What do all these numbers mean? Sensitivity, specificity, and likelihood ratios

Steve Simon

P.Mean Consulting

www.pmean.com

2. Why do I offer this webinar for free?

I offer free statistics webinars partly for fun and partly to build up goodwill for my consulting business,

- www.pmean.com/consult.html.

I also provide a free newsletter about Statistics, The Monthly Mean. To sign up for the newsletter, go to

- www.pmean.com/news

3. The next free webinars

- “Jumpstart statistics: Simple descriptive analyses” UMKC only!
  - Friday, March 5, noon-1pm, CST.
- “Putting it all together: meta-analyses and systematic overviews” Open to all!
  - Wednesday, March 31, 11am-noon, CST.

4. Abstract

- This one hour training class will give you a general introduction to numeric summary measures for diagnostic testing. You will learn how to distinguish between a diagnostic test that is useful for ruling in a diagnosis and one that is useful for ruling out a diagnosis. You will also see an illustration of how prevalence of disease affects the performance of a diagnostic test.

5. Objectives

Objectives: In this class you will learn how to:

- compute sensitivity and specificity;
- identify the problems with diagnosing a rare disease;
- understand which tests are useful for ruling in or ruling out a disease.

6. Sources

Part of the material for this webinar comes from:

- Stats #21: What Do All These Numbers Mean? Sensitivity and Specificity
  - www.childrens-mercy.org/stats/training/hand21.asp
- Stats #24: What Do All These Numbers Mean? Likelihood Ratios
  - www.childrens-mercy.org/stats/training/hand24.asp
- An introduction to diagnostic testing
  - www.childrensmercy.org/stats/diagnostic.asp

7. Outline of this talk

1. Very bad joke
2. Pop quiz (no grades handed out)
3. What is a diagnostic test
4. Definitions of sensitivity and others
5. Performance under low prevalence
6. Likelihood ratios
7. Repeat of pop quiz

8. Very bad joke

There are three types of statisticians in the world...

9. Very bad joke

There are three types of statisticians in the world...

those who can count,

10. Very bad joke

There are three types of statisticians in the world...

those who can count,

and those who can’t.

11. Pop quiz #1

The denominator for sensitivity is

1. All patients who are disease-free
2. All patients who have the disease
3. All patients who test negative
4. All patients who test positive

12. Pop quiz #2

If a disease has very low prevalence, you typically get poor values for

1. Negative predictive value
2. Positive predictive value
3. Sensitivity
4. Specificity

13. Pop quiz #3

A diagnostic test is useful for “ruling out” a disease when:

1. Sensitivity is large
2. Sensitivity is small
3. Specificity is large
4. Specificity is small

14. What is a diagnostic test?

A diagnostic test is a procedure which gives a

- rapid,
- convenient, and/or
- inexpensive

indication of whether a patient has a certain disease.

15. What is a diagnostic test?

Example: A standard electrocardiogram can produce a measure called QTc dispersion. In a study of 49 patients with peripheral vascular disease (Darbar 1996), all were assessed for their QTc dispersion values. These patients were then followed for 52 to 77 months. During this time, there were 12 cardiac deaths, 3 non-cardiac deaths, and 34 survivors. A value of QTc dispersion of 60 ms or more did quite well in predicting cardiac death.

16. What is a diagnostic test?

Example: The Yale-Brown obsessive/compulsive scale, a simple yes/no answer to the following question: Do you often feel sad or depressed? In a study of stroke patients at the Royal Liverpool and Broadgreen University Hospitals (Watkins 2001), this test was shown to perform well compared to a more complex measure, the Montgomery Asberg depression rating scale.

17. What is a diagnostic test?

Example: Patients with rectal bleeding will sometimes develop colorectal cancer. In a study at a network of practices in Belgium (Wauters 2000), 386 patients presented with rectal bleeding between 1993 and 1994. After following these patients for 18 to 30 months, only a few developed colorectal cancer.

18. What is a diagnostic test?

To assess the quality of a diagnostic test, you need to compare it to a gold standard. This is a measurement that is

- slower,
- less convenient,
- or more expensive

than the diagnostic test, but which also gives a definitive indication of disease status. The gold standard might involve invasive procedures like a biopsy or could mean waiting for several years until the disease status becomes obvious.

19. What is a diagnostic test?

The crossclassification of the diagnostic test and the gold standard leads to four possible categories.

- TP (true positive) = # who test positive and who have the disease,
- FN (false negative) = # who test negative and who have the disease,
- FP (false positive) = # who test positive and who are healthy, and
- TN (true negative) = # who test negative and who are healthy.

	Test Positive (T+)	Test Negative (T-)
Disease Present (D+)	True Positive (TP)	False Negative (FN)
Disease Absent (D-)	False Positive (FP)	True Negative (TN)

20. What is sensitivity?

The sensitivity (Sn) of a test is the probability that the test is positive when given to a group of patients with the disease. Notice that the denominator for sensitivity is the number of patients who have the disease.

	Test Positive (T+)	Test Negative (T-)
Disease Present (D+)	True Positive (TP)	False Negative (FN)
Disease Absent (D-)	False Positive (FP)	True Negative (TN)

Sensitivity (Sn) = TP / (TP + FN)

21. What is sensitivity?

A large sensitivity means that a negative test can rule out the disease. David Sackett coined the acronym "SnNOut" to help us remember this. Ruling out a disease is very important if the cost of treating a disease is low, but leaving a diseased patient untreated is very bad. Example: cervical fractures in patients reporting to ER.

22. What is sensitivity?

Example: In a study of 5,113 subjects checked for gastric cancer by endoscopy (Kitahara 1999), serum pepsinogen concentrations were also measured. A pepsinogen I concentration of less than 70 ng/ml and a ratio of pepsinogen I to pepsinogen II of less than 3 was considered a positive test. There were 13 patients with gastric cancer confirmed by endoscopy. 11 of these patients were positive on the test. The sensitivity is 11/13 = 85%.

23. What is specificity?

The specificity of a test is the probability that the test will be negative among patients who do not have the disease. Notice that the denominator for specificity is the number of healthy patients.

	Test Positive (T+)	Test Negative (T-)
Disease Present (D+)	True Positive (TP)	False Negative (FN)
Disease Absent (D-)	False Positive (FP)	True Negative (TN)

Specificity (Sp) = TN / (TN + FP)

24. What is specificity?

A large specificity means that a positive test can rule in the disease. David Sackett coined the acronym "SpPIn" to help us remember this. Ruling in a disease is very important if the cost of treating a disease is high, but leaving the diseased patient untreated is not so bad. Example: surgery for apendicitis.

25. What is specificity?

Example: In a study of the urine latex agglutination test (reference misplaced, sorry!), children were tested for H. influenzae using blood, urine, cerebrospinal fluid, or some combination of these. Of all the children tested, 1,352 did not have H. influenzae in any of these fluids. Only 9 of these patients tested positive on the urine latex agglutination test, the remaining 1,343 tested negative. The specificity is 1343 / 1352 = 99.3%.

26. What is positive predictive value?

The positive predictive value (PPV) of a test is the probability that the patient has the disease when restricted to those patients who test positive.

		Test Positive (T+)	Test Negative (T-)
Disease Present (D+)		True Positive (TP)	False Negative (FN)
Disease Absent (D-)		False Positive (FP)	True Negative (TN)

Positive Predictive Value (PPV) = TP / (TP + FP)

27. What is positive predictive value?

Do not calculate the positive predictive value on a sample where the prevalence of the disease was artificially controlled. For example, the PPV is meaningless in a study where you artificially recruited healthy and diseased patients in a one to one ratio.

28. What is positive predictive value?

Example: In a study of patients in a network of sentinel practices in Belgium (Wauters 2000), 386 patients presented with rectal bleeding. These patients were followed from 18 to 30 months and 27 of them developed colorectal cancer. The positive predictive value for rectal bleeding is 27 / 386 = 7%.

29. What is negative predictive value?

The negative predictive value (NPV) of a test is the probability that the patient will not have the disease when restricted to those patients who test negative.

	Test Positive (T+)	Test Negative (T-)
Disease Present (D+)	True Positive (TP)	False Negative (FN)
Disease Absent (D-)	False Positive (FP)	True Negative (TN)

Negative Predictive Value (NPV) = TN / (TN + FN)

30. What is negative predictive value?

Do not calculate the negative predictive value on a sample where the prevalence of the disease was artificially controlled. For example, the NPV is meaningless in a study where you artificially recruited healthy and diseased patients in a one to one ratio.

31. What is negative predictive value?

Example: In a study of depression among 79 patients hospitalized for stroke (Watkins 2001), 34 patients responded "no" to the question: Do you often feel sad or depressed? Among these 34 patients who tested negative, 6 had clinical depression as defined by a more complex measure, the Montgomery Asberg depression rating scale. Since 28 did not have depression, the negative predictive value is 28 / 34 = 82%.

32. Performance under low prevalence

Prevalence plays a large role in determining how effective a diagnostic test is. In general, when the prevalence of the disease you are testing is rare, it becomes harder to positively diagnose that disease.

Let's look at a hypothetical situation. In the graph shown on the next page, patients on the left have the disease and patients on the right are healthy.

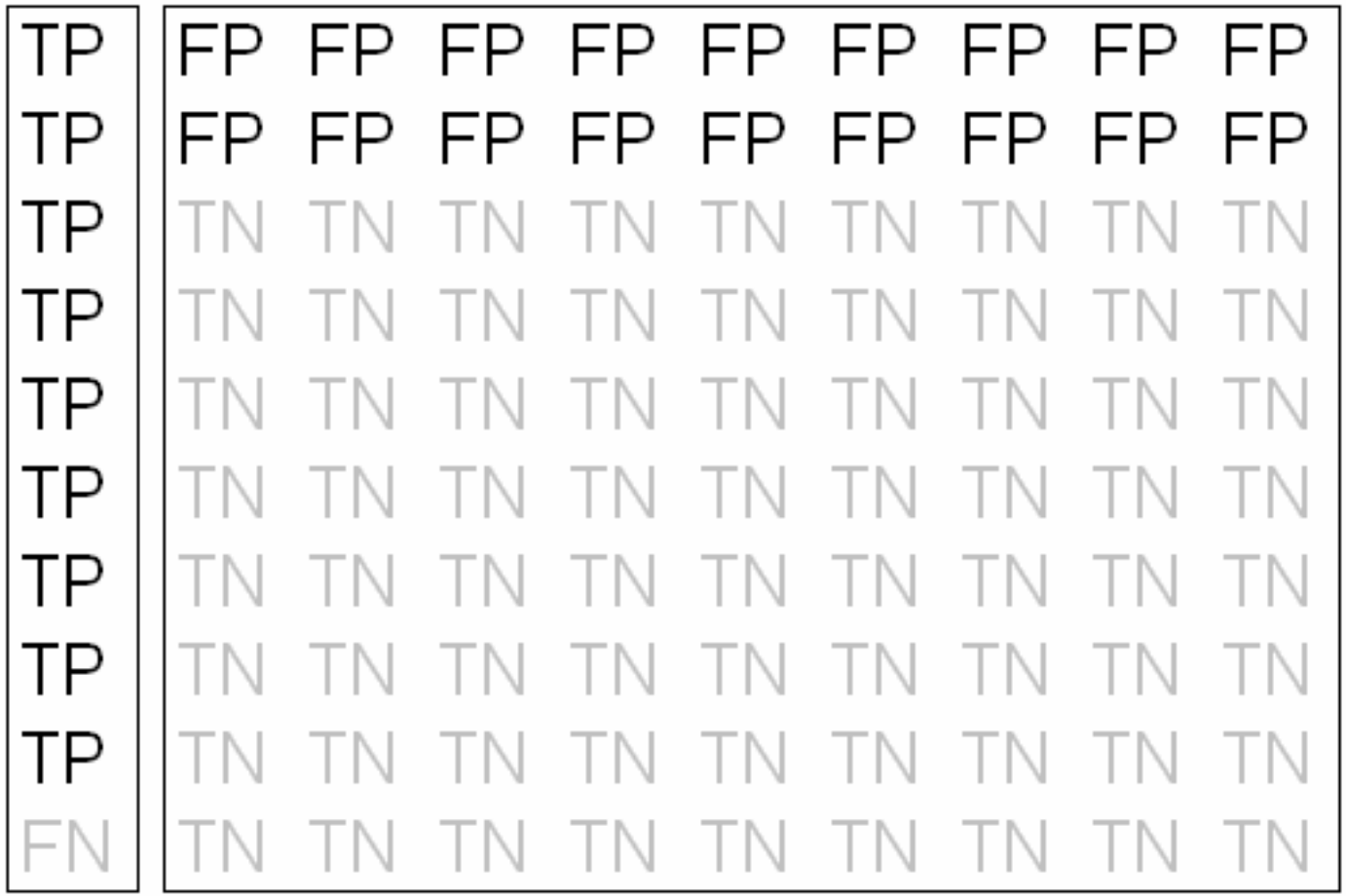
33. Performance under low prevalence

34. Performance under low prevalence

It is easy to see in this example that Sn=90%, Sp=80%. The results are not perfect, but they are pretty good.

In this example, prevalence of the disease is 40%. What happens if the prevalence drops to 10%?

35. Performance under low prevalence



36. Performance under low prevalence

In this example, the number of false positives swamps the number of true positives. This tells you that the Positive Predictive Value

- PPV = TP/(TP+FP)

will suffer in a low prevalence condition.

37. Likelihood ratios

You can summarize information about the diagnostic test itself using a measure called the likelihood ratio. The likelihood ratio combines information about the sensitivity and specificity. It tells you how much a positive or negative result changes the likelihood that a patient would have the disease.

38. Likelihood ratios

The positive likelihood ratio is

- $LR^+ = Sn / (1 - Sp)$.

You want to see a large value for LR+. The simplest way to make this fraction large is to make the denominator small, or make Sp close to one. This is consistent with the David Sackett acronym SpPIn (if the specificity of test is large, then a positive test will help rule in the diagnosis.

39. Likelihood ratios

The negative likelihood ratio is

- $LR^- = (1 - Sn) / Sp$.

You want to see a small value for LR-. The simplest way to make this happen is to make the numerator small, or make Sn close to 1. This is consistent with the David Sackett acronym SnNOut (if the sensitivity of a test is large, then a negative test will help rule out the diagnosis).

40. Likelihood ratios

What's a good value for a likelihood ratio?

There are no absolute boundaries, but here are some general rules. For a positive likelihood ratio, anything less than 2 is worthless. A good likelihood ratio should be 10 or higher. Anything bigger than 50 represents an excellent diagnostic test. For a negative likelihood ratio (LR-), the corresponding boundaries are 0.5 (1/2), 0.1 (1/10), and 0.02 (1/50).

41. Likelihood ratios

You combine the likelihood ratio with information about

1. the prevalence of the disease,
2. characteristics of your patient pool, and
3. information about this particular patient

to determine the post-test odds of disease.

42. Likelihood ratios

Example: An early test for developmental dysplasia of the hip. The test has 92% sensitivity and 86% specificity in boys (AJPH 1998; 88(2): 285-288). This paper does not compute likelihood ratios, so you have to do a few calculations yourself.

- LR+ = Sn / (1 - Sp) = 0.92 / 0.14 = 6.6.
- LR- = (1 - Sn) / Sp = 0.08 / 0.86 = 0.09.

43. Likelihood ratios

Example: Suppose one of our patients is a boy with no special risk factors. The diagnostic test is positive. What can we say about the chances that this boy will develop hip dysplasia? The prevalence of this condition is 1.5% in boys. This corresponds to an odds of one to 66. Multiply the odds by the likelihood ratio, you get 6.6 to 66 or roughly 1 to 10. The post test odds of having the disease is 1 to 10 which corresponds to a probability of 9%.

44. Likelihood ratios

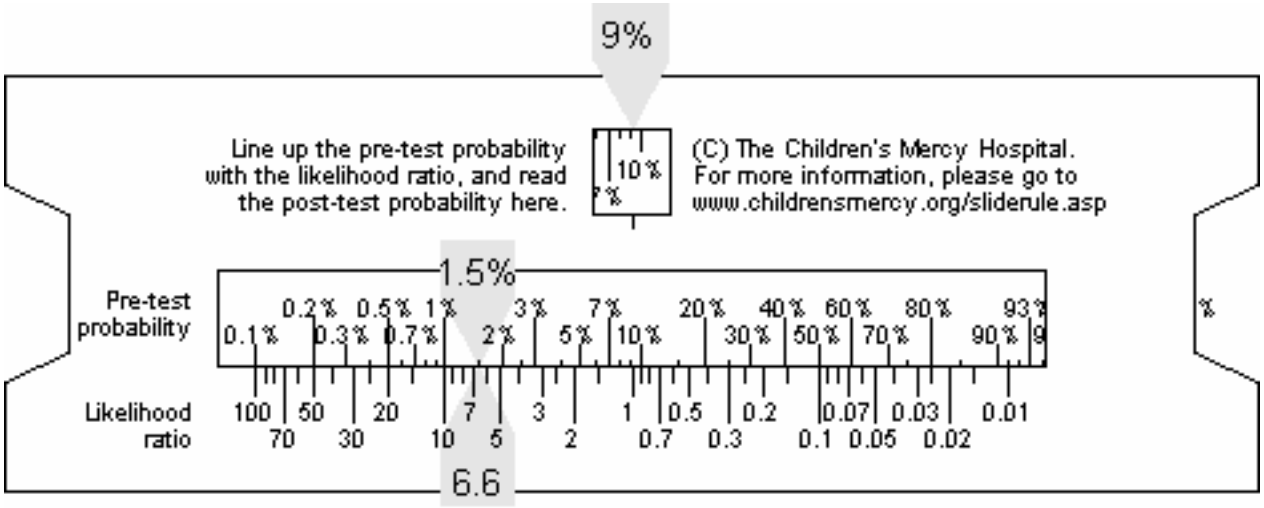
Example: Suppose we had a negative result, but it was with a boy who had a family history of hip dysplasia. Suppose the family history would change the pre-test probability to 25%. How likely is hip dysplasia, factoring in both the family history and the negative test result? A probability of 25% corresponds to an odds of 1 to 3. The likelihood ratio for a negative result is 0.09 or 1/11. So the post-test odds would be roughly 1 to 33, which corresponds to a probability of 3%.

45. Likelihood ratios

The use of likelihood ratios requires a bit of tedious calculations. I have developed a simple slide rule that will do likelihood ratio calculations for you.

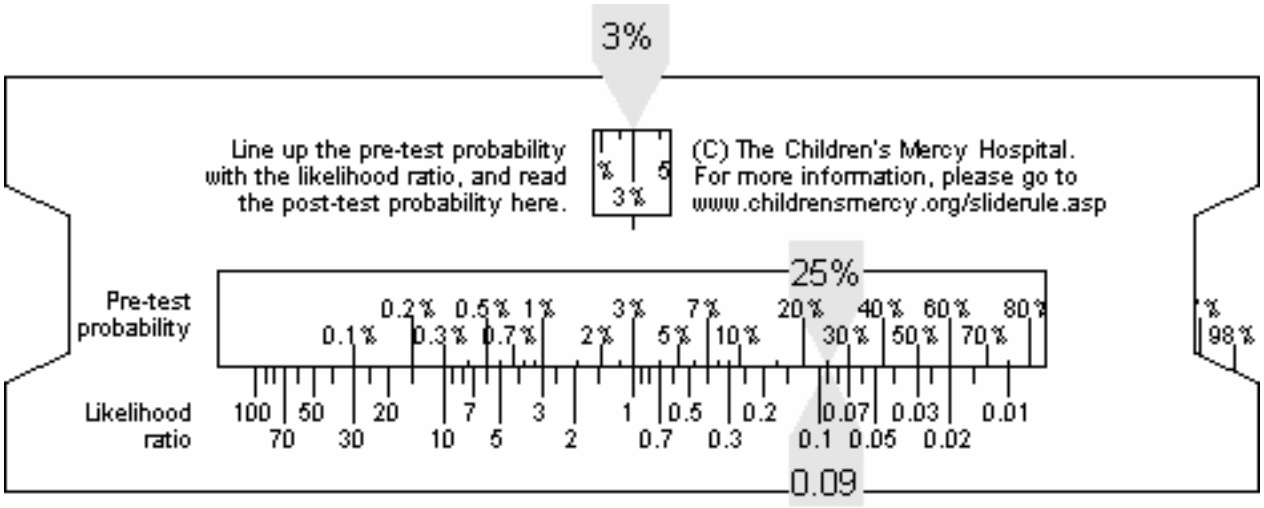
46. Likelihood ratios

For a pre-test probability of 1.5% and a positive likelihood ratio of 6.6, the post test probability is 9%.



47. Likelihood ratios

For a pre-test probability of 25% and a negative likelihood ratio of 0.09, the post test probability is 3%.



48. Repeat of pop quiz #1

The denominator for sensitivity is

1. All patients who are disease-free
2. All patients who have the disease
3. All patients who test negative
4. All patients who test positive

49. Repeat of pop quiz #2

If a disease has very low prevalence, you typically get poor values for

1. Negative predictive value
2. Positive predictive value
3. Sensitivity
4. Specificity

50. Repeat of pop quiz #3

A diagnostic test is useful for “ruling out” a disease when:

1. Sensitivity is large
2. Sensitivity is small
3. Specificity is large
4. Specificity is small

51. The next free webinars

- “Jumpstart statistics: Simple descriptive analyses” UMKC only!
  - Friday, March 5, noon-1pm, CST.
- “Putting it all together: meta-analyses and systematic overviews” Open to all!
  - Wednesday, March 31, 11am-noon, CST.