

Evolving Interpretation of Screening and Diagnostic Tests in Allergy



Elissa M. Abrams, MD, FRCPC, MPH^a, Edmond S. Chan, MD, FRCPC^b, and Jay Portnoy, MD^c Winnipeg, MB and Vancouver, BC, Canada; and Kansas City, Mo

INFORMATION FOR CATEGORY 1 CME CREDIT

Credit can now be obtained, free for a limited time, by reading the review articles in this issue. Please note the following instructions.

Method of Physician Participation in Learning Process: The core material for these activities can be read in this issue of the Journal or online at the *JACI: In Practice* Web site: www.jaci-inpractice.org/. The accompanying tests may only be submitted online at www.jaci-inpractice.org/. Fax or other copies will not be accepted.

Date of Original Release: December 1, 2021. Credit may be obtained for these courses until November 30, 2022.

Copyright Statement: Copyright © 2021-2023. All rights reserved.

Overall Purpose/Goal: To provide excellent reviews on key aspects of allergic disease to those who research, treat, or manage allergic disease.

Target Audience: Physicians and researchers within the field of allergic disease.

Accreditation/Provider Statements and Credit Designation: The American Academy of Allergy, Asthma & Immunology (AAAAI) is accredited by the Accreditation Council for Continuing Medical

Education (ACCME) to provide continuing medical education for physicians. The AAAAI designates this journal-based CME activity for 1.00 *AMA PRA Category 1 Credit*[™]. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

List of Design Committee Members: Elissa M. Abrams, MD, FRCPC, MPH, Edmond S. Chan, MD, FRCPC, and Jay Portnoy, MD (authors); Michael Schatz, MD, MS (editor)

Learning objectives:

1. To select appropriate diagnostic tests for a particular patient.
2. To interpret the results of diagnostic tests to make the most likely diagnosis.
3. To determine how well a diagnostic test performs relative to other tests for a given diagnosis.

Recognition of Commercial Support: This CME has not received external commercial support.

Disclosure of Relevant Financial Relationships with Commercial Interests: All authors and reviewers reported no relevant financial relationships.

Diagnostic tests for allergy usually are performed to confirm a diagnosis of an allergic disease. If a food allergy suspected, a test can help to determine whether it is present, to monitor its activity over time, and to determine whether the allergy is resolving. In this way, tests are used for diagnosis, monitoring, screening, and prognosis. There are 2 schools of thought for using tests: Frequentist and Bayesian approaches. The Frequentist approach defines probability in terms of the frequency of an event if it were to be repeated numerous times and uses parameters such as sensitivity, specificity, and predictive values to make a diagnosis. In contrast, the Bayesian approach defines probability as the degree of belief or disbelief regarding the diagnosis and asserts that only data are real and that test parameters are to be inferred from the

data. There are strengths and limitations to each approach; however, the Bayesian approach provides an algorithm leading to a disease probability. To use the Bayesian approach, test results need to be expressed as a likelihood ratio. This helps to determine how much the result of a test changes the probability of a particular diagnosis. Once a probability of disease is determined, decision thresholds need to be defined so that a treatment decision can be made. Using this Bayesian approach, the concept of a false-positive or false-negative test result becomes obsolete. © 2021 American Academy of Allergy, Asthma & Immunology (*J Allergy Clin Immunol Pract* 2021;9:4183-91)

Key words: Testing; Sensitivity; Specificity; Likelihood ratio; Bayes theorem

^aDepartment of Pediatrics, Section of Allergy and Clinical Immunology, University of Manitoba, Winnipeg, MB, Canada

^bDivision of Allergy & Immunology, Department of Pediatrics, the University of British Columbia, BC Children's Hospital, Vancouver, BC, Canada

^cSection of Allergy/Immunology, Children's Mercy Hospital, Kansas City, Mo. No funding was received for this work.

Conflicts of interest: E. M. Abrams declares no relevant conflicts of interest. E. S. Chan has received research support from DBV Technologies; has been a member of advisory boards for Pfizer, Padiapharm, Leo Pharma, Kaleo, DBV, AllerGenis, Sanofi Genzyme, Bausch Health, and Avir Pharma; and is a member of the

healthcare advisory board for Food Allergy Canada. J. Portnoy has been a speaker for Thermo Fisher.

Received for publication March 26, 2021; revised May 17, 2021; accepted for publication May 23, 2021.

Corresponding author: Jay Portnoy, MD, Section of Allergy/Immunology, Children's Mercy Hospital, 2401 Gillham Road, Kansas City, MO 64108. E-mail: jportnoy@cmh.edu.

2213-2198

© 2021 American Academy of Allergy, Asthma & Immunology
<https://doi.org/10.1016/j.jaip.2021.05.018>

Abbreviations used

LR- Likelihood ratio
 NPV- Negative predictive value
 OFC- Oral food challenge
 PPV- Positive predictive value
 ROC- Receiver operator characteristics
 sIgE- Specific IgE
 SPT- Skin prick test

Diagnostic and screening tests for allergic conditions generally involve detection and measurement of allergen specific IgE (sIgE) antibodies using either *in vivo* (ie, prick skin test and intracutaneous test) or *in vitro* methods. The presence of sIgE is called sensitization that is a necessary but not sufficient condition for a patient to become symptomatic when he or she is exposed to the allergen. Allergy tests are performed when a patient is suspected to have symptomatic type I allergies either to confirm the presence of sIgE or to monitor disease activity to determine whether the allergy is resolving as might be the case in food allergy.¹ When to perform an allergy test and which test to use depends on the reason for performing the test and on the performance characteristics of the specific test. Interpretation of the test results requires an understanding of what a certain result means and how to use it when making decisions about treatment. In this paper, we will review the characteristics of allergy tests and discuss their uses for diagnosing allergic conditions.

PURPOSE OF TESTS

In general, tests are used for the purpose of diagnosing, monitoring, screening, and estimating a prognosis. For allergic conditions that are mediated by allergen sIgE, tests are used to identify relevant allergens and to determine the probability that they are (and will continue to be) the cause of allergic symptoms.

Diagnosis is the process of finding out whether a patient has a specific disease (eg, food allergy). A diagnosis represents an attempt to classify an individual's condition into categories that permit decisions to be made about treatment. Once a condition has been diagnosed, monitoring can be used to determine how well the disease is controlled. For example, serial measurement of food sIgE is often used to determine the status of a patient's food allergy.²

Screening consists of testing patients who do not (yet) have signs or symptoms of a specific illness to identify it before the patient becomes symptomatic. This is done because treatment is likely to be more effective when started early. Because screening tests are used in large segments of the population, they should be simple and inexpensive. Such tests are designed to be very sensitive so that uncommon conditions can reliably be detected. This is usually followed by a more specific test to rule out the diagnosis in patients who were incorrectly identified with the condition by the first test. An example is newborn screening for congenital severe combined immunodeficiency using T-receptor excision circles.³ In contrast, screening of sIgE for low-prevalence allergic conditions such as food allergy is not recommended because of the high likelihood of false-positive results.⁴

A test for prognosis is used to determine the probability that a patient either will develop a disease in the future or to predict the likely progress of a disorder that is already present. For example, a

patient who initially has a low level of food sIgE that decreases over time has an increased likelihood of outgrowing his or her food allergy.^{5,6}

Regardless of its immediate purpose, a test provides information that can be used to recommend treatment.

FREQUENTIST VERSUS BAYESIAN APPROACH TO TESTING

Currently, there are 2 schools of thought in statistics for making a diagnosis using tests: Frequentist and Bayesian. The difference between these 2 approaches is in the way they interpret probability. To a Frequentist, probability is defined in terms of the frequency of an event if it were to be repeated numerous times. Test parameters are believed to be correct with a known error and the patient's results are compared with them to determine a confidence interval in which the patient's result falls. Therefore, a Frequentist assumes that there are true values for the parameters of a test, and that they can be used to estimate the likelihood that an individual patient's result falls within a certain range.

A Bayesian defines probability as the degree of belief or disbelief regarding the diagnosis. A Bayesian asserts that only data are real and treats the test parameters as probability distributions that are to be inferred from the data. For a Bayesian, there are no true values for test parameters as such, just that some values are more probable than the others. Tests are used to update a prior belief to a subsequent posterior belief using evidence provided by the test.

Frequentist approach

When evaluating the results of an allergy test, whether it be a screening or diagnostic test, it is essential to have a sense of how well that test performs. The validity of a test is its ability to differentiate between individuals who have a disease such as peanut allergy and those who do not. This is usually determined using a confusion matrix (Figure 1). A confusion matrix compares the results of a reference test (often referred to as the "gold standard" test) with the results of an index test that is used as a proxy for the reference test. Although a reference test could reliably classify patients either as having or not having a diagnosis, it is often either too expensive, laborious, unavailable, or risky to use for routine diagnostics.⁷ Most confusion matrices compare dichotomous results (ie, positive or negative) of the reference test to an index test; however, the actual values for a reference test also could be used for this classification instead of categorizing its results.

The parameters of a test are usually described in terms of sensitivity and specificity.⁸ Sensitivity (true positive rate) is the ability of a test to identify individuals who have the disease and test positive, whereas specificity (true negative rate) is the ability of a test to identify individuals who do not have the disease and test negative.⁷ We know that both peanut sIgE and skin prick test (SPT) are highly sensitive (>90%) for the diagnosis of IgE-mediated peanut allergy.⁹⁻¹¹ In contrast, the specificity of both tests is less than 50%.⁶ This means that many children with a positive test will likely tolerate peanut.¹² In a review of 125 children aged 1 to 19 years who underwent an oral food challenge (OFC), 80% to 100% of children who were avoiding foods largely because of a positive SPT or sIgE in the absence of a convincing clinical history were able to eat the food without a reaction. This demonstrates the importance of careful selection of

		Disease (as determined by the reference test)		
Total population		Disease present	Disease absent	Prevalence = $\frac{\sum \text{Disease present}}{\sum \text{Disease present} + \sum \text{Disease absent}}$
Index Test outcome	Test positive	True positive	False positive (Type I error)	Positive predictive value (PPV) = $\frac{\sum \text{True positive}}{\sum \text{Test positive}}$
	Test negative	False negative (Type II error)	True negative	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Test negative}}$
Positive likelihood ratio (LR+) = TPR / FPR		True positive rate (TPR) = $\frac{\sum \text{True positive}}{\sum \text{Disease present}}$	False positive rate (FPR) = $\frac{\sum \text{False positive}}{\sum \text{Disease absent}}$	Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$
Negative likelihood ratio (LR-) = FNR / TNR		False negative rate (FNR) = $\frac{\sum \text{False negative}}{\sum \text{Disease present}}$	True negative rate (TNR) = $\frac{\sum \text{True negative}}{\sum \text{Disease absent}}$	
Diagnostic odds ratio (DOR) = LR+ / LR-		Sensitivity = TPR	Specificity = TNR	

FIGURE 1. Confusion matrix showing various characteristics of a binary diagnostic test. The name stems from the fact that it makes it easy to see if the system is confusing 2 classes (ie, commonly mislabeling one as another).

food SPT and sIgE tests based on history as opposed to testing to a panel of foods, which should be discouraged.^{13,14} Sensitivity and specificity can also be used to determine a test's accuracy that is the proportion of individuals tested whose diagnosis is correctly classified by the test.^{8,15}

Other parameters of a test are its positive predictive value (PPV) and negative predictive value (NPV). The PPV is the percent of patients with a positive result who have the disease, whereas the NPV is the percent of patients with a negative result who do not have the disease.⁸ A limitation of predictive values is that they are dependent on the prevalence of the disease. It has been noted that predictive values need to be "determined on the basis of careful clinical deliberation"¹⁵ and that ultimately in the field of food allergy the "best test" that we truly have is our own clinical judgment.^{7,12}

Assume, for example, that a patient presents for evaluation of possible peanut allergy. In the absence of other information such as what happens with peanut ingestion, what is the probability that this patient will have a systemic reaction if peanut is ingested? Figure 2 shows a confusion matrix from a study by Beyer et al¹⁶ of peanut allergy that compared the results of oral challenge (reference test) with peanut sIgE (index test) using a cutoff of 0.35 kU/L for a positive result (matrix A). From this matrix we can conclude that 86% of patients with peanut allergy had a positive test (sensitivity) and that 86% of patients without peanut allergy had a negative test (specificity). In addition, 82% of patients with a positive test had peanut allergy (PPV) and 86% with a negative test did not have peanut allergy (NPV). Because our patient has a peanut sIgE >0.35 kU/L, you might conclude from this that the patient has an 82% probability of having peanut allergy, which would be correct for patients in the Beyer study but incorrect for patients in the general population.

The problem with this approach is that the overall prevalence of peanut allergy is only approximately 1.5%,¹⁷ whereas the

prevalence of peanut allergy among subjects in the Beyer study was 42.9%.¹⁶ This means that of 210 patients in the general population, we would expect to find approximately 3 with peanut allergy instead of the 90 found in the Beyer study. A confusion matrix that has the same sensitivity and specificity but with prevalence-adjusted rates for peanut allergy is shown in Figure 2, matrix B. This low prevalence of peanut allergy in the general population explains why there are so many false-positive tests when screening is done in an unselected population, as predictive values are impacted by disease prevalence. In other words, use of an allergy test on an unselected population will overestimate the number of patients with the allergy. To answer the question posed above, an unselected patient with a positive peanut sIgE only has a 3.7% probability of having peanut allergy as opposed to the 82% probability based on the unadjusted PPV.

For most *in vitro* allergy tests, the cutoff for a positive test is set at 0.35 kU/L. Historically, this was the level of detection of the test rather than a clinically proven best value for making a diagnosis. If a different cutoff is used to categorize positive and negative test results, the test parameters will change. The test can be made more sensitive by reducing the cutoff and more specific by increasing it, as shown in Table I.¹⁸ A cutoff could be selected to maximize the number of correctly classified patients, and a different cutoff could be selected to maximize a test's sensitivity to screen for a condition that you do not want to miss such as anaphylaxis after peanut ingestion. A curve that represents the trade-off between false positives (1–specificity) and true positives (sensitivity) is called a receiver operator characteristics (ROC) curve (Figure 3). It is called that because it was originally developed for military radar receiver operators to classify signals during World War II.

There are limitations to the data we have on the parameters of food allergy tests that would be present if the reference test (in

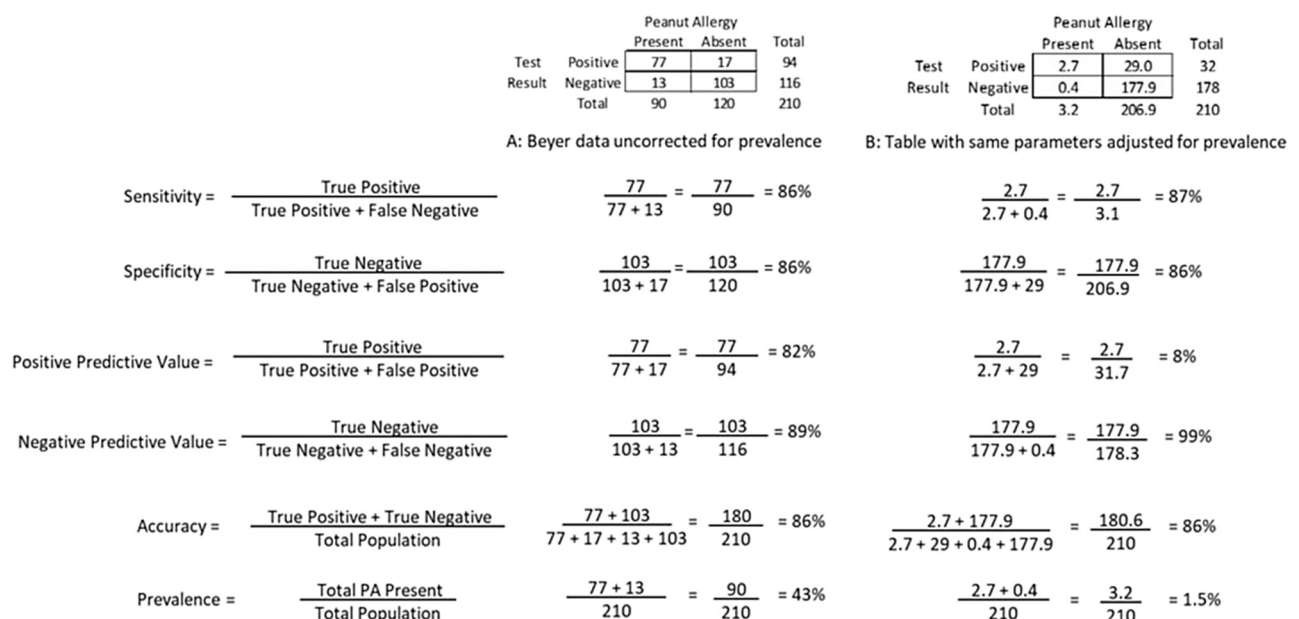


FIGURE 2. Confusion matrix of test parameters for data from the Beyer study (matrix A) and a confusion matrix with the same sensitivity and specificity but adjusted to have a disease prevalence of 1.5%. This illustrates the importance of accounting for prevalence when interpreting test results.

this case, an OFC) were used.⁷ For example, factors that can influence the results of SPT include the location of the test on a patient's body, the quality of the extracts used, and potential suppression by medications. Results of sIgE can depend on which assay is used, the machine's calibration, technician training and expertise, and the quality of the solid phase allergen.¹⁹ In addition, because certain populations may differ from others in their response to a test, generalization from one population to another must be done with caution.

Bayesian approach

Logistic regression to categorize test results is a Frequentist approach. It assumes that there are enough measurements in a sample of subjects to infer something meaningful about a general population and that this can be applied to individual patients. The Frequentist approach also assumes that the measurements are representative of the general population, which is why it is important to take into account the prevalence of the disease in the studied population when interpreting test results. In the Bayesian approach, test results are supplemented with additional information in the form of a prior probability distribution. A prior belief about the population is combined with the likelihood function of the test according to the Bayes theorem to yield a posterior belief about an individual patient's diagnosis.

The Bayesian approach relies on a sequential series of steps that begin with an initial belief about the probability of a diagnosis that, to continue with our example, could be a systemic reaction after peanut ingestion. This is called the prior or pretest odds. In the absence of additional information such as what happens if peanut is eaten, the pretest odds is simply the prevalence of the disease in a population that is similar to the patient being tested.²⁰ For small values (ie, <5%), the probability and odds are approximately the same. To a Bayesian, a diagnostic test

updates the prior odds using evidence provided by the test. The amount of change provided by the test is determined by the Bayes factor or likelihood ratio (LR). Unlike predictive values, LRs are not affected by the disease prevalence and therefore can be used to adopt the results from other investigators to your own patient population.

Binary tests have a specific cutoff. Therefore, any particular test result usually is expressed either as positive or negative depending on where it falls relative to the cutoff (Figure 4). The LR is expressed either as LR+ for a positive test or LR− for a negative test. The most informative way to determine the LR for a particular value of the test is to set the cutoff at that value and then to determine the sensitivity and specificity of the test. This LR is the slope of the ROC curve that is sensitivity/(1−specificity). Most studies evaluate test parameters at a single cutoff (eg, 0.35 kU/L) that can lead to significant differences for the LR if the observed test result differs substantially from the cutoff that was studied.

Bayesian analysis can be done by following the steps below (shown with more detail in Figure 5). We will illustrate the process using the results from the Nicolaou study shown in Table 1.¹⁸

- (1) Determine the pretest probability. If no information, such as what happens with peanut ingestion, is available, then the prevalence of the condition can be used for this value. For peanut allergy, the prevalence is 1.5%. This serves as the prior belief that is to be updated with evidence provided by the test.
- (2) Convert the pretest probability into pretest odds. This is necessary because probability is a fraction between 0 and 1 and odds is a ratio of occurrence to nonoccurrence that can be multiplied by the LR to get posttest odds.

TABLE I. Sensitivity, specificity, and percent correctly classified for various cutoffs of peanut sIgE*

Antigen	Cutoff (kU/L)	Sensitivity (%)	Specificity (%)	Correctly classified (%)	Likelihood ratio
Peanut	0.35	96.55	26.92	51.85	1.32
	3.91	79.31	84.62	82.72	5.16
	5.00	75.86	90.38	85.19	7.89
	5.30	75.86	94.23	87.65	13.15
	5.96	72.41	94.23	86.42	12.55
	7.81	72.41	96.15	87.65	18.81
	15.00	55.17	96.15	81.48	14.33
Ara h2	43.86	34.85	98.08	75.31	18.15
	0.30	100.00	90.20	93.75	10.20
	0.32	100.00	94.12	95.00	17.01
	0.35	100.00	96.08	97.50	25.51
	0.38	96.55	96.08	96.25	24.63
	0.40	93.10	98.04	96.25	47.50
	0.55	93.10	100.00	97.50	ND
	0.87	89.66	100.00	96.25	ND

ND, Not defined; sIgE, specific IgE.

*Likelihood ratios were calculated using the formula $LR = \text{sensitivity} / (1 - \text{specificity})$.¹⁸

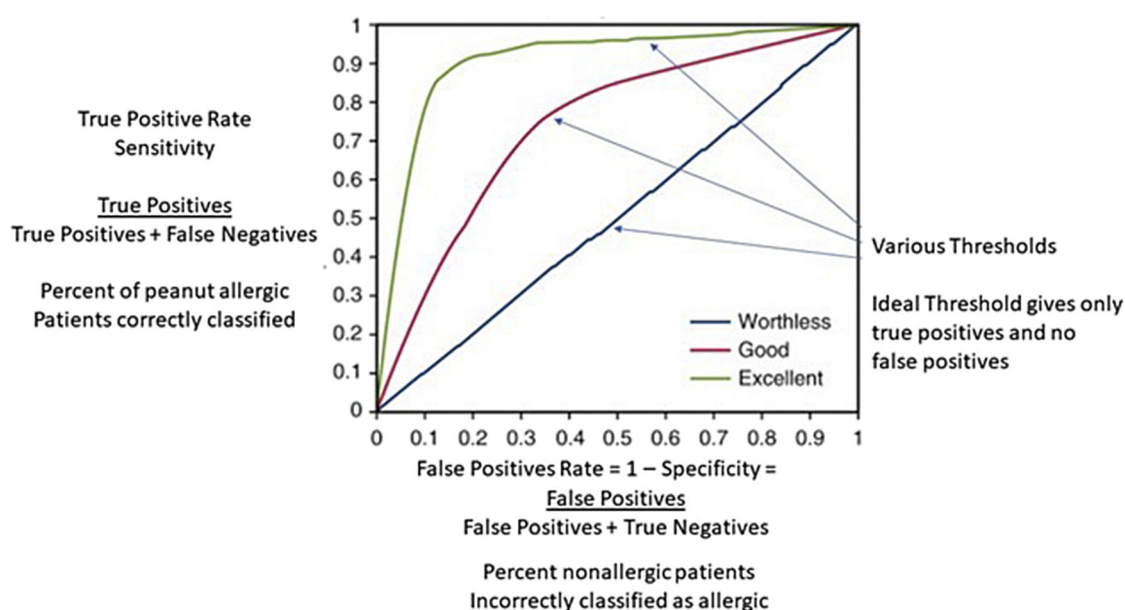


FIGURE 3. Characteristics of a receiver operator characteristics curve.

- (3) Determine the LR (otherwise known as the Bayes factor) for the test result. The LR is calculated from the sensitivity and specificity of the test, and it does not depend on prevalence in the reference group, or on the pretest probability. The LR tells you how much more (or less) likely a patient with the given test result is to have the disease.
- (4) Determine the posttest odds. This is the product of the pretest odds with the LR.
- (5) Determine the posttest probability by converting from odds to probability.

The primary limitation to this approach is the need for a prior or pretest probability. Although an estimate of the disease prevalence is a reasonable starting place, a closer approximation

would account for prevalence differences based on geographic, ethnic, age, sex, and other factors. In addition, it is unclear whether the history of reaction should be used to estimate the prior probability or whether it should be associated with an LR. Another limitation is the need to determine LRs for different cutoff values of sIgE.

Reflex testing

Component testing provides an opportunity to refine diagnostic sensitivity and specificity for food allergy. One example is Ara h2-sIgE as a means of peanut allergy diagnosis. A recent 2020 practice parameter update with a GRADE analysis noted that Ara h2 provides the highest diagnostic accuracy for peanut

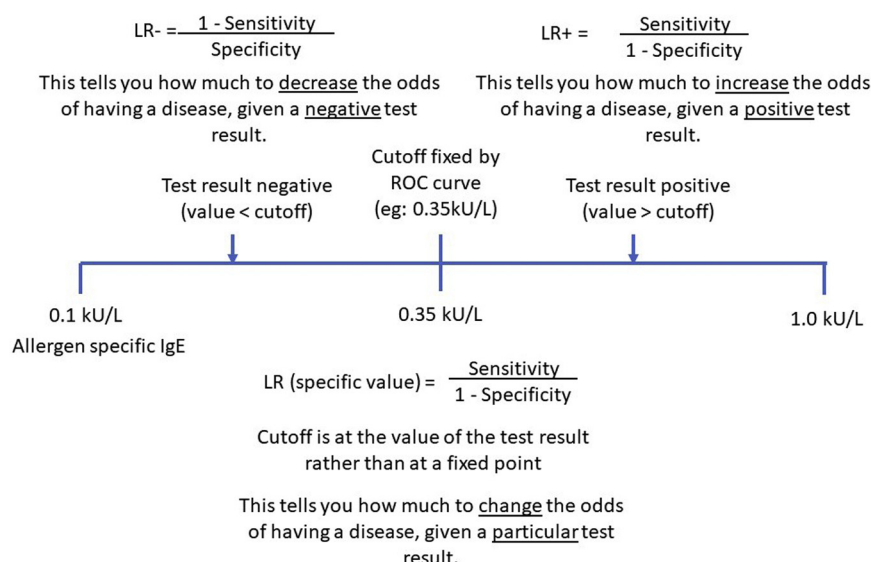


FIGURE 4. How LR+ and LR- differ from the LR for various cutoff values of the test. *LR*, Likelihood ratio; *ROC*, receiver operator characteristics.

allergy with a higher specificity (thereby reducing false-positive cases).¹⁴ However, this practice parameter also noted that although Ara h2 has the highest specificity, it has the lowest sensitivity compared with SPT and peanut sIgE, and therefore if pretest probability is high, any of the first-line screening measures could be used. In addition, it is possible for a patient to have sIgE to Ara h1 and Ara h3 but not to Ara h2. Such individuals would be missed if Ara h2 alone were used as the initial test for peanut allergy. This is accounted for in the sensitivity and specificity of the test when measured against the gold standard of an OFC.

Although screening for peanut allergy in general is not recommended, recent literature suggests that if screening to peanut is done, sIgE for Ara h2 may have a larger LR than sIgE for whole peanut. In a recent study of infants with moderate-to-severe eczema, history of food allergy or immediate family history of food allergy, Ara h2 had the highest specificity (98%) compared with peanut sIgE (86%) or SPT (88%) although once again had a lower sensitivity.²¹

A more efficient and cost-effective way to implement this sequence of tests is with reflex testing.²² A reflex test is performed when an initial test result meets predetermined criteria (eg, positive or outside normal parameters), and the test result is inconclusive without the reflex or follow-up test. It is performed automatically without the intervention of the ordering physician. For example, if total peanut sIgE is <0.10 kU/L or very low, it is unlikely that Ara h2-sIgE would be elevated so there is no reason to perform the test.²³ An elevated value for peanut sIgE could be followed by measurement of Ara h2-sIgE to provide a larger LR. Both Quest and Labcorp offer an ability to order component-resolved diagnostics reflexively in the United States, and both will customize such tests for individual clients.

In practice, one could use the posttest odds resulting from an initial test as the pretest odds for another independent test. For example, if the patient in the previous example had a peanut IgE of 7.8 kU/L and an Ara h2 of 0.35 kU/L, the LRs for these tests from Table 1 would be 18.8 and 25.5, respectively. The combined LR for both tests would be $18.8 \times 25.5 = 480$. This

means that a patient with these results is 480 times more likely to have peanut allergy than to not have peanut allergy.

THE FUTURE OF TEST REPORTING AND INTERPRETATION

Currently, test reports are formatted to provide the results of the test, a normal range (usually 95% confidence interval), and a flag for abnormal results (Figure 6, A). This leads to what we refer to as the “red circle phenomenon” in which a provider circles the “abnormal” results with a red pen and tells the patients that they have a diagnosis based on the abnormal flag in the absence of supporting clinical information. This is something even the most well-meaning clinician tends to do (personal observation). The problem with this approach is that it does not account for the prior probability of the diagnosis that might not be readily available. In addition, it does not provide an estimate of how probable the diagnosis is, and it does not necessarily lead to the most effective recommendation for treatment. Simply circling “abnormal” values can lead to misdiagnosis and to unnecessary additional tests.

For those reasons, we recommend that reports should be formatted or at least supplemented with formatting as shown in Figure 6, B. Here the test result is accompanied by the LR for the result. This tells the clinician how much more (or less) likely the patient is to have a particular diagnosis given the test result. Because some test results can be associated with various diagnoses, it may be necessary to provide a separate LR for each diagnosis. For example, LRs for food sIgE are likely to be different for anaphylaxis where they are highly predictive, than for predicting atopic dermatitis where they are poorly predictive. Because most studies evaluate test parameters for a single cutoff, the LRs are usually expressed as LR+ for a positive test and LR- for a negative test. Such LRs have been described for SPT and *in vitro* tests for aeroallergens that trigger rhinitis²⁵; however, a consensus is

To determine the probability of PA if the peanut sIgE = 7.8 kU/L, following these steps:

1. Determine the pre-test probability. If no information is available, then the prevalence of the condition can be used. For peanut allergy the prevalence is 1.5%.
2. Convert the pretest probability into pretest odds:

$$\text{Pretest Odds} = \frac{\text{Probability of having the disease}}{\text{Probability of not having the disease}} = \frac{P(\text{peanut allergy})}{1 - P(\text{peanut allergy})} = \frac{1.5}{98.5} = 0.015$$

3. Determine the likelihood ratio for the test result:

For a binary test, the LR can be determined from sensitivity and specificity. Ideally, these would be determined with a cutoff equal to the test result, however, most studies only evaluate LR at a cutoff of 0.35 kU/L. Table 1 shows sensitivity and specificity for various cutoffs including 7.8 kU/L so we can use those results as follows:

$$\frac{\text{Likelihood Ratio}}{\text{(Bayes' Factor)}} = \frac{P(\text{Test result} | \text{Patients with PA})}{P(\text{Test result} | \text{Patients without PA})} = \frac{\text{True Positive Rate}}{\text{False Positive Rate}} = \frac{\text{Sensitivity}}{1 - \text{Specificity}} = \frac{72.4\%}{96.1\%} = 18.8$$

4. Determine the posttest odds:

$$\text{Posttest Odds} = \text{Likelihood ratio} \times \text{Pretest Odds} = 18.8 \times 0.015 = 0.286$$

5. Determine the posttest probability

$$\text{Posttest Probability} = \frac{\text{Posttest Odds}}{1 + \text{Posttest Odds}} = \frac{0.286}{1 + 0.286} = \frac{0.286}{1.286} = 22.3\%$$

6. The posttest odds can also be used as the pretest odds for another independent test. For example, if the patient had an Ara h2 of 0.35, Table 1 shows that the LR would be 25.5. Bayesian analysis with 2 tests would give a posttest probability of 88% as follows:

$$\text{Odds (posterior)} = \text{Odds (prior)} \times \text{LR1 (peanut)} \times \text{LR2 (Ara h2)}$$

	Test Value	Pretest Probability	Pretest Odds	LR	Posttest odds	Posttest probability
Peanut IgE	7.8 kU/L	1.5%	0.015	18.8	0.286	22.3%
Ara h2	0.35 kU/L	22.3%	0.286	25.5	7.301	88.0%

FIGURE 5. The procedure for determining the probability of disease for a given test result using the Bayes theorem. LR, Likelihood ratio; PA, peanut allergy; sIgE, specific IgE.

needed for other allergens such as foods. A better approach would be to determine an LR for each value of the test as was described in a recent paper by Nalin et al.¹² Although a consensus does not yet exist regarding such LRs, we consider this to be an aspirational goal that merits consideration. This, along with a prior probability based either on the patient's

history or on the prevalence in the population, can be used to determine the posttest probability of disease.

Although these calculations can be complicated to perform, they could be automated using decision assistance algorithms implemented in an electronic medical record. Until then, an estimate of the probability of a diagnosis using the pretest

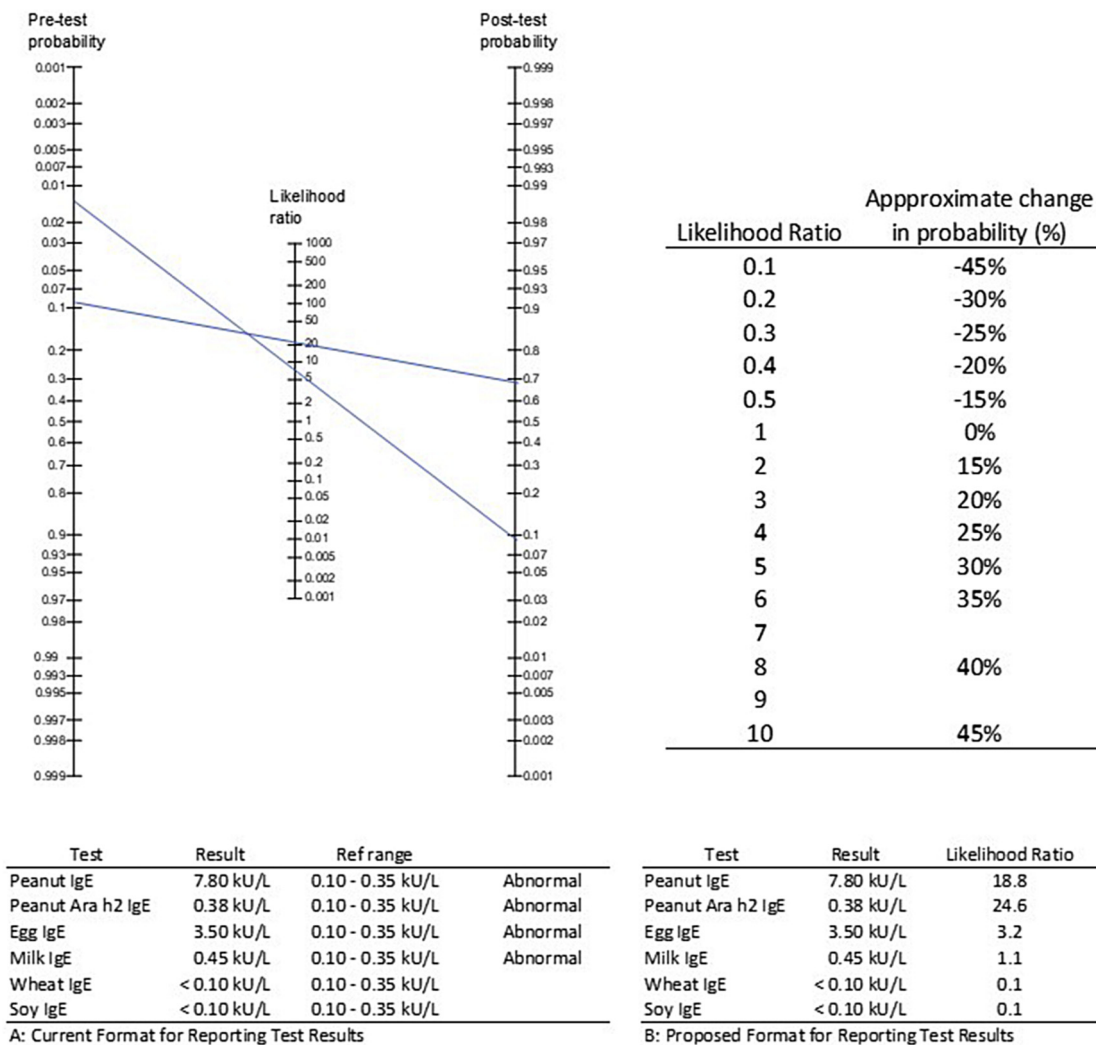


FIGURE 6. Current (Table A) and proposed (Table B) format for reporting test results for use with a Fagan nomogram.²⁴ Note: the LRs in the right lower pane for peanut and peanut Ara h2 are from Table 1. The other LRs are hypothetical and are used to demonstrate the principle. Note that LRs that are <1 are associated with a lower odds of food allergy, whereas LRs that are >1 are associated with an increased odds of food allergy. *LR*, Likelihood ratio.

probability and likelihood ratio can be made using a Fagan nomogram (Figure 6).²⁴ In addition, a simplified method for using LRs to estimate the change in disease probability has been described by McGee.²⁶ This method uses a table in which LRs between 0.1 and 0.5 are associated with changes in disease probability of -45 to -15, respectively, and LRs between 2 and 10 with changes in disease probability of +15 to +45, respectively. An important issue that remains is to determine how probable the diagnosis of a disease needs to be for different recommendations to be made. This will require further research into decision analysis.

CONCLUSION

Diagnostic testing can help to make a diagnosis and to monitor the disease activity. Test reports that provide a test's result, 95% confidence interval (so-called "normal" range), and a flag for abnormal values do not easily permit the clinician to estimate the probability of disease, nor do they help to identify the best treatment. An alternative approach is

a report to the clinician that includes the LR for each value of the test and a Fagan nomogram. Studies need to be designed to measure and analyze LRs to determine these values. Consensus meta-analyses of these studies can then be used to inform test reports. Once this is done, it is still necessary to determine what ranges of probabilities are associated with treatments that can provide the best outcome for patients. All of this will likely be made possible using automated clinical decision assistance algorithms in future EHRs. To enable this, additional research is needed. By using the Bayesian approach to estimate the probability of disease, the concept of false-positive and -negative test results hopefully will become obsolete. In other words, a test could stand on its own and would be interpreted based on likelihood or probability of a diagnosis.

REFERENCES

1. Sampson HA, Aceves S, Bock SA, James J, Jones S, Lang D, et al. Food allergy: a practice parameter update-2014. *J Allergy Clin Immunol* 2014;134:1016-1025.e43.

2. Sicherer SH, Sampson HA. Food allergy: a review and update on epidemiology, pathogenesis, diagnosis, prevention, and management. *J Allergy Clin Immunol* 2018;141:41-58.
3. van der Spek J, Groenwold RH, van der Burg M, van Montfrans JM. TREC based newborn screening for severe combined immunodeficiency disease: a systematic review. *J Clin Immunol* 2015;35:416-30.
4. Abrams EM, Brough HA, Keet C, Shaker MS, Venter C, Greenhawt M. Pros and cons of pre-emptive screening programmes before peanut introduction in infancy. *Lancet Child Adolesc Health* 2020;4:526-35.
5. Longo G, Berti I, Burks AW, Krauss B, Barbi E. IgE-mediated food allergy in children. *Lancet* 2013;382:1656-64.
6. Abrams EM, Sicherer SH. Diagnosis and management of food allergy. *CMAJ* 2016;188:1087-93.
7. Sicherer SH, Wood RA. Advances in diagnosing peanut allergy. *J Allergy Clin Immunol Pract* 2013;1:1-13;quiz 4.
8. Celentano D, Szklo M. *Gordis Epidemiology*. 6th ed. Philadelphia: Elsevier; 2014.
9. Cox L, Williams B, Sicherer S, Oppenheimer J, Sher L, Hamilton R, et al. Pearls and pitfalls of allergy diagnostic testing: report from the American College of Allergy, Asthma and Immunology/American Academy of Allergy, Asthma and Immunology Specific IgE Test Task Force. *Ann Allergy Asthma Immunol* 2008;101:580-92.
10. Kattan JD, Sicherer SH. Optimizing the diagnosis of food allergy. *Immunol Allergy Clin North Am* 2015;35:61-76.
11. Sicherer SH, Sampson HA. Food allergy. *J Allergy Clin Immunol* 2010;125(Suppl 2):S116-25.
12. Nalin UG, Shroba J, Pandya A, Portnoy J. Diagnosis of peanut allergy using continuous likelihood ratios. *Ann Allergy Asthma Immunol* 2020;125:680-5.
13. Fleischer DM, Bock SA, Spears GC, Wilson CG, Miyazawa NK, Gleason MC, et al. Oral food challenges in children with a diagnosis of food allergy. *J Pediatr* 2011;158:578-583.e1.
14. Greenhawt M, Shaker M, Wang J, Oppenheimer JJ, Sicherer S, Keet C, et al. Peanut allergy diagnosis: a 2020 practice parameter update, systematic review, and GRADE analysis. *J Allergy Clin Immunol* 2020;146:1302-34.
15. Trevelyan R. Sensitivity, specificity, and predictive values: foundations, plabilities, and pitfalls in research and practice. *Front Public Health* 2017;5:307.
16. Beyer K, Grabenhenrich L, Hartl M, Beder A, Kalb B, Ziegert M, et al. Predictive values of component-specific IgE for the outcome of peanut and hazelnut food challenges in children. *Allergy* 2015;70:90-8.
17. Venter C, Maslin K, Patil V, Kurukulaaratchy R, Grundy J, Glasbey G, et al. The prevalence, natural history and time trends of peanut allergy over the first 10 years of life in two cohorts born in the same geographical location 12 years apart. *Pediatr Allergy Immunol* 2016;27:804-11.
18. Nicolaou N, Murray C, Belgrave D, Poorafshar M, Simpson A, Custovic A. Quantification of specific IgE to whole peanut extract and peanut components in prediction of peanut allergy. *J Allergy Clin Immunol* 2011;127:684-5.
19. Oppenheimer J, Durham S, Nelson H, Wolthers O. Allergy diagnostic testing. World Allergy Organization; 2014. Updated July 2014. Available at: <https://www.worldallergy.org/education-and-programs/education/allergic-disease-resource-center/professionals/allergy-diagnostic-testing>. Accessed June 7, 2021.
20. Shaker M, Hsu Blatman K, Abrams EM. Engaging patient partners in state-of-the-art allergy care: finding balance when discussing risk. *Ann Allergy Asthma Immunol* 2020;125:252-61.
21. Keet C, Plesa M, Szelag D, Shreffler W, Wood R, Dunlop J, et al. Ara h 2-specific IgE is superior to whole peanut extract-based serology or skin prick test for diagnosis of peanut allergy in infancy. *J Allergy Clin Immunol* 2021;147:977-983.e2.
22. Rubinstein M, Hirsch R, Bandyopadhyay K, Madison B, Taylor T, Ranne A, et al. Effectiveness of practices to support appropriate laboratory test utilization: a laboratory medicine best practices systematic review and meta-analysis. *Am J Clin Pathol* 2018;149:197-221.
23. Leo SH, Dean JM, Jung B, Kuzeljevic B, Chan ES. Utility of Ara h 2 sIgE levels to predict peanut allergy in Canadian children. *J Allergy Clin Immunol Pract* 2015;3:968-9.
24. Fagan TJ. Letter: Nomogram for Bayes theorem. *N Engl J Med* 1975;293:257.
25. Gendo K, Larson EB. Evidence-based diagnostic strategies for evaluating suspected allergic rhinitis. *Ann Intern Med* 2004;140:278-89.
26. McGee S. Simplifying likelihood ratios. *J Gen Intern Med* 2002;17:646-9.