What Are Prenatal Genetic Screening Tests
and How Should They be Interpreted?

Jazmin E. Juarez MD, Sally F. Burn PhD, Anika M. Deluca BFA, Kristie N. Smith MS, CGC, Marra S. Francis MD, Lisa Edelmann PhD, Gustavo A. Stolovitzky PhD, Eric E. Schadt PhD

Overview

In its recent article, "When They Warn of Rare Disorders, These Prenatal Tests Are Usually Wrong," *The New York Times* published a discussion on the performance of prenatal blood screening tests. The piece, which also looked at the potential impact of results on patients, has fueled an online debate on medical ethics and misinformation. As a patient-centered health intelligence company with deep expertise in reproductive health and medical testing, we see this as a perfect opportunity to inform our partners on population health topics to help them better understand the accuracies (and inaccuracies) of clinical testing.

In this guide, we aim to educate readers about the differences between screening and diagnostic tests and, crucially, why "positive" screening results should be interpreted as *"further testing is recommended"* and *"negative"* results as *"no abnormalities detected; no further testing recommended."*

Executive Summary

- ***Screening*** tests evaluate the risk of having a disease now or in the future. They personalize a patient's ***pre-test probability*** of disease to weigh the risk and benefits of invasive diagnostic testing.
- Test characteristics such as ***sensitivity*** and ***specificity*** represent the probability of yielding ***false negative*** and ***false positive*** rates, respectively.
- Screening tests are highly sensitive to maximize the likelihood of capturing all true positive events at the expense of lower specificity and higher false positive rates.
- Test ***accuracy*** estimates the strength of association between pre-test probability and ***post-test probability.***
- ***Positive*** and ***negative predictive values*** are dependent on ***disease prevalence***, or pre-test probability. As prevalence decreases, PPV decreases, increasing the false positive rate.
- ***Noninvasive prenatal tests (NIPT)*** for microdeletions have low PPVs (due to these conditions' low frequency in the general population) and low sensitivity (due to limitations of today's technology), yielding higher false positive rates.
- A low PPV is tolerated to maximize knowledge of health risks and enable more targeted follow-up diagnostic testing in a smaller, higher-risk population.
- **"Positive"** screening results should be interpreted as ***"further testing is recommended"*** and **"negative"** results as ***"no abnormalities detected; no further testing recommended."***

Screening vs. diagnostic testing

Medical tests look for answers to specific health questions. ***Screening*** tests are performed to evaluate the risk of having a disease now or in the future, while ***diagnostic*** tests are performed to pinpoint the cause of a disease process and narrow down a diagnosis. Diagnostic tests also tend to be more invasive, associated with some risk, and costly. Screening tests, on the other hand, provide a cost-effective way to identify individuals at higher risk of disease, shifting the benefit-to-risk ratio to justify diagnostic testing and invasive treatments. The goal of screening tests is early diagnosis for early intervention: they provide valuable information that can indicate when diagnostic testing is needed. After diagnostic testing, an early diagnosis can help minimize long-term adverse health outcomes and optimize potential treatment options.

Noninvasive prenatal testing, or NIPT, is a blood-based screening test for expectant mothers. It evaluates the risk of fetal anomalies caused by certain chromosomal aberrations, such as aneuploidies (extra or missing chromosomes) and microdeletions (missing portions of chromosomes). Professional medical societies like the American College of Obstetricians and Gynecologists (ACOG) and the Society for Maternal-Fetal Medicine (SMFM) recommend that all pregnant women be offered prenatal screening or diagnostic testing for common aneuploidies to provide expectant parents with information about their pregnancy's risk of a genetic condition. After informed discussions with their healthcare providers about the benefits and risks of prenatal screening, this information can help parents prepare for potential complications at birth since conditions like Patau syndrome and Edwards syndrome (trisomy 13 and 18) are usually fatal by the first year of life. Early knowledge of potential complications can enable parents to deliver at a facility equipped to deal with high-risk births and connect with specialists who can guide their child's medical care.

As with all screening tests, patients opting to undergo NIPT are evaluated in the context of a larger population, based on a ***pre-test probability*** or the ***disease prevalence*** (i.e., how often a disease is present in that specific population). For example, Edwards syndrome (or trisomy 18) is the second most common fetal aneuploidy after Down syndrome, with a prevalence of one in 5,500 live births. Therefore, one newborn in 5,500 will be expected to have the condition among affected pregnancies that make it to term without screening. However, the pre-test probability also depends on patient-specific factors like ethnic background and maternal age. As the risk of fetal aneuploidy rises with age, expectant mothers are encouraged to undergo prenatal screening to detect the potential risk of fetal genetic abnormalities.

In short, screening tests are not intended to diagnose a particular condition. Instead, they serve to better personalize a person's pre-test probability of disease and weigh the need for further invasive testing.

Evaluating test utility

Test performance is evaluated based on the ability to reliably distinguish high-risk individuals or patients with disease from healthy people. A test's ***sensitivity*** gauges how well it can detect risk or disease in people who are, in fact, at high risk or sick. We can think of test sensitivity like modern home security systems that detect unexpected entries. Screening tests will detect some level of ***false positive*** alarms (perhaps, in the case of a home security system, a visiting relative), but better to be safe than sorry. The relationship between a test's sensitivity and ***false negative*** rate (the probability that an affected person incorrectly tests 'negative' or, in this case, that an intruder is not detected) can be calculated by subtracting sensitivity from 1. Because no test is perfect, positive screening results should be followed up with diagnostic tests, which are usually more specific, to facilitate a diagnosis.

Specificity gauges how well a test identifies low-risk or unaffected people who are, in fact, truly negative. When a test result is positive, specificity helps us understand whether a positive result is likely due to that disease versus other factors. The ***false positive*** rate (the probability an unaffected person incorrectly tests positive) is calculated by subtracting specificity from 1. In an ideal world, medical tests would be both perfectly sensitive and perfectly specific so that all affected individuals reliably have positive results and all healthy individuals confidently test negative. However, adjusting one of these values will generally inversely affect the other. Think of an email spam filter: setting the filter such that all emails received are sent to the spam folder would result in a filter with 100% sensitivity but at the cost of lowering specificity to 0%, which is equivalent to a false positive rate of 100%. In this case, all vital emails would be labeled as spam. In clinical practice, we are forced to strike a balance between sensitivity and specificity.

Sensitivity and specificity help speak to the ***accuracy*** of a test. Accuracy is the ability to differentiate healthy or low-risk individuals from high-risk or sick patients. It serves to estimate the strength of association between pre-test probability and ***post-test probability,*** a person's 'updated' probability of having a disease after testing. However, even in screening tests with more than 99% sensitivity, specificity, or accuracy, "positive" or "increased risk" results only tell us that an individual falls among the true positive or false positive cases. Screening tests are tuned to have higher sensitivity to cast as broad a net as possible to capture all potentially positive events, even if it means the test has slightly lower specificity, yielding a higher false positive rate. Higher sensitivity helps to maximize the likelihood of capturing all truly positive cases. This approach contrasts with diagnostic tests that aim to diagnose or "rule out" disease with near-perfect sensitivity and specificity.

Positive and negative predictive values (PPV and NPV, respectively) help determine the predictive utility of screening tests more confidently, especially in the context of screening for low-population frequency events (where even tests with high sensitivity and specificity will yield more false positive than true positive results). PPV, or test **precision**, reveals the proportion of positive test results that are true positive after considering factors like a person's medical history. For example, since the risk of aneuploidy increases with maternal age in the general population, the PPV of Edwards syndrome in a 20-year old's pregnancy is 14% compared to 69% at 40-years of age[1]. We can estimate that 69% of pregnancies with positive screens in a given population of 40-year-old patients are truly affected.

Predictive values are also highly dependent on disease prevalence *(Figure 1)*. If a disease is highly prevalent and a person tests positive for it, a high PPV means a higher likelihood that this person truly has the disease. However, as the prevalence of a disease decreases, the PPV also decreases because we can expect more false positive results for every true positive result, even with very high sensitivity and specificity. A lower PPV also increases the NPV since, for rare diseases, we can expect more individuals who are truly negative for every false negative result.

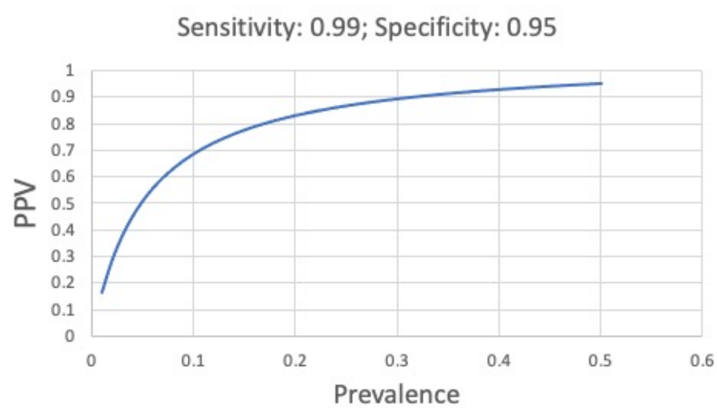

Figure 1. PPV is dependent upon disease prevalence. For equally sensitive and specific tests, PPV decreases with decreasing disease prevalence.

PPV and NPV help assess whether a given screening test has good enough utility to be broadly applied. A lower PPV along with a very high NPV is often tolerated to maximize knowledge about health risks, help capture more true positive cases, and enable more targeted follow-up diagnostic testing in a smaller, higher-risk subpopulation *(Figure 2)*. For example, mammography screening for breast cancer yields a high false positive rate in genetic females under 50. In this age group, we can expect more true negative results (high NPV) from mammography screening since the prevalence of breast cancer is lower in younger individuals. We can also expect more false positive than true positive results given the low population prevalence (low PPV), even though this warrants more invasive diagnostic testing (a situation similar to that seen with NIPT). Yet, the American Cancer Society recommends breast cancer screenings start at age 45 because the benefit of early screening on mortality heavily outweighs higher false positive rates in the 45-50-year old age group.

Therefore, the best way to think about *"positive"* screening test results is to interpret the result as *"further testing is recommended"* and *"negative"* results as *"no abnormalities detected; no further testing recommended."*

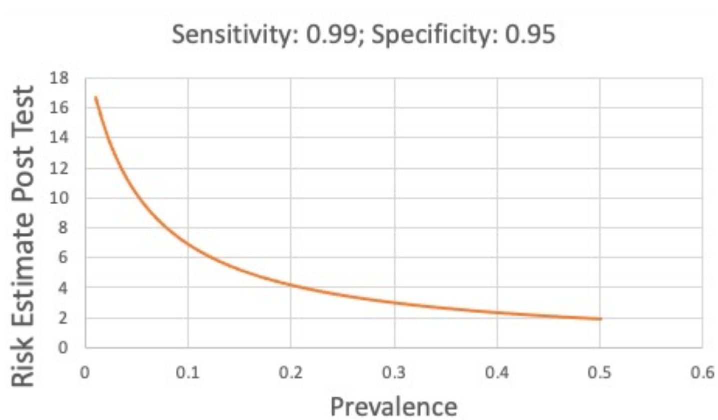

Figure 2. Post-test risk estimate is dependent on disease prevalence. Screening tests can increase our ability to detect at-risk individuals or pregnancies beyond the pre-test risk estimation for lower prevalence diseases. Pre-test, identifying individuals at risk for a rare disease is like trying to find a needle (positive case) in a haystack (larger population). However, post-test, the pile of hay (sub-population) is much smaller, and finding the needles becomes a more manageable task.

sema4

- True Positive
- True Negative
- False Positive
- False Negative

How Do We Find a Needle in a Haystack?

Or: How Do Medical Tests Identify At-Risk Individuals?

In a population of **10,000** people, a 1% disease prevalence

means we can expect **100** to be affected & **9,900** to be unaffected

A screening test with **99% sensitivity**

will **correctly** identify **99 OUT OF 100** with the disease as **true positive**

&

incorrectly label **1 OUT OF 100** as **"negative" (false negative)**

A screening test with **95% specificity**

will **correctly** identify **9,405** people **without the disease** as **true negative**

&

incorrectly label **495 healthy people as "positive" (false positive)**

resulting in **5 times as many false positives** as **true positives**

495 > 99

Screening with this test identifies **594** individuals at higher risk (5.5% of the population)

$$PPV = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

$$\frac{99}{594} = 17\%$$

= 17x enrichment in this **sub-population**, making it easier to find ***a needle in the haystack*** (true positives among false positives)

Diagnostic tests are indicated to **differentiate the true positives** among **all with positive results.**

Positive screening results should be interpreted as ***'further testing is recommended,'*** and **negative** results as **'no abnormalities detected; no further testing is recommended.'**

Patients should engage in conversations with their HCPs* about the goals, risks, and benefits of screening tests.

*healthcare providers

Understanding screening test results

For rare disorders and conditions with a high disease burden, the goal is to detect as many suspected cases as possible to maximize knowledge about related health risks. Take a group of 10,000 pregnant individuals in their first trimester. In this group, suppose the prevalence of a rare disorder with a high risk of miscarriage, stillbirth, developmental disability, and poor quality of life is roughly 1% (100 out of 10,000 pregnancies). Suppose these 10,000 expectant patients undergo screening for this disorder with a 99% sensitive and 95% specific test. In this scenario, we expect 100 individuals to be truly positive. With a 1% false negative rate *(1-sensitivity)*, 99 of these 100 individuals would correctly receive a positive result. However, one affected person would incorrectly receive a negative result. Of 9,900 unaffected pregnancies, screening would correctly identify 9,405 as true negatives, although it might incorrectly label 495 healthy pregnancies as positive. We can expect about five times as many false positives as true positives and an estimated 17% PPV despite this test's excellent sensitivity and specificity.

At first glance, the above screening test may not appear informative, but the insights gained from these results are crucial. This test has captured nearly all true positive cases at higher risk of this rare disorder. Out of 10,000 individuals with a pre-test probability of 1% of having this rare disorder, it identified a subpopulation of 594 (99 true positives plus 450 false positives) individuals as "positive" with a 17% post-test probability (PPV). That is, the screening test identified a subpopulation that is 17-times more likely to have a severe genetic disorder than the general population that was screened. The resulting 594 "positive" individuals (5.9% of the population) are now armed with results to inform critical discussions with their healthcare providers about the need for confirmatory diagnostic testing. What is important is the degree to which a screening test reliably enriches for positive cases with respect to population prevalence, allowing for a more targeted diagnostic test in a high-risk subpopulation.

In sum, the reliability of test results depends on the test characteristics discussed above. We use screening tests, combined with an individual's medical history and population variables, to determine the risk and benefit of further testing. Medical professionals, ethicists, and society at large help determine whether screening tests with particular characteristics provide enough benefit to offset the risk of potentially more invasive but more accurate testing. These foundations help establish a screening test, such as NIPT, as a standard of care in medical practice.

Noninvasive prenatal testing (NIPT)

NIPT is one form of fetal screening in which cell-free DNA from a maternal blood sample is analyzed for potential chromosome abnormalities. These screenings can also detect microdeletion syndromes characterized by physical and intellectual impairment and a higher risk of adverse health effects. Microdeletions and aneuploidies are rare in the general population, so tests for these conditions naturally have lower PPVs. Microdeletion tests are also less sensitive than aneuploidy tests, given the rarity of microdeletions and the limitations of today's technology. Their false positive rates are higher because of this lower sensitivity. However, microdeletion tests still provide significant information from false negative rates. As these tests continue to evolve, their resolution and the trade-off between sensitivity and specificity will continue to improve.

Screening tests should always be followed up with formal diagnostic testing. Patients with positive NIPT results are encouraged to undergo diagnostic procedures like amniocentesis or chorionic villus sampling for more insight into their pregnancy's risk of a genetic condition. Because these diagnostic tests carry a small risk to the pregnancy, screening tests are essential for identifying high-risk individuals who warrant more invasive testing. In addition, we encourage all families to engage in substantiative conversations with their healthcare providers so they can be better informed of their child's health now and in the future. These conversations can help them prepare for their pregnancy and make the best-informed decisions.

Sema4, 333 Ludlow Street, Stamford CT 06902

1. Calculated using the National Society of Genetic Counselors (NSGC) NIPT/Cell Free DNA Screening Predictive Value Calculator