Making Sense of the Numbers: Interpreting Multi-Cancer Early Detection Test Performance

Megan P. Hall, PhD
VP, Medical & Corporate Affairs, Distinguished Scientist

The Galleri® multi-cancer early detection (MCED) test is ushering in a transformative era in cancer screening. With a single blood draw, Galleri® is capable of detecting DNA signals associated with numerous deadly cancers before symptoms emerge, offering an opportunity for a cure. The test is particularly sensitive to aggressive cancers that are often diagnosed too late, such as pancreatic, liver, and ovarian cancers, and it rarely detects a false positive signal in individuals without cancer. When a cancer signal is detected, Galleri® also predicts the Cancer Signal Origin (CSO)—a feature increasingly recognized as important to help healthcare professionals determine where that signal is coming from in the body, to help enable targeted, efficient follow-up (Ahlquist 2018, FDA Advisory Committee Meeting 2023, Brito-Rocha 2023, Constantin 2022, Hackshaw 2022, Putcha 2021, Schrag 2023).

As the field evolves, and as MCED testing becomes more widely adopted, it becomes increasingly important to understand how to interpret results across different types of studies and technologies. This necessitates a clear understanding of the underlying performance metrics and their implications for multi-cancer screening. A detailed examination of these metrics can help clarify both individual- and population-level benefits, and what MCED results truly mean for healthcare professionals and individuals.

The Goal of Multi-Cancer Early Detection

We believe that the goal of any multi-cancer early detection test used for population screening should be two-fold: detect as many aggressive cancers as possible, and maximize the likelihood that a positive test result is actually cancer. These two metrics, Cancer Detection Rate (CDR) and Positive Predictive Value (PPV), respectively, reflect critical population-level and individual-level measures. They are also not traditionally considered in our current single-cancer screening paradigm. We explore why in the following sections.

Traditional Measures in Single-Cancer Screening—And Why Multi-Cancer Tests Are Different

The current recommended cancer screening tests focus on a single organ or tissue type. As such, the effectiveness of these single-cancer screening tests has historically been measured by how well they detect a single type of cancer. Well-known examples include mammograms for breast cancer, or colonoscopies or fecal immunochemical testing/fecal occult blood testing (FIT/ FoBT) for colon cancer. For these single-cancer tests, reporting performance measures is well-established: the presence or absence of just one type of cancer is considered, and all test results relate directly to that cancer.

This landscape is evolving with the introduction of MCED tests. These tests are designed to simultaneously screen for many types of cancer, each with its own prevalence, risk factors, and clinical consequences. Using the same traditional metrics (calculated as if a test were looking for just one kind of cancer) can be inappropriate when applied to MCED technology. As such, these tests demand a new approach—one that considers how performance metrics are impacted by cancer types and prevalence, the need for accurate cancer signal origin information, and the broader population-scale impact of screening for many diseases at once.

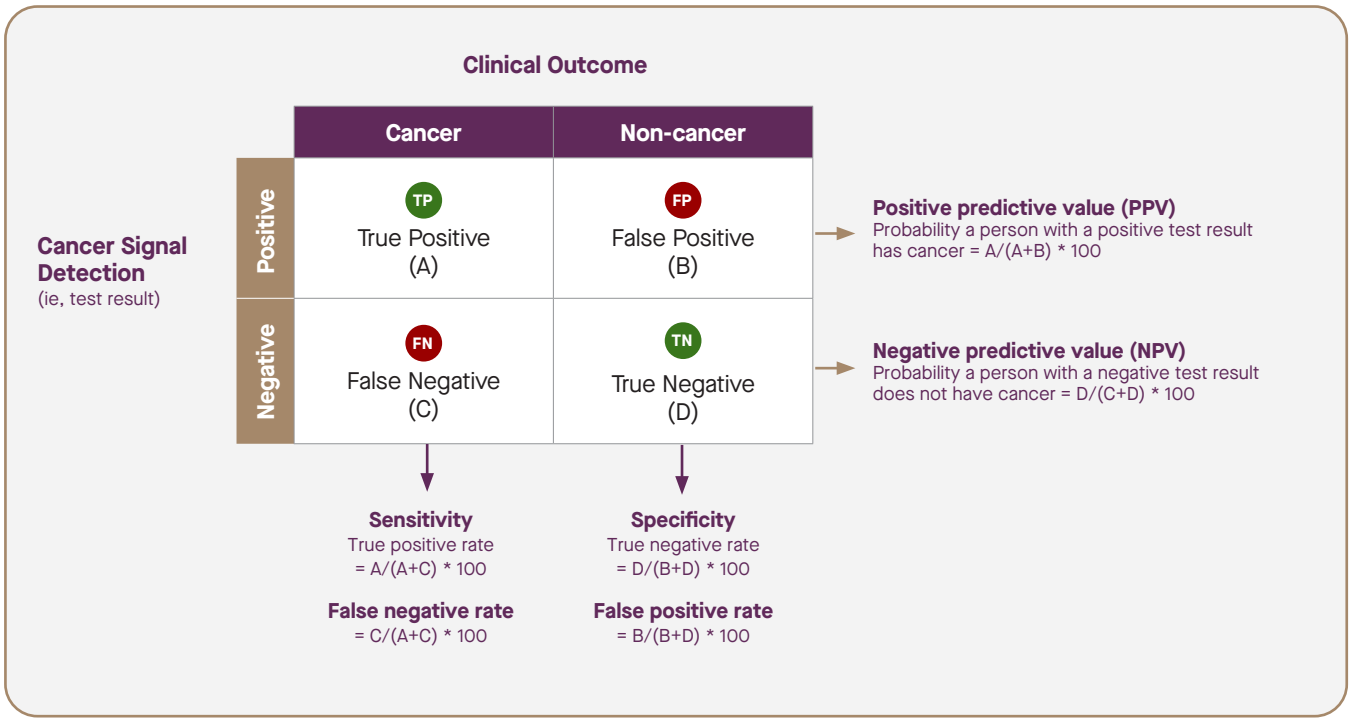
Test Outcomes: The Foundation of Performance Metrics

To understand and interpret performance metrics, we must first define the possible outcomes. The evaluation of a diagnostic test (Hintze NCSS) compares the test's result—either "cancer" or "no cancer"—against the individual's actual condition. The four possible outcomes are:

- **True Positive (TP):** The test indicates a cancer signal, and cancer is present.
- **True Negative (TN):** The test indicates no cancer signal, and cancer is absent.
- **False Positive (FP):** The test indicates a cancer signal, but cancer is absent.
- **False Negative (FN):** The test indicates no cancer signal, but cancer is present.

These outcomes are usually summarized in a "2×2" classification table (Figure 1).

Figure 1. A 2×2 classification table describing the key performance characteristics of a binary diagnostic test.



Sensitivity and Specificity: Traditional Measures That Require Knowing an Individual's Cancer Status

Sensitivity: How Often Does the Test Detect Cancer?

Sensitivity (or the true positive rate) measures how likely the test is to find cancer in individuals who actually have cancer (i.e., their cancer status is known). As such, it requires knowing whether someone has cancer at the time of the test. For example, if a test returns positive results for 71 out of 100 people who have cancer, its sensitivity is 71%. High sensitivity maximizes the cancers that are detected—and single-cancer tests, like mammography, are designed to maximize sensitivity. This metric is best assessed in case-control studies, where the cancer status of each individual is known.

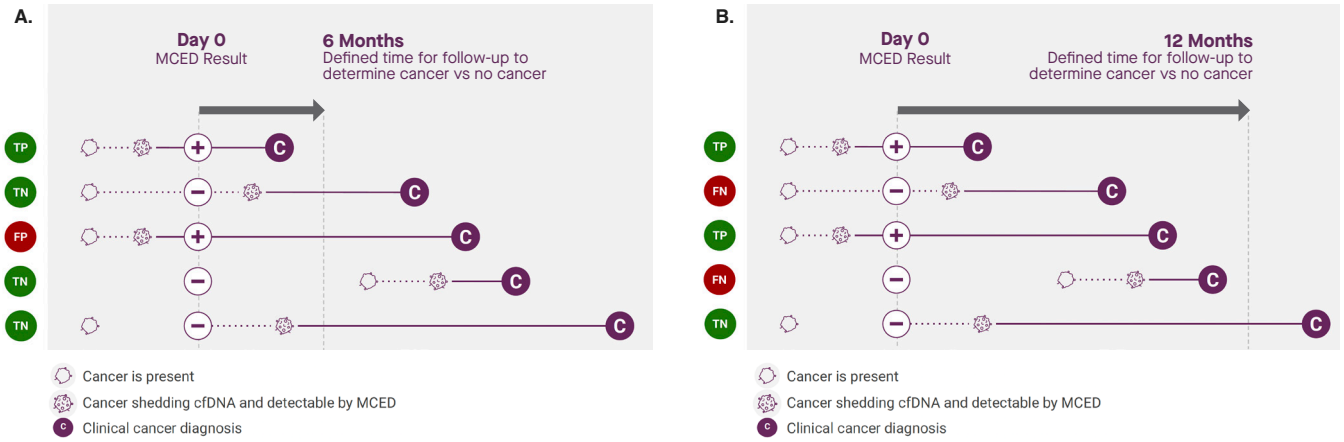
Translating sensitivity to a multi-cancer test, however, is not straightforward. For this kind of test, the sensitivity measure reflects all of the cancer types the test detects: for Galleri®, that's more than 50 cancer types—and those cancer types are broad buckets that include numerous cancer subtypes (for example, HPV-positive and HPV-negative head & neck cancers). And—when calculating sensitivity, the denominator should be all cancers. Exclusion of cancers that are known to have low signal, such as early-stage prostate and breast cancers, or thyroid cancers, from this calculation can artificially elevate sensitivity. But—there is an important trade-off: excluding cancer types to artificially elevate sensitivity (or because the test is only designed to detect a smaller number of cancers) reduces the overall Cancer Detection Rate (CDR), which is described below. And—CDR is one of the most important measures of a multi-cancer test's efficacy.

Episode Sensitivity: Understanding Performance in a Real-World Screening Setting

Most laboratory measures, such as sensitivity and specificity, are calculated in controlled settings like case-controlled studies, where the cancer status of all individuals is known at the time the test is performed. By contrast, the objective of cancer screening in the general population is to find cancers that we don't yet know about—it involves screening individuals without a known diagnosis at the time of testing. Here, the concept of episode sensitivity becomes relevant. Episode sensitivity refers to the proportion of cancers detected at the time of the initial screening test, out of all cancers diagnosed in individuals during a pre-defined follow-up period. In an intended use screening population, an individual's cancer status is not known; this means that individuals must be followed for a pre-defined period of time to determine their cancer status—and whether that original test detected or missed that cancer. The time of follow-up used to define cancer status can impact performance metrics. A shorter follow-up period can minimize false negatives, whereas a longer follow-up period minimizes false positives (Figure 2). A one-year follow-up period has been used in other large-scale cancer screening studies (eg, Lehman et al, Menon et al, Pinsky et al), including the PATHFINDER and PATHFINDER 2 studies.

Episode sensitivity yields a more conservative measure than traditional 'test sensitivity' (which is what is measured in case-control studies), but more accurately reflects the impact of a test on population health. Critically, for a multi-cancer test to be considered fully clinically validated, we believe it should be assessed in the intended use population—not just the case-control setting.

Figure 2. Episode sensitivity as measured in interventional trials in the intended use population. True positives (TP), false positives (FP), true negative (TN), and false negatives (FN) are indicated to the left. The MCED test result is indicated at Day 0 as positive (+) or negative (-). Onset of cancer development undetectable by MCED is indicated by the open cells. Cancer shedding cfDNA and detectable by MCED is indicated by the rough cells. A clinical cancer diagnosis is indicated with a purple "C". The impact on TPs, FPs, TNs, and FNs is depicted for (A) a 6-month follow-up period, and (B) a one-year follow-up period.



Cancer Detection Rate: How Many Cancer Cases Does the Test Detect? A Critical Population-Level MCED Test Metric.

The Cancer Detection Rate (CDR) expresses the number of cancers identified in the screened population, which can be expressed as a percentage. For example, if a test uncovers 120 cancers among 10,000 screened subjects, the CDR is 1.2%. This links statistical analysis to a straightforward, tangible outcome: the number of cancers actually detected via screening.

For a multi-cancer test, CDR is a critical population-scale performance measure. A multi-cancer test designed to detect many cancer types—like Galleri®—will optimize the CDR. By contrast, tests that only focus on a few cancer types will have a lower CDR—they will miss cancers in the population that they are not 'looking for.' Another critical consideration is that CDR and sensitivity are inversely related. Above we described how sensitivity can be artificially elevated by excluding cancers that have lower sensitivities; this approach has a direct negative impact on CDR. Excluding low-signal cancers, like prostate cancer, to achieve a high sensitivity means a lower CDR—fewer cancers detected in the population. A test that detects 20% of all cancers will identify significantly more cancers in the population than a test that detects 100% of one cancer (regardless, no single-cancer screening test exists with 100% sensitivity) (Figure 3).

Specificity: How Often Does the Test Avoid False Positives?

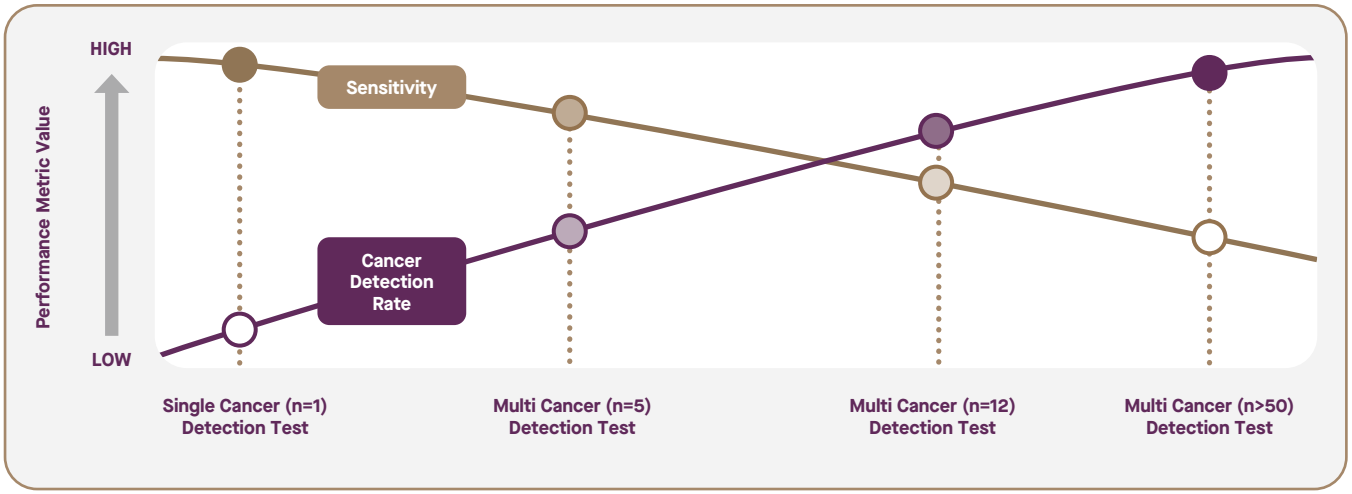
Specificity (or the true negative rate) measures how likely the test is to return a negative result in individuals without cancer. Similarly to sensitivity, it requires understanding an individual's cancer status at the time of the test, and is measured in individuals without a cancer diagnosis. For example, if a test provides negative results for 74 out of 100 people without cancer, its specificity is 74%. Specificity can also be expressed as the inverse of the false positive rate (FPR, where specificity = 1-FPR). In this example, a test providing negative results for 74 out of 100 individuals would thus return incorrect positive results for 26 out of 100 individuals without cancer—and the FPR for this test would be 26%.

High specificity (or a low FPR) minimizes the occurrence of false positives, and potentially unnecessary follow-up procedures. It is a critical safety metric for all screening tests, but especially for MCED tests. There is an inherent trade-off between sensitivity and specificity; because single-cancer tests are designed to maximize sensitivity, they tolerate a lower specificity—meaning they detect more cancers at the expense of more false positives.

By contrast, multi-cancer tests should be designed to maximize specificity. This is intentional, to minimize false positives and avoid unnecessary diagnostic procedures. And—these specificity metrics are extremely high: 99.5% for Galleri® (Klein 2021, Schrag 2023), and 98.5% (or below) for other tests in development (Abraham 2025, Seeking Alpha). Although the CancerSEEK assay originally reported a specificity of greater than 99% (Cohen 2018), when studied in a clinical trial in the intended use population the specificity of the first blood test was 95.3% (Lennon 2020); the importance of robust clinical validation in the intended use population, as has been reported for the Galleri® test (Schrag 2023), is described in Ofman 2025. For the Galleri® test, that 99.5% specificity (or – a 0.5% FPR) means that only one out of 200 people without cancer will get a positive test result—the remaining 199 without cancer will get a correct negative test result.

These may seem like minor differences, but on a population-scale, what appear as minor differences in specificities translate to a large number of impacted individuals. Compared to a test reporting a 99.5% specificity (or a 0.5% FPR), tests reporting a 98.5% specificity will have three times as many false positives, and a test with a 95.3% specificity will have at least 9.4 times as many false positives. On a population-scale, this translates to between 1,000 and 4,200 additional false positives per 100,000 screened individuals.

Figure 3. The inverse relationship between Sensitivity and Cancer Detection Rate (CDR).



The Predictive Values: What The Test Results Actually Mean for the Individual

Positive Predictive Value: What is the Likelihood of Cancer if the Test is Positive?

The positive predictive value (PPV) reports how likely it is that a person with a positive test result actually has cancer. PPV is a critical performance metric for the individual being screened, because it represents the proportion of positive test results that are actually cancer. Unlike sensitivity, the PPV doesn't require knowing an individual's cancer status at the time of the blood draw, which makes it a more meaningful metric for a screening test. For individuals, PPV shows how much they can trust a positive result. This metric depends mainly on specificity and on how common cancer is in the group being tested (prevalence). Higher specificity and higher prevalence both increase the PPV (Ahlquist 2018)—and, a high PPV translates to fewer false positives.

For multi-cancer tests, where screening is expected to be performed in a large number of people and where cancer is still fairly rare, PPV is the best metric to determine whether a positive test result can be "trusted", and predict whether a positive test means real disease. A high PPV means positive results are more likely to be true, helping to ensure genuine benefits without unnecessary false alarms and unnecessary diagnostic testing.

Multi-cancer test results should be trustworthy and impactful, and should be designed to maximize the PPV and therefore lower the risk of a false positive. This is achieved by maintaining an extremely high specificity (eg, >99%), and by detecting as many cancer types as possible. By contrast, single-cancer tests are designed for high sensitivity—this means their PPVs are often much lower (typically in the single digits, as is observed for mammography, which has a PPV of 4.4% [Lehman 2017]), because they tolerate a lower specificity and only target one cancer type (so have a lower prevalence). For example, MCED tests benefit from the aggregate prevalence of many cancers, which substantially increases the overall likelihood that a positive result corresponds to a true cancer signal. This aggregation effect can lead to significantly higher PPVs for MCED tests, making them informative and actionable for individuals and clinicians alike (Ahlquist 2018), particularly when the test also predicts with cancer signal origin with high accuracy.

Because Galleri® can detect so many cancer types with high specificity, it reached a PPV of 43% in previous clinical studies (Klein 2021, Schrag 2023), and even higher PPVs in real-world reports (Hurt 2025, O'Donnell 2025, Sehbai 2025, Atwood 2024). More recently, both PATHFINDER 2 and NHS-Galleri® trials reported PPVs that are "substantially higher" than observed in prior clinical studies (GRAIL 2025 [1], GRAIL 2025 [2]). These consistently high PPVs are a direct result of Galleri's high specificity and broad cancer coverage.

A critical point of confusion: An absolute number of false positives is often derived from the PPV, and incorrectly assumed to be the FPR. For example, if the PPV is 50%, it is often assumed that the remaining 50% represent the FPR. That remaining 50% is actually a separate metric called the false discovery rate, which is not addressed in detail here. As a reminder, PPV is results-centered, and it answers this question: "Among individuals with a positive test result, how often is it correct?" By contrast, FPR is cancer status-centered, asking: "Among individuals who don't have cancer, how often does the test falsely report a positive result?" The absolute number of false positive results calculated from the PPV cannot be directly translated to the FPR, because with predictive values a person's cancer status is not known.

Negative Predictive Value (NPV): How Likely is a Negative Test Result Not Cancer?

The negative predictive value represents the proportion of negative test results that are not cancer. Like PPV, the NPV does not require knowing a person's cancer status at the time of the blood draw. The NPV indicates the proportion of negative test results that correctly identify an individual without cancer. In most screening contexts where cancer is rare, the NPV is quite high, and may offer significant reassurance with a negative result. NPV may be more relevant in a symptomatic setting, where it can be used for triage and decision-making.

Summary

PPV and CDR are two of the most informative performance metrics for MCED tests because of the complexity and scope of detecting multiple cancer types simultaneously. No one gets to choose the cancer they get diagnosed with, so adding multi-cancer screening with a high CDR can lead to a paradigm shift that can provide great benefit to an individual while providing a major population health benefit. Unlike single-cancer screening tests, which focus on maximizing sensitivity for a specific cancer, MCED tests must balance broad detection across many aggressive cancers with minimizing false positives. As such, MCED tests must distinguish true cancer signals from a vast background of non-cancer variation across many tissues. High PPV is essential to avoid a flood of false positives that could lead to unnecessary diagnostic workups across multiple organ systems. Similarly, a strong CDR ensures that the test is actually effective at detecting a meaningful proportion of cancers, including those cancers without existing screening methods. Together, these metrics highlight the critical value of MCED testing: the potential to shift the paradigm of cancer screening by providing significant benefits for individuals while advancing population-level health outcomes.

References

1. 24 Hour Summary Molecular and Clinical Genetics Panel Advisory Committee Meeting. November 29, 2023 (https://www.fda.gov/media/174413/download).
2. Ahlquist DA. Universal cancer screening: revolutionary, rational, and realizable. NPJ Precis Oncol. 2018;2:23. doi: 10.1038/s41698-018-0066-x.
3. Atwood C, Moy S, Kindy MS, et al. REFLECTION: Initial Findings From a Real-World Evidence Study of Multi-Cancer Early Detection (MCED) and Toxic Exposures Among Veterans in the Veterans Affairs Healthcare System (VA). Presented at the Early Detection of Cancer Conference (EDCC); October 22 - 24, 2024; San Francisco, CA.
4. Adapted from "Binary Diagnostic Tests – Single Sample (Chapter 535)". Accessed September 12, 2025. https://www.scribd.com/document/386482716/Binary-Diagnostic-Tests-Single-Sample
5. Brito-Rocha T, et al. Shifting the Cancer Screening Paradigm: The Rising Potential of Blood-based Multi-Cancer Early Detection Tests. Cells. 2023;12,935.
6. 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.
7. Constantin N, et al. Opportunities for Early Cancer Detection: The Rise of ctDNA Methylation-based Pan-cancer Screening Technologies. Epigenomes. 2022;6(1):6.
8. GRAIL, Inc. [1] (2025, May 13). GRAIL Reports First Quarter 2025 Financial Results [press release]. https://grail.com/press-releases/grail-reports-first-quarter-2025-financial-results/
9. GRAIL, Inc. [2] (2025, June 18). GRAIL Announces Positive Top-Line Results From The Galleri® PATHFINDER 2 Registrational Study [Press release]. https://grail.com/press-releases/grail-announces-positive-top-line-results-from-the-galleri%e2%93%a1-pathfinder-2-registrational-study/
10. Seeking Alpha. Guardant Health, Inc. (GH) Q2 2025 Earnings Call Transcript. Seeking Alpha. July 31, 2025. Accessed September 22, 2025. https://seekingalpha.com/article/4806840-guardant-health-inc-gh-q2-2025-earnings-call-transcript
11. Hackshaw A, et al. New genomic technologies for multi-cancer early detection: Rethinking the scope of cancer screening. Cancer Cell. 2022;40(2):109-113.
12. Hintze JL. Binary Diagnostic Tests - Single Sample. NCSS Statistical Software. Accessed August 3, 2025. https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Binary_Diagnostic_Tests-Single_Sample.pdf
13. Hurt RT, Ghosh AK, Dougan BM, et al. Implementation of a Multicancer Detection (MCD) Test in a Tertiary Referral Center in Asymptomatic Patients: An 18-Month Prospective Cohort Study. J Primary Care Comm Health 2025;16:21501319251329290. Epub 2025 Mar 25. doi: 10.1177/21501319251329132.
14. Klein EA, Richards D, Cohn A, et al. Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann Oncol 2021 Sep;32(9):1167-1177. doi: 10.1016/j.annonc.2021.05.806.
15. Lehman CD, Arao RF, Sprague BL, et al. National performance benchmarks for modern screening digital mammography: update from the Breast Cancer Surveillance Consortium. Radiology. 2017; 283: 49–58.
16. Lennon AM, Buchanan AH, Kinde I, et al. Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science. 2020 Jul 3;369(6499):eabb9601. doi: 10.1126/science.abb9601. Epub 2020 Apr 28.
17. Menon U, Gentry-Maharaj A, Hallett R, et al. Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). Lancet Oncol 2009; 10: 327–40.
18. O'Donnell E, Kauffman T, Beckwith J, et al. Implementation and evaluation of multi-cancer early detection testing at the Dana-Farber Cancer Institute: A retrospective analysis of clinical outcomes and diagnostic pathways. J Clin Oncol. 2025;43(16):suppl_11159. doi: 10.1200/JCO.2025.43.16_suppl.11159.
19. Ofman J, Hall MP, Clarke-Dur C. Not All MCED Tests are Created Equal: The Realities of MCED Test Development and Validation. July 31, 2025. Accessed September 22, 2025. https://grail.com/stories/not-all-mced-tests-are-created-equal-the-realities-of-mced-test-development-and-validation/
20. Pinksy PF, Geirada DS, Black W, et al. Performance of Lung-RADS in the National Lung Screening Trial: a retrospective assessment. Ann Intern Med. 2015 Apr 7;162(7):485-91.
21. Putcha G, et al. Multi-cancer Screening: One Size Does Not Fit All. JCO Precis Oncol. 2021;5:574-576.
22. Schrag D, Beer TM, McDonnell CH, et al. Blood-based tests for multi-cancer early detection (PATHFINDER): a prospective cohort study. Lancet. 2023;402:1251-1260.
23. Sehbai AS, Asif H, Rashid M, et al. A clinical review of a novel blood test use in rural Alabama for multi-cancer detection, analyzing methylation patterns of cell-free DNA and future strategies. J Clin Oncol. 2025;43(16):suppl_e151. doi: 10.1200/JCO.2025.43.16_suppl.e151.

For further details about Galleri® test performance, visit grail.com and galleri.com.