



Multicancer Early Detection: Learning From the Past to Meet the Future

Ruth Etzioni , PhD,^{1,2,*} Roman Gulati , MS,¹ Noel S. Weiss, MD, DrPH^{1,3}

¹Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; ²Center for Early Detection Advanced Research, Knight Cancer Institute, Portland, OR, USA, and ³Department of Epidemiology, University of Washington, Seattle, WA, USA

*Correspondence to: Ruth Etzioni, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, M2-B230, Seattle, WA 98109-1024, USA (e-mail: retzioni@fredhutch.org).

Abstract

Multicancer early detection (MCED) tests may soon be available to screen for many cancers using a single blood test, yet little is known about these tests beyond their diagnostic performance. Taking lessons from the history of cancer early detection, we highlight 3 factors that influence how performance of early detection tests translates into benefit and benefit-harm trade-offs: the ability to readily confirm a cancer signal, the population testing strategy, and the natural histories of the targeted cancers. We explain why critical gaps in our current knowledge about each factor prevent reliably projecting the expected clinical impact of MCED testing at this point in time. Our goal is to communicate how much uncertainty there is about the possible effects of MCED tests on population health so that patients, providers, regulatory agencies, and the public are well informed about what is reasonable to expect from this potentially important technological advance. We also urge the community to invest in a coordinated effort to collect data on MCED test dissemination and outcomes so that these can be tracked and studied while the tests are rigorously evaluated for benefit, harm, and cost.

Multicancer early detection (MCED) tests are on the cusp of entering the marketplace. These tests are designed to screen for a basket of cancers with 1 blood draw by interrogating features of circulating cell-free DNA: mutation profiles, fragmentation patterns, and methylation signatures. MCED tests represent a major technological advance toward realizing a vision that cancer screening can be achieved with a simple blood test. This vision has motivated a decades-long search for cancer biomarkers, producing blood tests such as cancer antigen 125 for ovarian cancer and prostate-specific antigen (PSA) for prostate cancer, with some notable successes, some disappointments, and plenty of controversy.

Current knowledge about MCED tests is focused on diagnostic performance. Retrospective studies have evaluated test sensitivity in known cancer cases and specificity in controls without a cancer diagnosis. Results have shown variable sensitivity depending on the test product, cancer type, and cancer stage; reported specificity has been consistently high (98%-99%). In 2018, Cohen et al. (1) reported that the CancerSEEK test's sensitivity to detect 8 types of cancer was 62% in a sample of clinically detected, nonmetastatic (stage I-III) cases, with a specificity of 99% in noncancer controls. In 2019, Cristiano et al. (2) reported that the DELFI test's sensitivity was 73% to detect 7 types of cancer in a sample of majority stage I-III cases with 98% specificity. And in 2020, Liu et al. (3) reported that the Galleri

test's sensitivity was 44% to detect more than 50 types of stage I-III cancers with 99.3% specificity. In all tests, sensitivity was lower for stage I (around 40%) compared with stage III tumors (around 80%). In addition to producing a cancer signal, several tests also label the most likely tissue of origin (TOO); in their retrospective reports, the DELFI and Galleri tests identified the TOO correctly in 75% and 93% of cancer cases, respectively.

The diagnostic performance of MCED tests suggests that they may prove to be useful. First, reported specificity considerably exceeds that of existing tests, such as mammograms and PSA tests. It has long been recognized that high specificity is critical in cancer screening to control the number of false-positive tests. In general, the specificity of an MCED test will vary with the number of cancers targeted by the test; Liu et al. (3) reported a specificity of 99.3% for a test targeting more than 50 cancers, but this is likely higher than the specificity for their prespecified subset of 12 cancers. Second, although sensitivity for early stage disease is limited, these tests appear able to detect a variety of cancers, including some lethal cancers for which screening tests do not currently exist. These advantages stand to alter the way we diagnose these cancers and, if they can translate into positive clinical impact, would dramatically change the practice of early detection.

The investigation of whether diagnostic performance translates to meaningful clinical impact is typically a lengthy

endeavor. In 2002, the Early Detection Research Network of the National Cancer Institute codified the typical sequence of studies required to establish benefit into a pipeline, the Phases of Biomarker Development (PBD) (4). This pipeline labels retrospective diagnostic performance studies as early phase (phase 2) and randomized screening trials as the ultimate phase (phase 5) before a test can be judged beneficial. From a PBD perspective, MCEd research is indeed still early phase. We do not yet know what the clinical impact of the technology will be, nor are we yet able to assess whether it will produce a sustainable benefit. Furthermore, even if benefit can be established, a central lesson of single-cancer screening is that the value of early detection can be a double-edged sword. For many cancers, earlier detection and treatment is more likely to result in cure, yet for some cancers, it can also lead to overdiagnosis and overtreatment (5). As a consequence, evaluation of both benefit and harm of a novel early detection technology is required before it can be recommended for general use in the population.

In this commentary, we highlight lessons learned from the history of cancer early detection research about 3 factors impacting how diagnostic performance of a screening test translates to benefit and harm: the ability to readily confirm a cancer signal, the population testing strategy, and knowledge of the natural histories of the targeted cancers. We make the case that knowledge gaps surrounding each of these factors imply that any projection of the likely population impact of MCEd tests based on information published to date is highly uncertain and not reliable. We argue that greater awareness is needed of the complexities involved in translating from diagnostic performance to population benefit and harm if we want to ensure that patients, providers, and the public are well informed about what they can expect from MCEd technology.

The Ability to Readily Confirm a Cancer Signal

The ability to readily confirm a cancer signal is critical for promptly triaging patients with true malignancies vs those with spurious signals. Many studies have highlighted the anxiety that surrounds a positive test result and its impact on quality of life. For single-cancer screening tests, disease confirmation most commonly occurs via image-guided biopsy. Even in this setting, disease confirmation is not without its complications. The imaging test is frequently imperfect, a biopsy may be highly invasive, and a negative biopsy result in the presence of a persistently positive screening result may lead to an odyssey of confirmation testing. In ovarian cancer, for example, imaging of the ovaries is used to triage women to biopsy, but most aggressive ovarian tumors originate in the fallopian tube, which is difficult to image, and any biopsy requires a surgical procedure. In prostate cancer, repeated negative 6-core biopsies in the presence of persistently elevated PSA tests led to increased sampling with 12 and even 20 or more biopsy cores.

For multicancer tests, the problem of confirming the test result is compounded. If the test does not label the TOO, then whole-body scanning might be done, with its own sensitivity limitations and the risk of identifying incidental benign lesions. If the test labels the TOO but no tumor is visualized, this may be indicative that no cancer is present. However, it could also be a false-positive result. In practice, the patient must contend with a dilemma: whether to reevaluate the same TOO, move down the list of likely sites of the tumor produced by the test, or consider the test to be a false positive (which does not automatically imply that no cancer is present, only that none of the cancers targeted by the test are

present). And how to do this optimally—whether to proceed according to the test's ordering, or disease prevalence, or personal risk (eg, due to family history of a specific cancer)—could substantially impact benefit, harm, and cost.

In the absence of any guidance as to the most efficient procedure for confirming a positive MCEd test result, we are almost completely in the dark about how often the test might lead to unnecessary confirmation imaging exams and biopsies. Societal values about how much confirmation testing is tolerable—and how to follow patients with positive results but no confirmed cancer—are likely to evolve in tandem with protocols for MCEd test workup. We also do not know the extent to which the advanced imaging prompted by the test might reveal incidental lesions that are not cancer with their own demands for evaluation and further care. In the National Lung Screening Trial (6), the percentage of all screening tests that identified abnormalities not suspicious for lung cancer was more than 3 times as high in the group that received low-dose computed tomography screening as in the comparison group that received chest radiography. Beyond the test itself, the burden of confirming MCEd test results could turn out to be the costliest aspect of population MCEd screening.

The Population Testing Strategy

The population testing strategy also greatly impacts health outcomes. Even for single-cancer screening tests, the same test offered over different age ranges or at different intervals can produce quite different benefit-harm profiles. In prostate cancer screening, for example, we have projected that just extending the stopping age from 69 to 74 years increases lives saved by almost one-third, but the frequency of overdiagnosis is doubled unless higher PSA thresholds for biopsy referral are used for older men (7). In risk-targeted screening, the definition of what constitutes the high-risk, screening-eligible population can be similarly influential. New US Preventive Services Task Force guidelines for lung cancer screening expand eligibility from 30 to 20 or more pack-years and lower the starting age based on a projected 40% increase in lives saved under the expanded guidance (8,9).

Determining how to deploy a cancer screening test is a complex undertaking that must be tailored to disease dynamics in the population. Because disease features such as prevalence and natural history differ across cancers, strategies that sustainably balance benefit and harm differ as well. Thus, the interval recommended for colonoscopy is 10 years for average-risk individuals (10) because this cancer has a well-defined precursor lesion that takes many years to develop into a tumor, and much of the benefit of colonoscopy is attributable to the preventive effect of detecting and removing this lesion. In contrast, the recommended interval for mammography screening is 1 to 2 years (11) because the average preclinical latency of invasive breast tumors is much shorter—approximately 2 to 4 years according to estimates based on breast screening trials (12).

Determining how to deploy MCEd tests is more complex because the various target cancers may have different preferred strategies. If an MCEd test is offered annually, the harm-benefit trade-off will be very different than if it is offered every few years. Even if we knew for certain that the best way to utilize MCEd tests is as a complement to existing screening tests, determining a preferred strategy would require much more information about the natural history of the target cancers than is currently known.

Knowledge of the Natural Histories of the Targeted Cancers

Knowledge of the natural histories of the targeted cancers—when these cancers start and how quickly they progress in the absence of any intervention—is limited mostly to those cancers for which screening tests currently exist. Indeed, it is the availability of data on how screening alters disease incidence and stage at diagnosis that permits identification of the natural history. Without such data, we cannot decouple preclinical onset from preclinical latency; for a given patient, all we know is the sum of the 2. We have estimates of tumor latency for breast (12), lung (13), and prostate cancers (14), but we lack estimates for cancers not currently screened for, such as pancreatic and liver cancers. Further, little is known about the fraction of the latent period representing early stage disease. The window of opportunity for MCED tests to detect these cancers early is therefore also uncertain. Even if an MCED test shows high sensitivity to detect a certain cancer at an early stage in a retrospective study, the opportunity for that test to detect the cancer at an early stage under a prospective screening protocol will be lower when the early stage duration is shorter.

Given the gaps in our knowledge about natural history for most cancers included in MCED tests, we cannot reliably project how often the tests will successfully shift cases from advanced to early stage at diagnosis. Clarke et al. (15) projected the implied cancer mortality reduction from redistributing all stage IV cases equally across stages I-III for a range of cancers to be 24% assuming that cases shifted to an earlier stage by screening would receive a corresponding shift in disease-specific survival. We believe that this projection is likely to be optimistic.

In prior randomized trials of cancer screening, no modality has reduced the incidence of late-stage disease by even 50%. Across breast cancer screening trials, the median reduction in late-stage incidence was 15% (16). In the European Randomized Study of Screening for Prostate Cancer, the reduction was 30% (17), and in the National Lung Screening Trial, the reduction in stage IV incidence was 25% (6). In a modified calculation of Clarke et al. (15), if only 25% of stage IV cancers are equally redistributed across stages I-III, then the cancer-specific mortality reduction will be closer to 6% than 24%. In practice, the redistribution may not be equal across stages, and there may be additional shifts (eg, from stage III to stages I and II), but it is difficult to predict the likelihood of these outcomes.

Our discussion of the factors that impact how screening test performance translates into clinical outcomes suggests that, at present, the implications of published MCED test performance for population screening benefit are highly uncertain. Ongoing prospective studies, which are using these tests to screen large numbers of individuals in the target population, will provide important data regarding the ability of MCED tests to detect latent disease, the diagnostic testing pathways following a positive test, and the frequency of unnecessary confirmation tests. But even if these studies show that the tests can detect some targeted cancers before they would have been diagnosed clinically, this alone will not be a guarantee of adequate population benefit or sustainable benefit-harm trade-offs. With these single-arm prospective studies, we can know how many cancers are found by the test, what type, and what stage, but we do not know if or when those cancers would have been found without the test. Most importantly, we cannot know whether the fate of persons with these cancers would have been different in the absence of the test.

Only unbiased comparative studies will tell us the extent to which patients detected by the test are actually being helped. Observational studies that compare mortality between screened

and unscreened persons often are unable to achieve comparability of the 2 groups with regard to their risk of disease (18). And studies that compare screening histories between persons who die as a result of a malignancy and controls from the same population cannot provide an accurate assessment of screening efficacy until the prevalence of screening in the population has stabilized (19).

There is a reason why randomized screening trials are considered to be the gold standard for evaluating cancer screening tests. In principle, these trials provide empirical validation that the screening test and disease natural history are likely to come together in a way that translates favorable diagnostic performance into clinically significant benefit, as well as permitting assessment of the harms of the test. Randomized screening trials are the ultimate phase of the Early Detection Research Network's PBD, and they are an evidence gold standard used by national policy panels such as the US Preventive Services Task Force. Yet, randomized screening trials require vast sample sizes and (for many forms of malignancy) long follow-up. Further, they can only examine a small number of possible screening strategies. And, despite being simple in concept, they tend to be complex to implement, analyze, and interpret (20).

The promise of MCED testing and the uncertainty about its population impact have created a need for randomized trials. It is hoped that such trials can be conducted expeditiously and designed to terminate early in case of a strong early signal of benefit or lack thereof. Shortening the typically long timeline for screening trials would clearly be welcomed by the public, but this is challenging because benefit accrues over time, so short-duration trials may not show a large clinical benefit.

It is likely that MCED tests will be made available to the public and submitted to regulatory agencies for approval and reimbursement decisions long before trial results are available. We have been in a similar situation before. Following the approval of the PSA test for prostate cancer surveillance after diagnosis, PSA screening rapidly disseminated into the population, dramatically affecting disease incidence, stage distributions, and (eventually) mortality. Rates of PSA screening were not tracked in real time and had to be assembled retrospectively (21) to provide information about the connection between PSA testing and changes in prostate cancer outcomes (22,23). Screening trials were launched, but much of the impact of PSA testing had already been felt, and a great deal had already been learned, by the time their less-than-conclusive results were published (24).

We have learned, from the history of early detection, that cancer screening conducted in the general population is the ultimate uncontrolled experiment. When screening tests disseminate ahead of randomized trials, we are compelled to try and learn from this experiment. To do this for MCED tests, we need to be prepared to track their dissemination within the health system, catalog the populations that are being tested and why, and record the diagnostic pathways and patterns of care following a test. Figure 1 provides a schematic of the type of information that we would encourage collecting in a new registry of MCED test utilization and outcomes. Such a registry could help inform about demand for and costs of confirmation testing following MCED testing, detection of incidental conditions, access and equity, and health-related costs. Although the logistics may be challenging, experience clearly supports making a coordinated effort to collect these data if and when MCED tests disseminate into the population. These data will facilitate a wide range of studies that can be conducted in parallel with randomized trials. If we fail to do this, we will miss an opportunity to learn in real time what we need to know about what could arguably be the most important advance in cancer early detection technology in the last several decades.

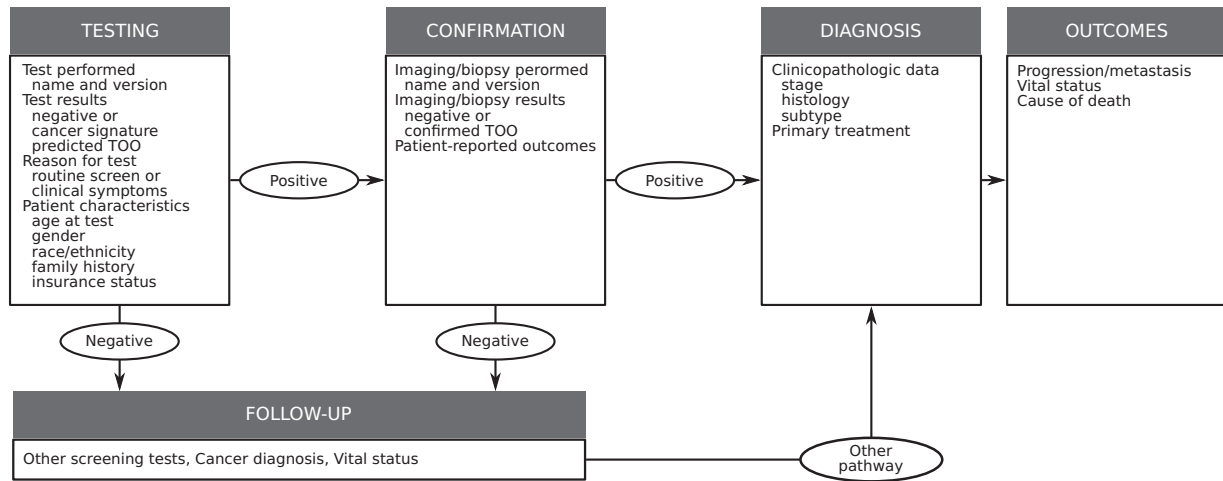


Figure 1. Schematic of the components of a data registry to track the utilization and outcomes of patients receiving multicancer early detection tests. TOO = tissue of origin.

Funding

This work was supported in part by the Rosalie and Harold Rea Brown Endowed Chair (to RE) and the National Cancer Institute at the National Institutes of Health (grant number R50 CA221836 to RG).

Notes

Role of the funders: The funding agencies had no role in the design of the study; the collection, analysis, or interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Disclosures: Dr Etzioni reported receiving personal fees from Grail prior to the submitted work. Dr Etzioni also holds shares in Seno Medical. The other authors declare no potential conflicts of interest.

Author contributions: Conceptualization, RE; Writing—Original Draft, RE; Writing—Review & Editing, all authors.

Data Availability

No primary data was used in this manuscript. All figures cited are from published sources.

References

- Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926–930.
- Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. 2019;570(7761):385–389.
- Liu MC, Oxnard GR, Klein EA, et al.; for the CCGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020;31(6):745–759.
- Pepe MS, Etzioni R, Feng Z, et al. Elements of study design for biomarker development. In: E Diamondis, HA Fritsche, H Lilja, et al., eds. *Tumor Markers: Physiology, Pathobiology, Technology, and Clinical Applications*. Washington DC: AACR Press; 2002: 141–150.
- Croswell JM, Ransohoff DF, Kramer BS. Principles of cancer screening: lessons from history and study design issues. *Semin Oncol*. 2010;37(3):202–215.
- Aberle DR, Adams AM, Berg CD, et al.; for 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.
- Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: model estimates of potential benefits and harms. *Ann Intern Med*. 2013;158(3):145–153.
- Krist AH, Davidson KW, Mangione CM, et al.; for the US Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2021;325(10):962–970.
- Meza R, Jeon J, Toumazis I, et al. Evaluation of the benefits and harms of lung cancer screening with low-dose computed tomography: modeling study for the US Preventive Services Task Force. *JAMA*. 2021;325(10):988–997.
- Bibbins-Domingo K, Grossman DC, Curry SJ, et al.; for the US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*. 2016;315(23):2564–2575.
- Siu AL; for the Force USPST. Screening for breast cancer: U.S. Preventive Services Task Force Recommendation statement. *Ann Intern Med*. 2016;164(4):279–296.
- Shen Y, Zelen M. Screening sensitivity and sojourn time from breast cancer early detection clinical trials: mammograms and physical examinations. *J Clin Oncol*. 2001;19(15):3490–3499.
- Meza R, ten Haaf K, Kong CY, et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer*. 2014;120(11):1713–1724.
- Draisma G, Boer R, Otto SJ, et al. Lead times and overdiagnosis due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. *J Natl Cancer Inst*. 2003;95(12):868–878.
- Clarke CA, Hubbell E, Kurian AW, et al. Projected reductions in absolute cancer-related deaths from diagnosing cancers before metastasis, 2006–2015. *Cancer Epidemiol Biomarkers Prev*. 2020;29(5):895–902.
- Autier P, Hery C, Haukka J, et al. Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. *J Clin Oncol*. 2009;27(35):5919–5923.
- Schroder FH, Hugosson J, Carlsson S, et al. Screening for prostate cancer decreases the risk of developing metastatic disease: findings from the European Randomized Study of Screening for Prostate Cancer (ERSPC). *Eur Urol*. 2012;62(5):745–752.
- Weiss NS. Commentary: cohort studies of the efficacy of screening for cancer. *Epidemiology*. 2015;26(3):362–364.
- Weiss NS, Dhillon PK, Etzioni R. Case-control studies of the efficacy of cancer screening: overcoming bias from nonrandom patterns of screening. *Epidemiology*. 2004;15(4):409–413.
- Etzioni R, Gulati R, Cooperberg MR, et al. Limitations of basing screening policies on screening trials: the US Preventive Services Task Force and prostate cancer screening. *Med Care*. 2013;51(4):295–300.
- Mariotto A, Etzioni R, Krapcho M, et al. Reconstructing prostate-specific antigen (PSA) testing patterns among Black and White men in the US from Medicare claims and the National Health Interview Survey. *Cancer*. 2007;109(9):1877–1886.
- Etzioni R, Gulati R, Falcon S, et al. Impact of PSA screening on the incidence of advanced stage prostate cancer in the United States: a surveillance modeling approach. *Med Decis Making*. 2008;28(3):323–331.
- Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics*. 2008;64(1):10–19.
- Barry MJ. Screening for prostate cancer—the controversy that refuses to die. *N Engl J Med*. 2009;360(13):1351–1354.