screens.

9.2.4 Interval Cancers

Interval cancers often are considered failings of cancer screening, even though cancer screening is not designed or expected to lead to detection of every Phase B cancer. Some conditions that lead to interval cancers, for example, errors in test interpretation and missed screens, may be addressable, but it is unrealistic to believe that interval cancers can be eliminated. Interval cancers are a reminder of the limits of cancer screening.

Cancer can be detected serendipitously, meaning that an unrelated diagnostic medical test or procedure inadvertently finds an abnormality that is suspicious for cancer. An MRI performed to investigate back pain could identify a colonic mass, for example. Whether serendipitously detected cancers are interval cancers is open to debate. They do not arise from symptoms but they may have been missed on the previous organ-specific cancer screening test.

9.3 Topics Regarding Policy

9.3.1 Selecting a Cancer Screening Interval

The phrase cancer screening interval refers to the time between screens. Though the choice of the screening interval should be based exclusively on the average length of Phase B and how variable it can be, historically, is has not. It is only recently that screening intervals have started to reflect the natural history of cancer. In the past, screening intervals were typically 1 year, probably because cancer screening was associated with the practice of having an annual physical.

The choice of screening interval will impact effectiveness and the magnitude of harms. It also will drive costs and availability of health care resources. Ideally, these factors are weighed in conjunction with knowledge of the natural history of cancer to arrive at a screening interval that affords benefit but does not strain a health care system.

9.3.2 De-implementation

De-implementation refers to the reduction or cessation of a service provided by health care practitioners. Calls for de-implementation may be made when practices do not benefit patients, including when they are harmful or wasteful. The need for de-implementation may arise in the instance of adoption of a practice whose benefit is uncertain, or if a practice observed to be efficacious is not effective. A well-known instance of de-implementation is the reduction in prescribing of postmenopausal hormone therapy after users experienced an increase in breast cancer risk [10].

De-implementation has been discussed in the context of cancer screening for a number of reasons. Some cancer screening tests have become widely adopted in clinical practice without strong or direct evidence that their use reduces cause-specific mortality; some also have been adopted without complete understanding of the harms they cause. A notable example of the former is thyroid cancer screening. Low-cost ultrasound thyroid cancer screening became available in South Korea in the 1990's even though the practice had never been evaluated in an RCT. Thyroid cancer incidence increased 15-fold from 1993 to 2011, although no change in thyroid cancer mortality occurred concurrently. In 2015, the Korean Committee for National Cancer Screening Guidelines issued a recommendation against thyroid cancer screening with ultrasonography for healthy individuals [11, 12].

De-implementation will result in reversal of the effects on intermediate outcomes described in Chap. 5. Incidence of invasive cancer (in the case of cancer screening that detects only invasive disease) and case survival will decrease, and assuming all else remains the same, should approach their pre-screening levels. The number of early stage cancers should decrease due to elimination of overdiagnosis. The number of late stage cancers will not change if cancer screening did not result in down staging, and will increase if it did.

As it is for implementation, it is critical to track the changes in both intermediate and definitive outcomes during a period of cancer screening de-implementation. Both implementation and de-implementation are by necessity based on certain assumptions; therefore, the impact cannot be predicted. It is particularly important to watch for unexpected consequences, be they favorable or deleterious.

9.3.3 Reduction in Advanced-Stage Cancer

A reduction in advanced-stage cancer, usually distant cancer, has been suggested as a surrogate for cause-specific mortality. The push to use advanced-stage cancer has to do, at least in part, with the desire to obtain answers regarding the impact of cancer screening without having to wait for a cause-specific mortality outcome. A reduction in the number of distant-stage cancers may be the best of the intermediate cancer screening outcomes in terms of correlation with reductions in cause-specific mortality, but it still does not reflect experience after diagnosis and does not measure how cancer screening alters length of life.

Legitimate use of a reduction in distant-stage cancers as what is, in effect, a definitive endpoint requires that those cancers are fatal, and often they are. It also assumes that non-distant-stage cancers have a better prognosis, which in most situations they do. Yet consider a cancer that, in the absence of cancer screening, would be diagnosed at a distant stage, but in the presence of cancer screening, is diagnosed at a regional stage. If the prognosis for regional stage cancer is the same as that of distant-stage cancer, no reduction in cause-specific mortality would occur even though the number of distant-stage cancers has decreased.

If the day comes when cancer is no longer fatal even at a distant stage, the goals of cancer screening will need to be reassessed. In the meantime, the choice of distant-stage disease as a definitive endpoint must be made carefully and on a situation-by-situation basis.

9.3.4 Benefit in the Absence of a Mortality Reduction

Once upon a time there was no cancer screening in the US. When discussions regarding establishment of population-based cancer screening began in earnest, the proposed metric of benefit was a reduction in cause-specific mortality, as cancer was considered to be a life-threatening disease. Diagnoses often occurred at late stages and few, if any, effective treatments were available once cancer spread beyond the organ of origin.

The first breast and colorectal cancer screening tests to become established in the US were shown to reduce cause-specific mortality in at least one RCT. Those tests, film-screen mammography and guaiac-based fecal occult blood testing, have since been replaced with tests that are more technologically advanced: digital mammography and breast tomosynthesis, and fecal immunochemical testing, flexible sigmoidoscopy, and colonoscopy. Yet none of the replacement tests was vetted in a study that assessed cause-specific mortality prior to adoption.

When replacement tests are adopted, it is done so under the assumption that the new test will confer the same or a greater reduction in cause-specific mortality as the test it is replacing. The replacement tests also have a characteristic that make them more desirable than the test they are replacing. They may have better performance measures, such as lower false positive rates, or they may be more acceptable to patients. They could be less expensive when all components of the screening process are considered.

In my opinion, future cancer screening tests that target an organ for which no efficacious screening test exists only should be implemented in clinical practice when high-level evidence is available to support a reduction in cause-specific mortality. Others may feel differently. Some have argued that a shift to a stage at diagnosis that is simpler to treat is benefit enough, though the consequences that come with a cancer diagnosis earlier in time must not be ignored. Those include intense surveillance regimens, chemoprevention strategies, and psychological challenges for periods of time that are longer than those that would have occurred if cancer had been diagnosed later.

Whether it is appropriate to adopt replacement tests in clinical practice without formal vetting using a cause-specific mortality endpoint or another measure of the benefit to harm is a matter of the cancer at hand and differences in the replacement and original test. There are some instances in which a strong argument can and have been made for adoption without full knowledge about the impact on benefits and harms. Data are available to retrospectively support some of the decisions made regarding replacement, including the choice to adopt colonoscopy screening for colorectal cancer.

References

1. The National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011;365(5):395–409.
2. Andriole GL, Crawford ED, Grubb RL 3rd, Buys SS, Chia D, Church TR, Fouad MN, Isaacs C, Kvale PA, Reding DJ, Weissfeld JL, Yokochi LA, O'Brien B, Ragard LR, Clapp JD, Rathmell JM, Riley TL, Hsing AW, Izmirlian G, Pinsky PF, Kramer BS, Miller AB, Gohagan JK, Prorok PC, PLCO Project Team. Prostate cancer screening in the randomized prostate, lung, colorectal, and ovarian cancer screening trial: mortality results after 13 years of follow-up. J Natl Cancer Inst. 2012;104(2):125–32.
3. Hugosson J, Roobol MJ, Månsson M, Tammela TLJ, Zappa M, Nelen V, Kwiatkowski M, Lujan M, Carlsson SV, Talala KM, Lilja H, Denis LJ, Recker F, Paez A, Puliti D, Villers A, Rebillard X, Kilpeläinen TP, Stenman UH, Godtman RA, Stinesen Kollberg K, Moss SM, Kujala P, Taari K, Huber A, van der Kwast T, Heijnsdijk EA, Bangma C, De Koning HJ, Schröder FH, Auvinen A, ERSPC Investigators. A 16-yr follow-up of the European randomized study of screening for prostate cancer. Eur Urol. 2019;76(1):43–51.
4. Autier P, Boniol M, Gavin A, Vatten LJ. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. BMJ. 2011;343:d4411.
5. Cancer Intervention and Surveillance Modeling Network. Overview of CISNET Modeling [Internet]. Bethesda, MD: National Cancer Institute [cited 2019 Oct 24]; [about 4 screens] Available from: https://resources.cisnet.cancer.gov/registry/learn/.
6. Modeling Report: Lung Cancer: Screening. [updated May 2019; cited 2019 Oct 24]. Available from: https://www.uspreventiveservicestaskforce.org/Page/Document/modeling-report/lung-cancer-screening.

7. Ripping TM, Ten Haaf K, Verbeek ALM, van Ravesteyn NT, Broeders MJM. Quantifying overdiagnosis in cancer screening: a systematic review to evaluate the methodology. J Natl Cancer Inst. 2017;109(10):djx060.
8. Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. BMJ. 2015;350:g7773.
9. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-cancer tumor size, overdiagnosis, and mammography screening effectiveness. N Engl J Med. 2016;375(15):1438–47.
10. Steinkellner AR, Denison SE, Eldridge SL, Lenzi LL, Chen W, Bowlin SJ. A decade of postmenopausal hormone therapy prescribing in the United States: long-term effects of the Women's Health Initiative. Menopause. 2012;19(6):616–21.
11. Ahn HS, Kim HJ, Kim KH, Lee YS, Han SJ, Kim Y, Ko MJ, Brito JP. Thyroid cancer screening in South Korea increases detection of papillary cancers with no impact on other subtypes or thyroid cancer mortality. Thyroid. 2016;26(11):1535–40.
12. Ahn HS, Welch HG. South Korea's thyroid-cancer "epidemic" – turning the tide. N Engl J Med. 2015;373(24):2389–90.