Selection of the Optimal 2x2 Table

Henry Glick

Epi 550

January 24, 2020

Continuously-Scaled Tests

- Last 2 sessions talked about dichotomous tests or tests treated dichotomously
- Consider use of WBC for diagnosis of bacteremia among children aged 3 to 36 months presenting with:

  A rectal temperature ≥39° C; no obvious focal infection; no "toxic" clinical appearance necessitating immediate hospitalization; and no specific viral infection, immune-deficiency condition, or chronic illness that alters standard approach to febrile illness
- Is WBC a dichotomous test?
- If not, what cut-off should we use to define a positive test?

WBC for Bacteremia

≥25	oooooo	6	■ oooooo	26
≥20, <25	oooo	4	■■ ooo	43
≥15, <20	ooooooo	7	■■■■■■ ooooooooo	129
≥10, <15	ooooooo	7	■■■■■■■ ■■■■■■■ ooooooooooo	292
≥0, <10	oo	2	■■■■■■■■■ ■■■■■■■■■ ooooooooo	369
	Disease	26	No Disease	859

■ = 20 observations; o = 1 observation

Bacteremia Example

Risk factors / symptoms	Pre-test	10*	15*	20*
Older age	0.01			
Average presentation	0.02			
Tachycardia	0.10			
Lethargy	0.20			
Indwelling catheter	0.50			

* Per high-powered field (HPF)

Summarizing Continuously Scaled Data

- At least 3 approaches available for summarizing continuously scaled data
- Identify "optimal" 2x2 table
  - Which of several candidate 2x2 tables -- each defined by a different cut-off for a positive test (e.g., $\geq$10, $\geq$15, or $\geq$20) -- should be used for which patients?
    - Address this question in this session
- Develop series of stratum-specific likelihood ratios (SSLR)
  - Which strata raise pre-test probability enough to treat?
  - Which lower pre-test probability enough to not treat?
- Use logistic regression to develop risk prediction rules

Multiple 2x2 Tables

- If using 1 of several potential 2 x 2 tables for a patient with a specific set of signs and symptoms, must determine:

Which of potential tables, each defined by a different criterion for a positive test, should we use?

What are some published methods for answering this question?

Which method is best?

Strategy

1) Start with data (Nx2 table) and evaluate sensitivity and specificity for each possible 2 x 2 table
2) Summarize test characteristics by developing a receiver operating characteristic (ROC) curve
3) Identify "optimal" 2 x 2 table
4) Describe a graphical summary of best table for different "types" of patients

No Recognized Convention for Nx2 Tables, But We Recommend...

Cut-off	D+	D-
Test results most indicative of disease (e.g., ≥25)		
•		
•		
•		
Test results least indicative of disease (e.g., <10)		

Test Results, WBC for Bacteremia

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia
≥25	6	26
≥20, <25	4	43
≥15, <20	7	129
≥10, <15	7	292
≥0, <10 *	2	369
Total	26	859

* Cells ≥0 and ≥5 combined; none of 26 children with disease had a WBC < 5

How many 2x2 tables can we construct?

Cut-Off ≥25

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	
≥25	6	26	+
≥20, <25	4	43	
≥15, <20	7	129	
≥10, <15	7	292	-
≥0, <10	2	369	

Red, classified as having a positive test (e.g., ≥25)
Blue, classified as having a negative test (e.g., <25)

Step 1. Estimate Sensitivities and Specificities

	D+	D-
≥25	6	26
<25	20	833
	0.2308	0.9697
		0.0303 (1-spec)

	D+	D-

	D+	D-

	D+	D-

	D+	D-

Cut-Off ≥20

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	
≥25	6	26	+
≥20, <25	4	43	
≥15, <20	7	129	
≥10, <15	7	292	-
≥0, <10	2	369	

Red, classified as having a positive test (e.g., ≥20)
Blue, classified as having a negative test (e.g., <20)

Estimate Sensitivities and Specificities

	D+	D-

	D+	D-
≥25	6	26
<25	20	833
	0.2308	0.9697
		0.0303

	D+	D-
≥20	10	69
<20	16	790
	.3846	.9197
		.0803

	D+	D-

	D+	D-

	D+	D-

4 Informative Tables (N-1 Tables for N Strata)

	D+	D-

	D+	D-
2) ≥25	6	26
<25	20	833
	0.2308	0.9697
		0.0303

	D+	D-
3) ≥20	10	69
<20	16	790
	.3846	.9197
		.0803

	D+	D-
4) ≥15	17	198
<15	9	661
	.6538	.7695
		.2305

	D+	D-
5) ≥10	24	490
<10	2	369
	.9231	.4296
		.5704

	D+	D-

"Informative": pre-test prob ≠ post-test prob

Cut-Off = ∞

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia
≥25	6	26
≥20, <25	4	43
≥15, <20	7	129
≥10, <15	7	292
≥0, <10	2	369

Red, classified as having a positive test (e.g., none)
Blue, classified as having a negative test (e.g., ≥0)

Cut-Off ≥ 0

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	
≥25	6	26	
≥20, <25	4	43	
≥15, <20	7	129	+
≥10, <15	7	292	
≥0, <10	2	369	

Red, classified as having a positive test (e.g., ≥0)
Blue, classified as having a negative test (e.g., none)

5 Strata, 4 Informative and 2 Uninformative Tables

	D+	D-
1) ≥∞	0	0
<∞	26	859
	Se, 0	Sp, 1 1-Sp,0

	D+	D-
2) ≥25	6	26
<25	20	833
	0.2308	0.9697 0.0303

	D+	D-
3) ≥20	10	69
<20	16	790
	.3846	.9197 .0803

	D+	D-
4) ≥15	17	198
<15	9	661
	.6538	.7695 .2305

	D+	D-
5) ≥10	24	490
<10	2	369
	.9231	.4296 .5704

	D+	D-
6) ≥0	26	859
<0	0	0
	SE, 1	SP, 0 1-Sp, 1

2x2 Table We ***NEVER*** Construct: Cut-Off ≥15, <20 *

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	
≥25	6	26	-
≥20, <25	4	43	-
≥15, <20	7	129	+
≥10, <15	7	292	-
≥0, <10	2	369	-

* Represents a stratum-specific likelihood ratio, not a 2x2 table

Sensitivities and Specificities of 6 Tables for 5 Strata

Cut-off	Sens	1-Spec	Spec
≥∞	0.000	0.000	1.000
≥25	0.231	0.030	0.970
≥20	0.385	0.080	0.920
≥15	0.654	0.231	0.769
≥10	0.923	0.570	0.430
≥0	1.000	1.000	0.000

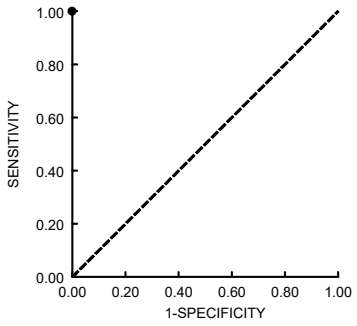
. tab score bact

	1=Bacteremia		
score	0	1	Total
1	369	2	371
2	292	7	299
3	129	7	136
4	43	4	47
5	26	6	32
Total	859	26	885

roctab bact score,detail table

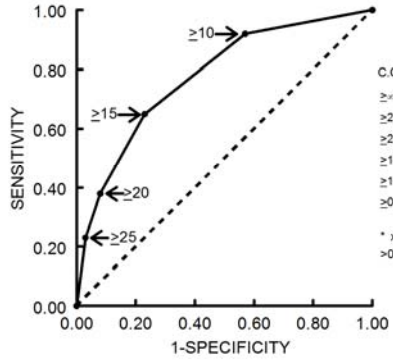
```
       |           WBC levels
1=Bact |    1     2     3     4     5 |  Total
-------+-------------------------------+----------
     0 |  369   292   129    43    26 |    859
     1 |    2     7     7     4     6 |     26
-------+-------------------------------+----------
 Total |   371   299  136    47    32 |    885

Detailed report of Sensitivity and Specificity
------------------------------------------------------------
                              Correct
Cutpoint   Sensit   Specif  Classif      LR+      LR-
------------------------------------------------------------
( >= 1 )  100.00%    0.00%    2.94%   1.0000
( >= 2 )   92.31%   42.96%   44.41%   1.6182   0.1791
( >= 3 )   65.38%   76.95%   76.61%   2.8366   0.4498
( >= 4 )   38.46%   91.97%   90.40%   4.7882   0.6691
( >= 5 )   23.08%   96.97%   94.80%   7.6243   0.7932
( >  5 )    0.00%  100.00%   97.06%            1.0000
------------------------------------------------------------
```

Step 2. Summarize Multiple 2x2 Tables



- Horizontal axis
- Vertical axis
- Best operating point
- 45° line
- Plotting operating points
- Points below 45° line

WBC ROC Curve



C.O. *	SENS	1-SPEC
≥∞	0.000	0.000
≥25	0.231	0.030
≥20	0.385	0.080
≥15	0.654	0.231
≥10	0.923	0.570
≥0	1.000	1.000

* x1000/mm³; Cells >0 and >5 combined

ROC "History"

- Receiver operating characteristic (ROC) curves developed in signal detection theory for, among other things, determining optimal settings for a radar "receiver"
- If radar set with too low a sensitivity, approaching planes or missiles overhead before defenses deployed
- If set with too low a specificity, defenses deployed only to find flocks of birds
- ROC curve plots trade-offs in sensitivity and specificity, and helps determine optimal trade-off between them

Decision Problem

- Tell story not to recount history of signal detection theory, but to identify nature of decision problem:
  - Radar monitors could make one grab at information and then had to make a decision whether or not to deploy planes and missiles
  - What was important?
    - A precise estimate of probability of enemy planes?

      OR
    - Estimate of whether expected value of deployment greater than expected value of not deploying?
      - Restated: Was expected value of deployment greater or less than expected value of withholding deployment?

”One (Test) and Done” Decision Making

- Refer to this latter form of decision making as "One (test) and Done"
- Characterized by need to make treatment decision quickly without a large number of opportunities to collect additional data
- Primary concern is ***NOT*** calculation of exact post-test probability of outcome
- Instead ***IS*** determination of whether post-test probability is high enough to initiate treatment

Step 3: Identify “Optimal” 2 X 2 Table

- If able to adopt any one of 6 2×2 tables from 5 strata, how should optimal table for a particular patient or class of patients be identified?
- Criterion for selection: Maximize expected value of treatment decision (minimize expected cost of mistakes)
  - Value defined in short term: make treatment decision based on current test result (i.e., "One (test) and Done")
- Expected value is a function of:
  - Frequency of correctly diagnosing presence and absence of disease (test accuracy)
  - Incremental value of correct diagnoses (cost of incorrect diagnoses) when they occur

Frequency of Correct Diagnosis (Test Accuracy)

- Proportion of correct classifications made using a particular set of test results

  $Accuracy_j$ = (Pre-test × $Sens_j$) + ([1-Pre-test] × $Spec_j$)
- Example #1: If pre-test probability = 0.20 and cut-off ≥ 10 (sensitivity = 0.923 and specificity = 0.430), accuracy equals:

  (0.20 * 0.923) + (0.80 * 0.430) = 0.529
- Example #2: If pre-test probability = 0.30 and cut-off ≥ 15 (sensitivity = 0.654 and specificity = 0.769), accuracy equals:

  (0.30 * 0.654) + (0.70 * 0.769) = 0.735

Accuracy of WBC for Bacteremia

			Accuracy		
			Pre-test Probability		
Cut-off	Sens	Spec	0.2	0.3	0.4
≥∞	0.000	1.000	0.800	0.700	0.600
≥25	0.231	0.970	***0.822***	0.748	0.674
≥20	0.385	0.920	0.813	***0.760***	0.706
≥15	0.654	0.769	0.746	0.735	***0.723***
≥10	0.923	0.430	0.529	0.578	0.627
≥0	1.000	0.000	0.200	0.300	0.400

- Which 2x2 table has greatest accuracy?
- Why?

Is Maximizing Accuracy Sufficient?

- Should we always use 2x2 table with greatest accuracy?
- In what situation would we want to use a less accurate test?

Value of Correct and Incorrect Diagnoses

- May not always want to maximize accuracy if value of true positive and true negative diagnoses differ
- For choice of an optimal 2x2 table, ***DON'T*** require knowledge of absolute value of difference in outcome for those with (ΔO_{D+}) and without (ΔO_{D-}) disease
  - ***INSTEAD*** require information about relative value
    - e.g., that difference in value of correctly treating someone (ΔO_{D+}) is twice difference in value of correctly withholding treatment (ΔO_{D-})
      - $\Delta O_{D+} = 2\ \Delta O_{D-}$ or $\frac{1}{2}\ \Delta O_{D+} = \Delta O_{D-}$
      - ***NOT***: $2\ \Delta O_{D+} = \Delta O_{D-}$ (i.e., ΔO_{D-} is twice ΔO_{D+})

$\Delta O_{D-} = 1$; $\Delta O_{D+} = 2$

- ΔO_{D+} is twice ΔO_{D-}
- $2\ \Delta O_{D-} = \Delta O_{D+}$
- Ratio of difference in value for disease and difference in value for no disease is 2 (2/1)
- ΔO_{D-} is half (50%) of ΔO_{D+}
- $\Delta O_{D-} = 0.5\ \Delta O_{D+}$
- ΔO_{D-} is 50% less than ΔO_{D+}
- Ratio of difference in value for disease and difference in value for no disease 0.5 (1/2)
- Easy way to not have to think about it: simply substitute 1 for the other
  - NBPT = se * p * $2\ \Delta O_{D-}$ + (1-sp) * (1-p) * ΔO_{D-}
  - In this equation ΔO_{D-} is less than ΔO_{D+}

Expressions of Value

- 4 expressions of value:
  - Expected cost of mistakes (minimize)
    - p (1-sens) ΔO_{D+} + (1-p) (1-spec) ΔO_{D-}
    - Used by Metz (1978) and others, (until last few years, always used this expression with this material)
  - Expected net benefit of positive test (maximize)
    - p sens ΔO_{D+} - (1-p) (1-spec) ΔO_{D-}
    - Used by Pauker and Kassirer (1980) and others
  - Expected net benefit of negative test (maximize)
    - (1-p) spec ΔO_{D-} - p (1-sens) ΔO_{D+}
  - Expected benefit of correct diagnoses (maximize)
    - p sens ΔO_{D+} + (1-p) spec ΔO_{D-}

Expressions of Value (2)

- When choosing among 2x2 tables, numeric results for each of 4 expressions differ
- But for all 4 expressions, **DIFFERENCE** in value between any pair of 2x2 tables is always identical

Expressions of Value (3)

- For example, if pre-test probability = 0.2 and ΔO_{D+} is twice ΔO_{D-}, then compared to the table defined by a cut-off of $\geq$15, the table defined by a cut-off of $\geq$20:
  - Minimizes expected cost of mistakes
    - 0.31 vs 0.3232: 0.0132
  - Maximizes expected net benefit of positive test
    - 0.09 vs 0.0768: 0.0132
  - Maximizes expected net benefit of a negative test
    - 0.49 vs 0.4768: 0.0132
  - Maximizes expected benefit of correct diagnoses
    - 0.89 vs 0.8768: 0.0132

  i.e., For all 4 expressions, there is a 0.0132 difference in value between $\geq$20 and $\geq$15 tables

Expected Net Benefit of Positive Test (NBPT)

- NBPT made up of 2 components:
  - Expected benefit among those with disease:

    Pre-test probability × ΔO_{D+} × Sensitivity
    - i.e., proportion of patients with disease (i.e., in whom true positives occur) × incremental value of positive test given disease × proportion of true positives among those with disease
  - Expected loss among those without disease:

    (1 - Pre-test probability) × ΔO_{D-} × (1 - Specificity)
    - i.e., proportion of patients without disease (i.e., in whom false positives occur) × incremental loss from positive test given no disease × proportion of false positives among those without disease

NBPT (2)

- NBPT equals difference between expected benefit from a positive test among those with disease and expected loss from a positive test among those without disease

  $NBPT_j = p\ \Delta O_{D+}\ Sensitivity_j - (1-p)\ \Delta O_{D-}\ (1-Specificity_j)$

- One means of identifying optimal 2x2 table uses this formula to calculate NBPT for each candidate 2x2 table

Net Benefit-Maximizing 2x2 Table

- If p = 0.2, ΔO_{D+} = 2, and ΔO_{D-} = 1, WBC 2x2 table that maximizes net benefit of a positive test is:

Cut-off	True Positive		False Positive		NBPT
≥∞	0.2 * 2 * 0.000	-	0.8 * 1 * 0.000	=	0
≥25	0.2 * 2 * 0.231	-	0.8 * 1 * 0.030	=	0.0684
≥20	***0.2 * 2 * 0.385***	***-***	***0.8 * 1 * 0.080***	***=***	***0.09***
≥15	0.2 * 2 * 0.654	-	0.8 * 1 * 0.231	=	0.0768
≥10	0.2 * 2 * 0.923	-	0.8 * 1 * 0.570	=	-0.0868
≥0	0.2 * 2 * 1.000	-	0.8 * 1 * 1.000	=	-0.4

- Note that for a pretest probability of 2, a cut-off of ≥25 has greatest accuracy, but not maximum NBPT

Optimal Trade-off Between Sensitivity And Specificity

- Calculating NBPT for ***ALL*** candidate tables works, but not simplest method for selecting optimal 2x2 table
  - Requires calculation of net benefit for **EVERY TABLE**, ***including noninformative tables***
  - Requires **RECALCULATION** of net benefit for each pre-test probability and each ΔO_{D+} and ΔO_{D-} that we consider
- More efficient method for identifying optimal table uses slopes of ROC curve/stratum specific likelihood ratios (SSLR)
  - SSLR represent characteristics of test that we discuss in next lecture

More Commonly Recommended Strategy

- Define trade-off between sensitivity and specificity that maintains a constant expected net benefit of a positive test (referred to as "optimal" operating slope)
- Identify tangency between this trade-off and test's ROC curve
  - Equivalent to comparing OOS and SSLR

Step 1. Define Trade-off Between Sensitivity and Specificity that Maintains a Constant Net Benefit of a Positive Test

- Start with equation for expected net benefit of positive test formula:

$$NBPT_j = p\,\Delta O_{D+}\,se_j - (1-p)\,\Delta O_{D-}\,(1-sp_j)$$

- And solve for sensitivity (sej):

$$se_j = \frac{(1-p)\,\Delta O_{D-}}{p\,\Delta O_{D+}}\,(1-sp_j) + b_j$$

$$\text{where } b_j = \frac{NBPT_j}{p\,\Delta O_{D+}}$$

Optimal Operating Slope

- Defines a line (y = mx + b) with a fixed NBPT when sensitivity and 1-specificity traded off by use of "optimal operating slope":

$$m = OOS = (1-p)\,\Delta O_{D-} / \; p\,\Delta O_{D+}$$

  - e.g., if p = .2, $\Delta O_{D+} = 2$ and $\Delta O_{D-} = 1$, then:

$$OOS = \frac{0.8 * 1}{0.2 * 2} = \frac{0.8}{0.4} = 2.0$$

- OOS maintains constant NBPT by trading off sensitivity and specificity in proportion to:
  1. Size of population among whom false positive and false negative mistakes can be made ([1-p]/p)
  2. Relative difference in outcomes (ΔO_{D-} / ΔO_{D+})

Interpretation of Optimal Operating Slope

- OOS of 2 means 2 additional units of sensitivity are needed for every 1 additional unit of 1-specificity to maintain constant net benefit of a positive test
  - If only wanted to maintain a constant accuracy, would need 4 additional units of sensitivity for every 1 unit of 1-specificity
    - Because in this example false negative mistakes made in 20% of population while false positive mistakes made in 80% of population
  - But when also considering differences in value, 4 additional units are reduced to 2
    - Because in example, difference in outcome among those with disease is twice difference in outcome among those without disease

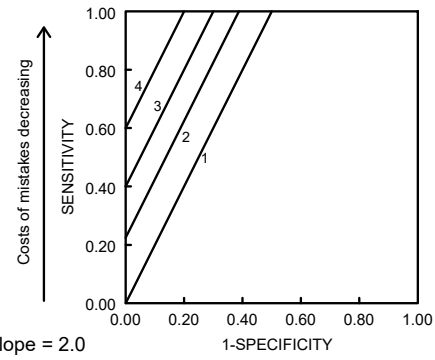
Step 2. Develop Family of Lines with OOS

- Impose series of lines, all with OOS, on ROC graph
  - In following example, lines have slope of 2
- Slope of 2 appropriate when pre-test probability of disease is 20% and ΔO_{D+} is twice ΔO_{D-} (i.e., $\frac{\Delta O_{D+}}{\Delta O_{D-}} = 2$)
- Also appropriate when pre-test probability of disease is 50% and ΔO_{D+} is half of ΔO_{D-} ($\Delta O_{D+} = 0.5\ \Delta O_{D-}$)

$$OOS = \frac{0.5 * 1}{0.5 * 0.5} = \frac{0.5}{0.25} = 2.0$$

- Etc….

Step 2. Develop a Family of Lines with OOS (2)



- Slope = 2.0

NBPT At Intercept

- NBPT =p ΔO_{D+} Sensitivity$_j$ - (1-p) ΔO_{D-} (1-Specificity$_j$)
- Because at intercept, 1-Specificity = 0, second term drops out and NBPT reduces to:
  - NBPT = p ΔO_{D+} sensitivity OR
  - NBPT = p ΔO_{D+} Intercept
- Thus, (relative) expected NBPT of line 2 in previous figure, which has an intercept of 0.225, equals:

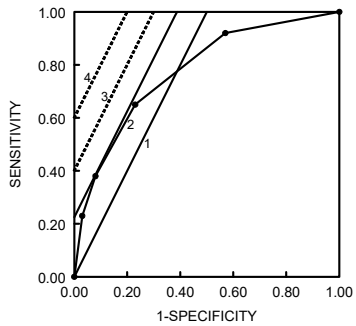
  $NBPT_{0.225}$ = .2 x 2 x .225 = 0.09
  - 0.09 equals "relative" expected NBPT because we didn't use relative rather than absolute magnitudes of ΔO_{D+} and ΔO_{D-} in calculation
    - i.e., 2 and 1 rather than, for example, $10,000 and $5,000

NBPT Maximization

- Among the family of lines defined by OOS, lines with larger intercepts have higher NBPT; those with smaller intercepts have lower NBPT

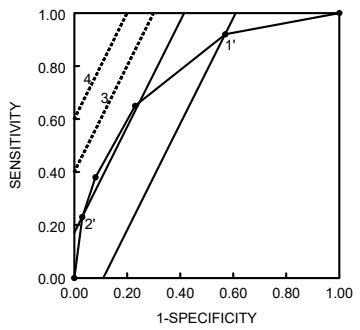
FOR A GIVEN OOS, WANT TO CHOOSE THE TEST CUT-OFF WITH LARGEST NBPT

Step 3. Identify Tangency



- Tangency at ≥20 operating point
- Intercept = 0.225
- NBPT = 0.09 (.2*2*.225)

Step 3. Other Lines Lower NBPT or Infeasible



- Line 2' intersects ≥25 operating point; line 1' intersects ≥10
- Intercepts = 0.171 (2') and -0.218 (1')
- Relative NBPT = 0.0684 (.2*2*.171) and -0.0872 (.2*2*-.218) respectively

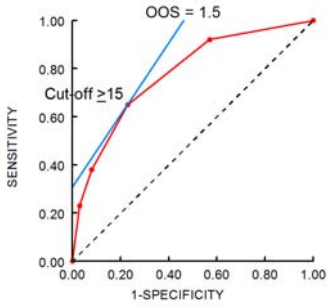
Intuition Behind Why We Want a Tangency

- As already noted, goal is to identify cut-off that maximizes NBPT arising from treatment decisions based on test result
- Among family of lines defined by OOS, one with highest obtainable intercept has largest NBPT
  - Because intercept (sensitivity) is maximized
- Largest obtainable intercept is defined by tangency

Don't expect anyone to 1) draw ROC curve, 2) draw family of lines with OOS, and 3) identify tangency

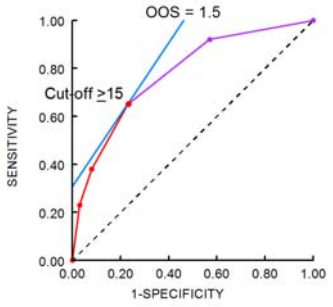
Practical Method for Identifying Cut-Off

- Is there something algebraic about ROC curve that is associated with positive and negative test results?



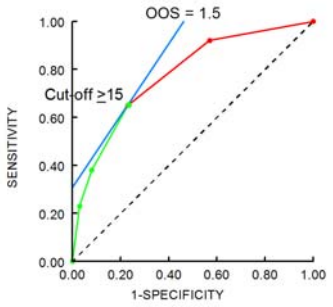
Slopes of ROC Curve < OOS

- Cut-offs with ROC curve slopes less than OOS (violet line segments) represent negative test results



Slopes of ROC Curve > OOS

- Cut-offs with ROC curve slopes greater than OOS (green line segments) represent positive test results



Can identify optimal cut-off (tangency) by comparing OOS and slopes of ROC curve

Cut-offs with slopes < OOS represent negative test results

Cut-offs with slopes > OOS represent positive test results

Cut-offs with slopes = OOS positive or negative

ROC Curve Slopes

- Slopes of ROC curve represent stratum-specific likelihood ratios (SSLR, More on this 1/29)
- Can calculate slopes/SSLR using multiple methods (e.g., $\Delta y / \Delta x = \Delta sens / \Delta(1\text{-}spec)$)
- If have original nx2 table, simplest method may be:

$$(a * f) / (b * e)$$

Cut-Off ≥25

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	Slope/ SSLR
≥25	(a) 6	(b) 26	7.6243
≥20, <25	4	43	
≥15, <20	7	129	
≥10, <15	7	292	
≥0, <10	2	369	
	(e) 26	(f) 859	

(6 * 859) / (26 * 26) = 7.6242604

Stata Command for CI for $LR_{\geq 25}$

csi 6 26 20 833

	Exposed	Unexp	Total	
Cases	6	26	32	
Noncases	20	833	853	
Total	26	859	885	
Risk	.2307692	.0302678	.0361582	
	Point Estimate		95% Conf.	Interval
Risk diff	.2005015		.0381476	.3628554
Risk ratio	**7.62426**		**3.434883**	**16.2766**
Attr frac ex	.8688397		.7088692	.9409097
Attr frac pop	.1629075			

chi2(1) = 29.11 Pr>chi2 = 0.0000

Exposed = D+; Cases = T+ Unexposed = D-; Noncases = T-

Cut-Off ≥20 to <25

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	Slope/ SSLR
≥25	6	26	7.6243
≥20, <25	(a) 4	(b) 43	3.0733
≥15, <20	7	129	
≥10, <15	7	292	
≥0, <10	2	369	
	(e) 26	(f) 859	

(4 * 859) / (43 * 26) = 3.0733453

Stata Command for CI for $LR_{\geq 20:<25}$

csi 4 43 22 816

	Exposed	Unexp	Total	
Cases	4	43	47	
Noncases	22	816	838	
Total	26	859	885	
Risk	.1538462	.0500582	.0531073	
	Point Estimate		95% Conf.	Interval
Risk diff	.1037879		-.0356616	.2432375
Risk ratio	**3.073345**		**1.191737**	**7.925783**
Attr frac ex	.6746217		.160889	.8738295
Attr frac pop	.0574146			

chi2(1) = 5.41 Pr>chi2 = 0.0201

Exposed = D+; Cases = T+ Unexposed = D-; Noncases = T-

Cut-Off ≥15 to <20

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	Slope/ SSLR
≥25	6	26	7.6243
≥20, <25	4	43	3.0733
≥15, <20	(a) 7	(b) 129	1.7928
≥10, <15	7	292	
≥0, <10	2	369	
	(e) 26	(f) 859	

(7 * 859) / (129 * 26) = 1.7927847

Stata Command for CI for $LR_{\geq 15:<20}$

csi 6 26 20 833

	Exposed	Unexp	Total	
Cases	7	129	136	
Noncases	19	730	749	
Total	26	859	885	
Risk	.2692308	.1501746	.1536723	
	Point Estimate		95% Conf. Interval	
Risk diff	.1190561		-.0531053	.2912176
Risk ratio	**1.792785**		**0.9331628**	**3.444283**
Attr frac ex	.4422085		-.0716244	.7096638
Attr frac pop	.0227607			

chi2(1) = 2.75 Pr>chi2 = 0.0927

Exposed = D+; Cases = T+ Unexposed = D-; Noncases = T-

Cut-Off ≥15

Cut-off	Children w/ Bacteremia	Children w/o Bacteremia	Slope/ SSLR
≥25	6	26	7.6243
≥20, <25	4	43	3.0733
≥15, <20	7	129	1.7928
≥10, <15	7	292	0.7920
≥0, <10	2	369	0.1791
	26	859	

ΔSensitivity / Δ1-Specificity

- Alternatively:

Cut-off	$Se_{j=i} - Se_{j=i-1}$		$1\text{-}Sp_{j=i} - 1\text{-}Sp_{j=i-1}$	Slope
≥25 - ∞	0.23077 - 0.00000	/	0.03027 - 0.00000	7.624
≥20 - ≥25	0.38462 - 0.23077	/	0.08033 - 0.03027	3.073
≥15 - ≥20	0.65385 - 0.38462	/	0.23050 - 0.08033	1.793
≥10 - ≥15	0.92308 - 0.65385	/	0.57043 - 0.23050	0.792
≥0 - ≥10	1.00000 - 0.92308	/	1.00000 - 0.57043	0.179

Compare Magnitude of OOS and Slopes

- OOS might be:
  - Equal to slope of line between two contiguous operating points
  - Greater than one slope, but less than contiguous slope
  - Greater than all slopes of ROC curve
  - Smaller than all slopes of ROC curve

OOS Equals Slope of Line Between 2+ Cut-offs

- Operating points on line segment have same (maximum) NBPT and can use any of 2+ cutoff on line segment to define a positive test
  - e.g., if slope equals 3.073, use either ≥20 or ≥25 to define a positive test
  - If slope equals 1.793, use either ≥15 or ≥20 to define a positive test
- Even though NBPT is equal for 2 operating points, their combinations of gains from true positive and losses from false positive differ

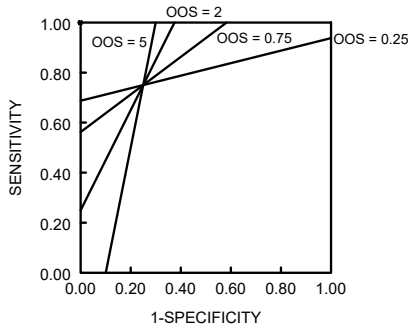
OOS Greater than One Slope and Less than Another

- Test cut-offs with slopes greater than OOS represent positive test results, while cut-offs with slopes less than OOS represent negative test results
  - e.g., an OOS of 2 is greater than 1.792 (≥15) but less than 3.073 (≥20). Test results ≥20 represent positive tests; test results <20 represent negative tests
    - At this operating point, optimal sensitivity and specificity are 0.385 and 0.920
  - An OOS of 1 is greater than 0.792 (≥10) but less than 1.792 (≥15). Test results ≥15 represent positive tests and test results <15 represent negative tests
    - At this operating point, optimal sensitivity and specificity are 0.654 and 0.769

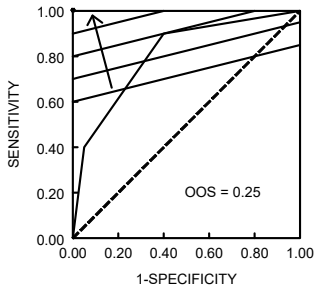
OOS Greater than or Less than All Slopes

- If OOS is greater than all ROC curve slopes (e.g., 10), tangency occurs at origin (sensitivity = 0; 1-specificity = 0; specificity = 1) and all test results represent negative tests
- If OOS is less than all ROC curve slopes, tangency occurs at upper right hand corner of curve (sensitivity = 1; 1-specificity = 1; specificity = 0) and all test results represent positive tests
- Because these operating points both fall on 45° line of no information, use of test at these cut-offs provides no additional information for making treatment decision
  - If main goal of testing is to gain clinical certainty, better to assume patient does not have / has disease than to test

Optimal Operating Slopes



Shallower OOS Emphasize Sensitivity

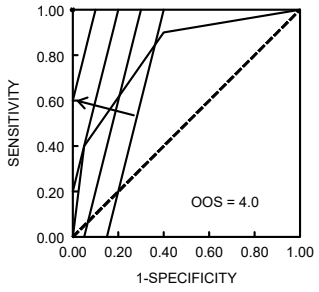


- When OOS is flatter (more horizontal), tangency occurs by moving lines toward top of X,Y graph
- Emphasizes greater sensitivity and lesser specificity

Intuition

- All else equal, as pre-test probability increases OR as ΔO_{D+} increases compared to ΔO_{D-}, accept weaker evidence of disease to decide to treat patient
  - Thus use less stringent cut-off for disease which raises sensitivity of test
- Rationale for raising sensitivity when pre-test probability or difference in outcomes among those with disease is relatively high is that value of false negative mistakes avoided due to gains in sensitivity offsets value of newly induced false positive mistakes

Steeper OOS Emphasize Specificity



- When OOS is steeper (more vertical), tangency occurs by moving lines toward left of X,Y graph
- Emphasizes greater specificity and lesser sensitivity

Intuition

- All else equal, as pre-test probability decreases OR as ΔO_{D-} increases compared to ΔO_{D+}, require stronger evidence of disease before we decide to treat patient
  - Thus use a more stringent cut-off for disease which raises specificity of test
- Rationale for raising specificity when pre-test probability is relatively low or difference in outcomes among those without disease is relatively high is that value of false positive mistakes avoided due to gains in specificity offsets value of newly induced false negative mistakes

Summary of OOS and Test Characteristics

	ΔO_{D+} High	ΔO_{D-} High
High pre-test prob	Sensitivity + Sensitivity +	Sensitivity - Sensitivity +
Low pre-test prob	Sensitivity + Sensitivity -	Sensitivity - Sensitivity -

Point "Closest" to Northwest Corner

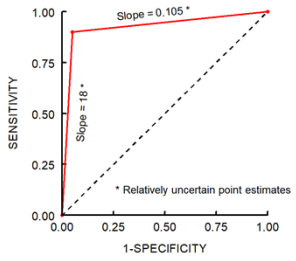
- Is a sense in which it is correct that cut-off "closest" to upper left corner is preferred operating point
- If so, should we always use cut-off of $\geq$15, which has smallest Euclidian distance (0.416) between cut-off and upper left corner?
  - NO, closeness isn't measured by Euclidian distance between operating point and corner
- When OOS is 2, intercept of line tangent to $\geq$20 is closer to upper left corner than is intercept of any other line that intersects ROC curve
  - Thus, "closeness" relates to intercepts of lines drawn with OOS, not to operating points on curve

ROC Curves and Dichotomous Tests

- In previous lectures, made no mention of ROC curves nor of identification of optimal 2x2 table
- But can define an ROC curve for truly dichotomous tests
- Given that a dichotomous test has 2 strata, it has 1 (2 - 1) informative 2x2 table and 2 noninformative tables

WBC ≥ 10 ROC Curve

- ROC curve for auditory biological marker if it were truly dichotomous with a cut-off of 0.596 on the regression score is shown below
- Its 2 slopes equal 18 (LR+) and 0.105 (LR-)



Interpreting Dichotomous Test ROC Curve

- Using what we've discussed today:
  - If optimal operating slope >18 tangency will occur at origin of curve
    - It would be better to assume patient does not have disease than it would be to order test
  - If optimal operating slope <0.105, tangency will occur at upper right hand corner of curve
    - It would be better to assume patient has disease than it would be to order test
  - If optimal operating slope is between 0.105 and 18, perform test and treat patient if result is positive while withholding treatment if result is negative

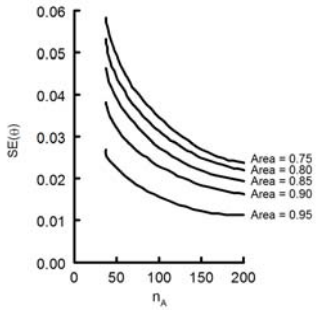
Making ROC Curve Analysis Clinically Relevant

- How many different types of tests do you order in your routine practice?
- How many different types make up 50% of total number of tests you order?
- Are you using different criteria for a positive test (i.e., different test results) for different pre-test probabilities for these tests?

Making ROC Curve Analysis Clinically Relevant (2)

- If not, consider:
  - Doing literature searches on top five tests you order to see whether or not ROC curves have been reported for these tests (if not you might want to do this evaluation)
  - Obtaining ROC curves for these tests and computing optimal operating slopes (and thus optimal operating points) for different pre-test probabilities
- Use this information when you order these common diagnostic tests

Samples Size for ROC Curve



Standard error (SE) for estimated area under ROC curve in relation to sample size (n_A = number of abnormal cases) and true area under ROC curve (θ). Calculations assume an equal number of normal cases.

Sample Size, Difference in Areas

- Sample size per test for comparison of ROC areas derived by use of two independent samples

$$n_d = \frac{\left(z_{1-\alpha/2}\sqrt{SE_1 + SE_2} + z_{1-b/2}\sqrt{SE_1 + SE_2 - \Delta^2}\right)}{\Delta^2}$$

where n_d equals number of subjects with/without condition; SE = standard error; and Δ equals difference in areas

  - If a case/control design is used, minimum total sample is $2n_d$ (i.e., n_d with disease and n_d without disease). If we are evaluating consecutive patients, total number of patients in sample needs to be larger of:

    N_{Sens} / Prevalence or N_{Spec} / 1-Prevalence

Samples Size for ROC Curve

- If two tests are being evaluated in a single sample, sample size is reduced because:
  - One, not two samples are required
  - Results of two tests possibly correlated
- Reduction due to correlation is calculated by subtracting 2 times covariance function from result under square root term (see pp. 209-210, Zhou)
  - e.g.,

$$z_{1-a/2}\sqrt{SE_1 + SE_2 - 2Cov_f}$$

where Cov_f = covariance function

Summary

- When a test is truly dichotomous, can:
  - Use it, with its sensitivity and 1-specificity or likelihood ratio positive and negative
  - Declare everyone has disease
  - Declare that no one has disease
- When a test is not truly dichotomous, and use a 2x2 approach for its interpretation, question that must be addressed is:

"Which of several candidate 2x2 tables should we use?"

Summary (2)

- In this presentation proposed a 2-step process for answering this question
- First, construct an ROC curve, which summarizes sensitivities and 1-specificities of candidate 2x2 tables
- Second, identify an optimal operating slope
  - Trades off sensitivity and 1-specificity in proportion to size of population among whom false positive and true positive mistakes can be made ([1-p]/p) and difference in outcome when they occur (ΔO_{D-} / ΔO_{D+})

Summary (3)

- Third, identify a tangency between ROC curve and a family of lines all defined by optimal operating slope
  - One means of identifying this tangency is to calculate slopes (SSLR) of ROC curve and compare them to OOS
- Resulting optimal 2x2 table maximizes net benefit of a positive test
- However, other approaches for interpreting continously scaled tests exist

APPENDIX: ROC AREA

ROC Area

- Area under ROC curve is a commonly reported statistic for diagnostic tests
- Many commentators believe it is an appropriate method of choosing between tests (i.e., choose test with larger area)
- ROC area is an indirect measure of test's "discriminating ability"

Discrimination

- Discrimination: Ability to assign different scores to those with and without disease
  - e.g., to assign generally lower scores to those without disease and to assign generally higher scores to those with disease
  - Discrimination is a property of scores
  - Given that predicted probabilities can be interpreted as scores, it applies to probabilities as well
    - For example, if a test score was 0.49 for everyone with disease and 0.51 for everyone without disease, predictions would be perfectly discriminating
      - That is, everyone with disease has a higher score than everyone without disease

Interpretation of ROC Area

- ROC areas can range between 0.5 (area under 45º line of no information) and 1.0 (area under ROC curve of a dichotomous test that has 100% sensitivity and specificity)
  - Area of 0.5 represents no ability to discriminate risk
    - Test assigns a similar distribution of scores to those in whom disease is present and those in whom disease is absent
  - Area of 1.0 represents perfect discrimination
    - No overlap in distribution of scores assigned to those in whom disease is present and those in whom disease is absent

Interpretation of ROC Area (2)

- Although curves with ROC areas of 0.5 and 1.0 are clearly distinguishable, there is little systematic information available about benefit of small increases in area under ROC curve (e.g., an increase from 0.75 to 0.77)
  - But, tests with larger areas under their ROC curve in general are more discriminating than are tests with smaller areas

Interpretation of ROC Area (3)

- Technically, ROC area equals probability that rule will correctly rank any randomly selected pair of persons, one in whom outcome of interest is present and one in whom it is absent
  - Nonparametric area represents p-value we derive from a Wilcoxon rank sum test
  - How often do pairs of patients walk into a provider's office; declare that one has disease while other does not, and then ask "which of us has a higher test score?"

Interpretation of ROC Area (4)

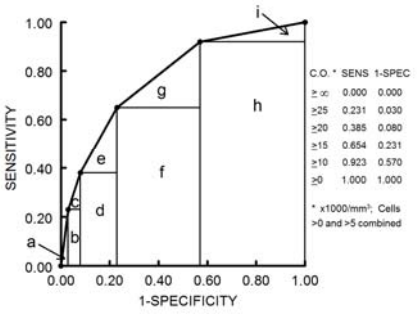
- ROC area is used as a measure of discrimination in many applications other than diagnostic test evaluation
  - C-statistic that is routinely reported by SAS as an index of discriminating ability of fitted logistic regressions models equals nonparametric area under logistic regression's ROC curve
  - Similarly, lroc command in STATA that can be run after logistic regression reports same area

Methods for Computing Area under ROC Curve

- Nonparametric or trapezoidal method
  - Calculates area of a series of triangles and squares
  - Formula or "connect dots"
- Maximum likelihood method
  - How easy is it to determine test cutoffs associated with unobserved continuous set of points on this curve?

Trapazoidal Method for Area

- Calculate area of a series of triangles (a, c, e, g, and i) and squares (b, d, f, h) and sum



Stata Area Calculation

. roctab bact score, hanley

ROC			-Asymptotic Normal-	
Obs	Area	Std. Err	[95% Conf.	Interval]
885	0.7779	.0457	0.68836	0.86752

Maximum Likelihood Area calculation

```
. rocfit bact score if test==0
Fitting binormal model:
Iteration 0:   log likelihood = -1130.9919
Iteration 1:   log likelihood = -1130.9858
Iteration 2:   log likelihood = -1130.9858
Binormal model of bact on score                 Number of obs   =        885
Goodness-of-fit chi2(2) =       0.10
Prob > chi2             =     0.9531
Log likelihood          =  -1130.9858
------------------------------------------------------------------------------
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   intercept |   1.195332   0.294345     4.06   0.000     0.618427    1.772238
   slope (*) |   1.044539   0.200562     0.22   0.824     0.651445    1.437633
-------------+----------------------------------------------------------------
       /cut1 |  -0.177857   0.042992    -4.14   0.000    -0.262119   -0.093595
       /cut2 |   0.738356   0.047040    15.70   0.000     0.646159    0.830552
       /cut3 |   1.403876   0.061660    22.77   0.000     1.283025    1.524728
       /cut4 |   1.872904   0.084042    22.29   0.000     1.708184    2.037624
------------------------------------------------------------------------------
------------------------------------------------------------------------------
             |                  Indices from binormal fit
       Index |   Estimate   Std. Err.                      [95% Conf. Interval]
-------------+----------------------------------------------------------------
    ROC area |   0.795773   0.044733                      0.708098    0.883449
    delta(m) |   1.144364   0.204372                      0.743802    1.544926
        d(e) |   1.169293   0.224126                      0.730014    1.608572
        d(a) |   1.169016   0.223157                      0.731635    1.606396
------------------------------------------------------------------------------
(*) z test for slope==1
```

rocplot,confband