Why Is It So Hard to Communicate the Value of a Diagnostic Test? (Part 1)

Henry Glick

02/14/20
(Happy Valentine's Day)

Outline (This and Next Lecture)

- (Brief) Identify commonly used approaches meant to communicate value of diagnostic tests and discuss limitations
- Suggest properties an ideal statistic should have
- Review better approaches
  - Test threshold graphs
  - Decision curve graphs
  - Net benefit "football" graphs
  - Decision Slope graphs
- Goal: identify a single table or graph that communicates value for a wide range of patients

Goal

- Identify a statistic/set of statistics that allow someone to obtain "best" outcome:
  - When there is no test, by either withholding treatment or treating
  - When there is a dichotomous test, by either withholding treatment, testing and making treatment decision based on test result, or treating
  - When there is a multi-outcome test, by either withholding treatment, testing and making treatment decision by use of "best" test cut-off, or treating
  - When there are 2 (dichotomous or multiple outcome) tests, by withholding treatment, testing and making a treatment decision by use of the "best" test and the best cut-off, or treating

WHAT DO WE MEAN BY "BEST" OUTCOME?

If value of correct positive and negative diagnoses are equal (or simply can't determine relative value), maximum accuracy. If relative value can be determined, maximum NBPT/minimum cost of mistakes

Common Approaches Meant to Communicate Diagnostic Test Value

- Comparison between tests
  - Largest area under receiver operating characteristic (ROC) curve
- Comparison between individual operating points from a single test or between multiple tests
  - Point on ROC curve closest to (with smallest Euclidian distance from) northwest corner of ROC graph
  - Largest Youden index (sens+spec-1)
  - Largest diagnostic odds ratio (sens*spec)/((1-sens)*(1-spec))
  - Point where sens ≈ spec

PROBLEMS WITH COMMON APPROACHES

Limited To Characteristics of Tests

- All 5 approaches limited to consideration of characteristics of test (often stable) BUT ignore characteristics of patients (often vary)
- Relevant patient characteristics include:
  - Pretest probability of disease
  - Value of test results among diseased/nondiseased
- Thus listed approaches:
  - Do not help with decision to test vs withhold testing
  - Either don't recommend test cut-off (area under ROC curve) or always recommend same test cut-off for all patients (other 4 approaches)

ANSWER TO QUESTION "WHY IS IT SO HARD TO COMMUNICATE THE VALUE OF A DIAGNOSTIC TEST?"

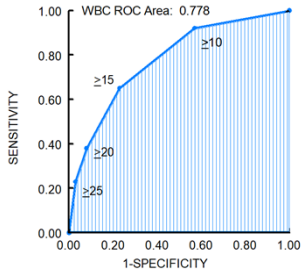
Because it depends on characteristics of both test and patients in whom test is being used!

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

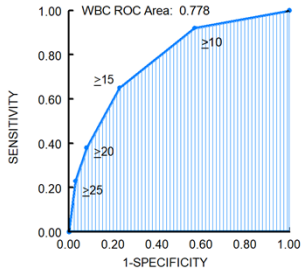
Method 1: Area Under ROC Curve

- Provides little or no information for choosing among tests or operating points



ROC Area: Selection of an Optimal Cut-Off

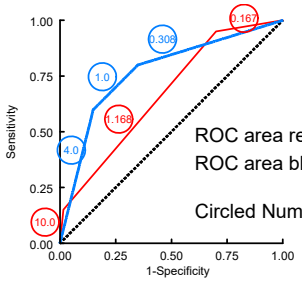
- Does knowing the test area help us decide which of the 6 potential cut-offs we should use?



ROC Area: Selection of an Optimal Test

- Test with smaller area can have cut-offs for some patients that are superior to all cut-offs from test with larger area

- ≤.167: Treat
- ≥.167; ≤0.429: Red
- ≥.429; ≤3.33: Blue
- ≥3.33; ≤10: Red
- ≥10: Do nothing



ROC area red: 0.67
ROC area blue: 0.77

Circled Numbers = SSLR

CONCLUSION, AREA UNDER ROC CURVE

- Provides no information for choice of cut-off for a single test
  - i.e., Provides no information that helps with decision to treat no one, use single informative test cut-off (dichotomous tests), use one of several informative cut-offs (continuous tests), or treat everyone
- Provides little or no information for choice among (optimal) tests/cut-offs among multiple tests

Method 2: Point Closest to Northwest Corner

- Sackett et al: "…the point on an ROC curve that is closest to this upper left-hand corner is the `best' cutoff in terms of making the fewest mistakes when prevalence is at or around 50%..."

Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology. Second Edition. Boston: Little Brown; 1991. (also first edition, p. 106)

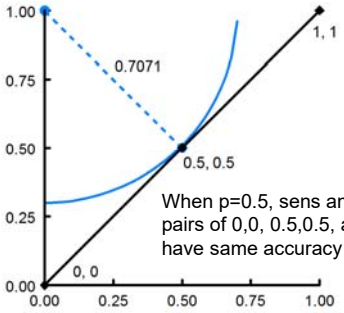
"Fewest Mistakes When Prevalence At Or Arouind 50%

- Simply not true
- Are we always making decisions for patients whose pretest probability is "at or around 50%"?
- Should fewest mistakes be our goal or should it be lowest cost of mistakes (greatest benefits)?

Counter Example #1

- Consider 3 operating points on ROC graph: 0,0; 0.5,0.5; and 1,1
  - All 3 lie on 45º line and have no information (post-test probability = pre-test probability)
- 0.5,0.5 is closer to upper left hand corner (0.707 distance) than 0,0 and 1,1 (1.0 distance each)
- When prevalence is around 50% (or any other percentage for that matter) Is 0.5,0.5 more accurate (i.e., makes fewer mistakes) than 0,0 or 1,1?

Counter Example 1: 0.5,0.5 Is Closest, But Isn't Most Accurate



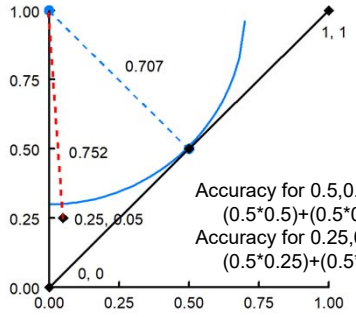
When p=0.5, sens and 1-spec pairs of 0,0, 0.5,0.5, and 1,1 all have same accuracy (0.5) *

* 0.5, 0.5: (0.5 * 0.5) + (0.5 * 0.5) = 0.5; 0.0, 0.0: (0.0 * 0.5) + (1.0 * 0.5) = 0.5; 1.0, 0.0: (1.0 *0.5) + (0.0 * 0.5) = 0.5

Counterexample 2

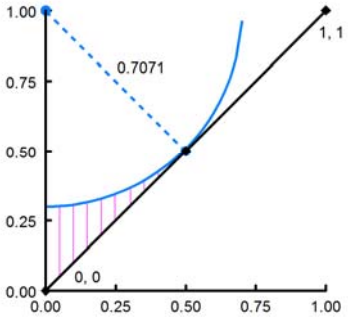
- Consider a 4th point, sens=0.25, 1-spec=0.05, which is modestly discriminating with a ROC area of 0.60
- Uninformative point 0.5,0.5 has smaller distance (0.707) than does 0.25,0.05 (0.752)
- When prevalence is around 50%, is 0.5,0.5 more accurate than 0.25,0.05?

NO! 0.25,0.05 Further Away But More Accurate

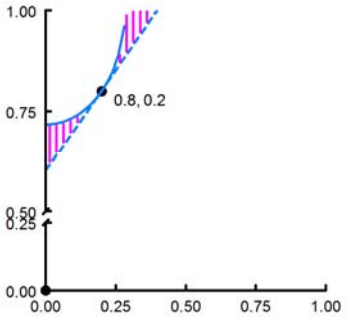


Accuracy for 0.5,0.5 = 0.5
(0.5*0.5)+(0.5*0.5)
Accuracy for 0.25,0.05 = 0.6
(0.5*0.25)+(0.5*0.95)

All Points in Shaded Area Feasible, Further From NE Corner, and More Accurate Than 0.5,0.5



More generally, if closest point is on a 45º line other than 0/0-1/1, any point in shaded area both further away and more accurate



Conclusions:

Point closest to northwest corner of ROC curve need not maximize accuracy when p=0.5

Not sure why accuracy at p=0.5 is correct statistic

Not sure why we aren't maximizing NBPT instead of accuracy

Method 3: Youden Index (sens + spec - 1) *

- Unlike Sackett, Youden Index does identify test cut-off that makes "the fewest mistakes when prevalence is at or around 50%"

Youden WJ. Index for rating diagnostic tests. Cancer. 1950; 3: 32-35.

Youden Index Example

Sens, Spec	Youden Index	Accuracy
0.9, 0.5	0.4	0.7 = (0.9 × 0.5) + (0.5 × 0.5)
0.8, 0.59	0.39	0.695 = (0.8 × 0.5) + (0.59 × 0.5)
0.59, 0.8	0.39	0.695 = (.59 × .5) + (.8 × 0.5)

- Youden Index succeeds because Youden Index = $(2*\text{Accuracy}_{p=0.5}) - 1$
- $\text{Accuracy}_{p=0.5} = (p*\text{sens}) + ((1-p)*\text{spec}$ where p=0.5
  - Multiply through by 2; subtract 1 from both sides
- Maximizing linear transformation of $\text{accuracy}_{p=0.5}$ necessarily maximizes $\text{accuracy}_{p=0.5}$

What If Prevalence ≠ 0.5 (e.g., P = 0.4)?

Sens, Spec	Youden Index	Accuracy
0.9, 0.5	0.4	0.66 = (0.9 × 0.4) + (0.5 × 0.6)
0.8, 0.59	0.39	0.674 = (0.8 × 0.4) + (0.59 × 0.6)
0.59, 0.8	0.39	0.716 = (.59 × .4) + (.8 × 0.6)

- If, however, p can take on values other than 0.5, accuracy formula (or if linear transformations are preferred, (2*accuracy – 1)) should be used to replace Youden index
- But if accuracy assessment is not limited to a single pre-test probability, would no longer have 1 statistic per cut-off
  - Instead would need multiple prevalence-dependent statistics

Conclusions:

Even though Youden index has limited use as a value metric, it's still cited in literature (686 Ovid references between 2015-2020

Not sure why linear transformation of accuracy at p=0.5 is correct statistic

Not sure why we're not maximizing NBPT rather than accuracy measures

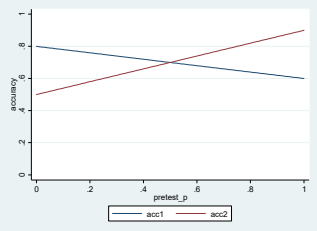
Method 4: Diagnostic Odds Ratio (DOR)

- DOR = (sens*spec)/((1-sens)*(1-spec))
- As with Youden index, DOR is independent of pre-test probability
- However,
  - For test with dichotomous results, always recommends testing
  - For test with multiple possible cut-offs, always identifies 1 that is best for all patients
  - For multiple tests (with one cut-off for each), always identifies 1 that is best for all patients
- But 2x2 table with largest DOR need not maximize accuracy for all pre-test probabilities of disease and need not have largest NBPT

Diagnostic Odds Ratio Example

- Suppose test 1 has a sens of 0.6 and a spec of 0.8 while test 2 has a sens of 0.9 and a spec of 0.5
- Test DORs equal:
  - Test 1 DOR = 6 ((0.6*0.8)/(0.4*0.2))
  - Test 2 DOR = 9 ((0.9*0.5)/(0.1*0.5))
- Based on larger DOR choose test 2

Larger DOR Not Necessarily Mean More Accurate



- Test 1 more accurate for p<0.5; Test 2 more accurate for p>0.5; accuracy identical for p=0.5

Conclusions:

Even though DOR has limited use as a value metric, it's still cited in literature (928 Ovid references between 2015-2020

Only maximizes accuracy for some pretest probabilities

Not sure why we're not maximizing NBPT rather than accuracy

Method 5: Point Where Sens = Spec

- 0.5, 0.5 is one of points where sens=spec
- (In response to method 2,) we've already seen that this point needn't maximize accuracy

PROPERTIES OF IDEAL STATISTIC

Properties of an Ideal Statistic (Ideal Statistics)

1) Single statistic that identifies best test for all patients, no matter what their pretest probability (p) nor what treatment threshold (p*) applies to them
   - Ideal **OBTAINABLE** statistic(s) – single graph or table that provides statistics for all relevant pretest probabilities and treatment thresholds
2) Statistics are characteristics of tests whose properties have same stability (instability) as sensitivity, specificity, and likelihood ratios

Properties of an Ideal Statistic (Ideal Statistics) (2)

3) Statistics independent of development of new cost-effective treatments
   - Cost-effectiveness should affect p* that is appropriate for a particular patient, but not height of curve at p*
4) Statistics allow determination of complete ranking of testing strategies
   - e.g., that testing is superior to treating no one which is superior to treating everyone
5) Statistics allow determination of relative (or absolute) difference in outcomes among testing strategies
   - Latter in part to address issues related to inclusion of cost of test

Properties of an Ideal Statistic (Ideal Statistics) (3)

6) Statistics unaffected by pre-test probability in sample used to develop test
   - "…index is independent of the relative sizes of the control and diseased groups" (Youden)
   - **True of any statistic based on sensitivity and specificity or LRs**
7) Statistics unaffected by treatment threshold
8) Possible to calculate a standard error for statistics (Youden)

BETTER METHODS

FUNDAMENTALS

Treatment Threshold (p*)

- p* where value of no Rx equals value of Rx:

$$\rightarrow p^* = \frac{\Delta O_{D-}}{\Delta O_{D-} + \Delta O_{D+}}$$

Same p* we've seen before

Additional Transformation †

- (Important for Vickers and Elkin's decision curves and net benefit football)

$$\rightarrow \frac{p^*}{1-p^*} = \frac{\Delta O_{D-}}{\Delta O_{D+}}$$

- Refer to $\Delta O_{D-} / \Delta O_{D+}$ as "ratio of differences in outcomes" or the "ratio of differences"
  - When either of the 2 ratios less than 1 indicate that the difference in outcomes among persons with disease is greater than the difference in outcomes among persons without disease
  - For ratios greater than 1, reverse is true

† Derivation in appendix

Relative Value?

- Previously noted **DON'T** need to know exact magnitudes of differences in outcomes
  - Instead only require information about relative cost of differences in outcomes
    - i.e., can set ΔO_{D+} equal to 1 and express ΔO_{D-} as a multiple of ΔO_{D+}
    - e.g., ΔO_{D-} is half ΔO_{D+} ($\Delta O_{D-} = 0.5$ and $\Delta O_{D+} = 1$)

$$\text{If } \Delta O_{D+} = 1\text{: } \frac{p^*}{1-p^*} = \Delta O_{D-}$$

- Issues arise for incorporating testing cost when Δs incorporated as relative rather than absolute terms?
  - i.e., need relative testing costs

Expected Outcome of Testing

- $p \times sens \times O_{D+T+} + p \times (1\text{-}sens) \times O_{D+T-} + (1\text{-}p) \times spec \times O_{D-T-} + (1\text{-}p) \times (1\text{-}spec) \times O_{D-T+}$
- Test cost? $[-T_c]$

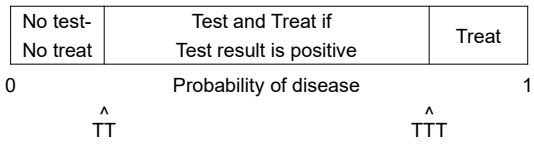
Better Method 1: Pauker and Kassirer Test Thresholds

Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980; 302: 1109-17.

Do Nothing / Test and Test / Treat Thresholds

- In 1980 Pauker and Kassirer used equations to define do nothing/test and test/treatment thresholds (here referred to more generally as “test thresholds”)
  - Early graphical method for describing value of a test
- Graph provides information about ranges of probabilities for which treating no one, testing, and treating everyone has greatest net benefit
- Typical graph compares these 3 strategies alone, but possible to evaluate more than 3 strategies (with ranges of probabilities where each of 3+ strategies has largest net benefit)

Test Threshold Graph (We’ve Seen This Before)



Do Nothing / Test Threshold

- Defined by setting expected value of treating no one equal to expected value of testing and solving for p

$$p \times O_{D+T-} + (1-p) \times O_{D-T-} = p \times sens \times O_{D+T+} + p \times (1-sens) \times O_{D+T-} + (1-p) \times spec \times O_{D-T-} + (1-p) \times (1-spec) \times O_{D-T+}$$

- Testing cost?

Solving for Do Nothing / Test Threshold

- Without testing Cost (e.g., testing is costless)

$$p_{tt} = \frac{(1\text{-spec})\ \Delta O_{D-}}{(1\text{-spec})\ \Delta O_{D-} + \text{sens}\ \Delta O_{D+}}$$

- With (positive) testing cost

$$p_{tt} = \frac{(1\text{-spec})\ \Delta O_{D-} + T_c}{(1\text{-spec})\ \Delta O_{D-} + \text{sens}\ \Delta O_{D+}}$$

- Except for addition of test characteristics and T_c, equations defining p_{tt} same as equation defining p^*
- T_c enlarges region where no test/no treat is preferred

Test / Treatment Threshold

- Defined by setting expected value of testing equal to expected value of treating everyone and solving for p

$$p \times \text{sens} \times O_{D+T+} + p \times (1\text{-sens}) \times O_{D+T-} + (1-p) \times \text{spec} \times O_{D-T-} + (1-p) \times (1\text{-spec}) \times O_{D-T+} = p \times O_{D+T+} + (1-p) \times O_{D-T+}$$

- Testing cost?

Solving for Test / Treatment Threshold

- Without testing Cost (e.g., testing is costless)

$$p_{ttt} = \frac{\text{spec}\ \Delta O_{D-}}{\text{spec}\ \Delta O_{D-} + (1\text{-sens})\ \Delta O_{D+}}$$

- With testing cost

$$p_{ttt} = \frac{\text{spec}\ \Delta O_{D-} - T_c}{\text{spec}\ \Delta O_{D-} + (1\text{-sens})\ \Delta O_{D+}}$$

- Except for addition of test characteristics and T_c, equations defining p_{ttt} same as equation defining p^*
- T_c enlarges region where treatment is preferred

Generalized Threshold Equations

- Threshold for any pair of (costless) testing strategies:

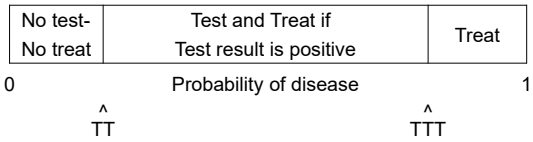
$$p_t = \frac{(spec_1 - spec_2)\ \Delta O_{D-}}{(spec_1 - spec_2)\ \Delta O_{D-} + (sens_2 - sens_1)\ \Delta O_{D+}}$$

- Threshold for any pair of strategies (with test cost):

$$p_t = \frac{(spec_1 - spec_2)\ \Delta O_{D-} + (T_{c2} - T_{c1})}{(spec_1 - spec_2)\ \Delta O_{D-} + (sens_2 - sens_1)\ \Delta O_{D+}}$$

- Formulae identify p_t but don't indicate which test is better when p greater than or less than p_t
  - Strategy with larger sensitivity superior for $p > p_t$
  - Strategy with smaller sensitivity superior for $p < p_t$

Test Threshold Graph



- Strength: reports most net beneficial diagnostic strategy for any pretest probability of disease
- Weakness: A single graph assumes a constant set of values for ΔO_{D-}, ΔO_{D+}, and test cost
  - Requires separate graphs as values change
  - Thus, no single graph or table provides statistics for all relevant pretest probabilities and treatment thresholds

Conclusions:

Test Thresholds identify strategies with maximum NBPT for full range of probability of disease

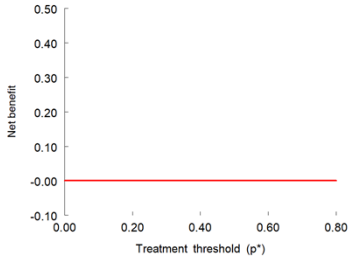
But different combinations of ΔO_{D-}, ΔO_{D+}, and test cost require different graphs

Better Method 2:
Vickers and Elkin's Decision Curves

Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. Med Decis Making. 2006; 26: 565-74.

Decision Curves (Vickers and Elkin, 2006)

- Plot net benefits of a positive test on graph whose X axis represents p* and whose Y-axis represents net benefits of positive test



Decision Curves (2)

- Originally proposed to compare value of different prediction rules
- While some differences may exist between valuation of prediction rules vs diagnostic tests, fundamental principles unchanged

Vickers and Elkin's X Axis (p*)

- In theory V&E's p* same as our p*
  - Represents probability above which treatment has a greater net benefit than withholding treatment and below which withholding treatment has a greater net benefit than treatment
  - Defined as $\Delta O_{D-} / (\Delta O_{D-} + \Delta O_{D+})$
  - Required if optimal test is to maximize net benefit of a positive test

Vickers and Elkin's X Axis (p*) (2)

- But Vickers and Elkin aren't so committed to our definition
  - Express concern (as have previous members of our class) about estimating p*
  - Discuss ranges of p* and clinicians' subjective judgments as reasonable values

Vickers and Elkin's Y Axis (NBPT or NB)

- As we've previously defined it, net benefit from a positive test equals:

  $NB = p \times sens \times \Delta O_{D+} - (1-p) \times (1-spec) \times \Delta O_{D-}$

- This version of NBPT equation does not allow plotting of NB as a function of p*

Vickers and Elkin's Y Axis (NBPT or NB)

- Vickers and Elkin's substitute following equation for NB:

$$NB = \frac{\text{true positive count}}{N} - \frac{\text{false positive count}}{N}\,\frac{p^*}{1-p^*}$$

- where N equals number of individuals in study sample, true positive count equals N p sens and false positive count equals N (1-p) (1-spec)

V&E INSIGHT: MAKE NB A FUNCTION OF P*

Which is possible because P* defined by cost differences

Alternative Version of Vickers and Elkins Equation

$$NB = \frac{\text{true positive count}}{N} - \frac{\text{false positive count}}{N}\,\frac{p^*}{1-p^*}$$

- Substitute N p sens for true positive count
- Substitute N (1-p) (1-spec) for false positive count
- Cancel N/N

$$NB = p\,sens - (1-p)\,(1-spec)\,\frac{p^*}{1-p^*}$$

- Begins to look similar to equations we use

Alternative Version of Vickers and Elkins Equation (2)

- Previously indicated that (derivation in appendix slides)

$$\rightarrow \frac{p^*}{1-p^*} = \frac{\Delta O_{D-}}{\Delta O_{D+}}$$

- Substitute $\Delta O_{D-} / \Delta O_{D+}$ for p*/1-p*
- Divide both ΔO_{D+} and ΔO_{D-} by ΔO_{D+} (allows cancellation of ΔO_{D+} and makes ΔO_{D-} a relative value of ΔO_{D+})
- Set ΔO_{D+} equal to 1 and multiply p sens by 1 (i.e., by ΔO_{D+}

$$NB = p\,sens\,\Delta O_{D+} - (1-p)\,(1-spec)\,\Delta O_{D-}$$

- Resulting equation matches our original formula for calculating NBPT (see ROC curve lecture notes)

Formula for Treat No One, Test, and Treat All

- Assume p = 0.2; treat no one sensitivity = 0.0; treat no one specificity = 1.0; test sensitivity = 0.75; test specificity = 0.85; treat everyone sensitivity = 1.0; treat everyone specificity = 0.0

Treat No one

$$NB = (0.2 * 0) - \left(0.8 * 0 * \frac{p^*}{1-p^*}\right)$$

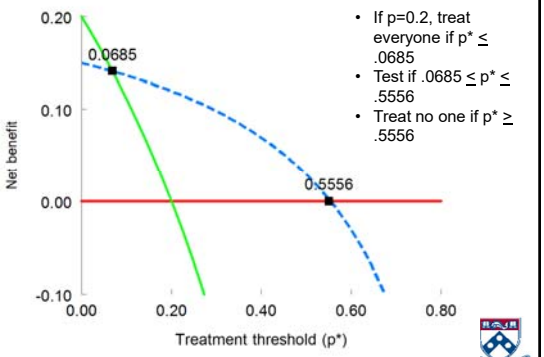
Test

$$NB = (0.2 * 0.75) - \left(0.8 * 0.15 * \frac{p^*}{1-p^*}\right)$$

Treat All

$$NB = (0.2 * 1.0) - \left(0.8 * 1 * \frac{p^*}{1-p^*}\right)$$

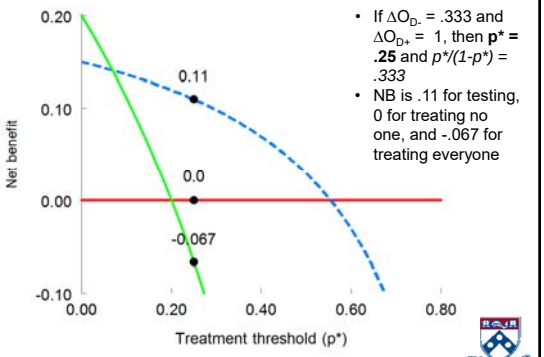
Decision Curve Graph (**p=0.2**)



- If p=0.2, treat everyone if p* ≤ .0685
- Test if .0685 ≤ p* ≤ .5556
- Treat no one if p* ≥ .5556

NB, treat no one; NB, Treat everyone; NB, test

Allows Calculation of NB for Any Operating Point

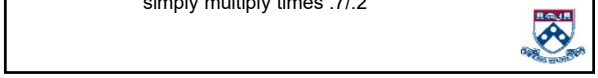


- If $\Delta O_{D-} = .333$ and $\Delta O_{D+} = 1$, then **p* = .25** and *p*/(1-p*) = .333*
- NB is .11 for testing, 0 for treating no one, and -.067 for treating everyone

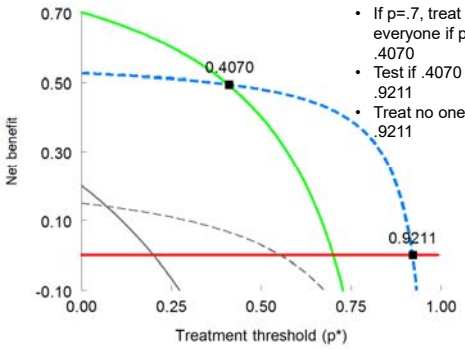
NB, treat no one; NB, Treat everyone; NB, test

Decision Curves and Pre-Test Probability

- Prior decision curve plotted for p=0.2
- Curve does not describe net benefits for other pre-test probabilities
  - Shapes of curves for other probabilities of disease, can change dramatically
  - No simple transformation that allows use of one curve to construct other curves
    - e.g., cannot simply multiply 20% curve times ratio of probabilities of disease used to construct the two sets of curves
      - i.e., to derive 70% curve from 20% curve, can't simply multiply times .7/.2



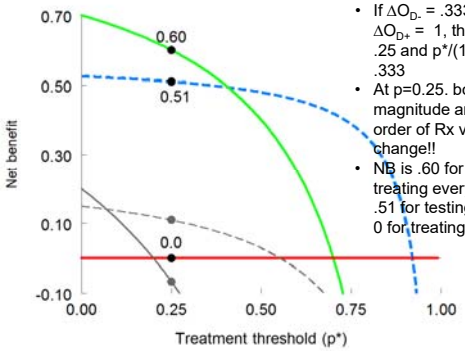
Decision Curve Graph (p=0.7)



- If p=.7, treat everyone if p* ≤ .4070
- Test if .4070 ≤ p* ≤ .9211
- Treat no one if p* ≥ .9211

NB, treat no one; NB, Treat everyone; NB, test; NB curves for p=0.2

Allows Calculation of NB for Any Operating Point



- If ΔO_{D-} = .333 and ΔO_{D+} = 1, then p* = .25 and p*/(1-p*) = .333
- At p=0.25. both magnitude and order of Rx values change!!
- NB is .60 for treating everyone, .51 for testing, and 0 for treating no one

NB, treat no one; NB, Treat everyone; NB, test; NB curves for p=0.2

Conclusions:

Decision Curves identify strategies with maximum NBPT for full range of (relative) costs

But different values of p require different graphs

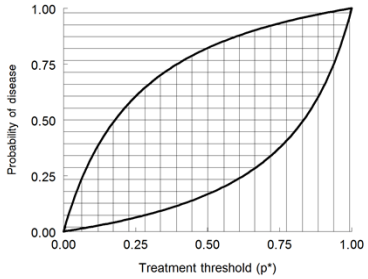
Better Method 3

Net Benefit Football

Net Benefit Football

- Recommended that reporting (relative) difference in value of test strategies is one of the properties for an ideal statistic, but…
- Can provide information about value for all p* (i.e., cost ratios) and all probabilities of disease in single graph if change Y-axis from NB to probability of disease
- Net benefit football would have treatment thresholds (p*) on X-axis, pretest probabilities of disease on Y-axis, and graph would identify p*/probability pairs for which the different strategies have the largest net benefit

Net Benefit "Football", 3 Diagnostic/Treament Stategies



- When test sensitivity=0.75 and test specificity=0.85, treat all in area with horizontal lines, test in area with cross-hatched lines, and treat no one in area with vertical lines

Football Based on Test Thresholds

Start With Do Nothing / Test Threshold Equation

- Without testing Cost (e.g., testing is costless)

$$p_{tt} = \frac{(1\text{-spec})\ \Delta O_{D-}}{(1\text{-spec})\ \Delta O_{D-} + \text{sens}\ \Delta O_{D+}}$$

- Divide numerator and denominator of threshold equation by ΔO_{D+}
- Yields $\Delta O_{D-}/\Delta O_{D+}$ in numerator and $\Delta O_{D-}/\Delta O_{D+}$ and 1 in the denominator
- On slide 34, we showed that $\Delta O_{D-}/\Delta O_{D+} = p^*/(1-p^*)$ which we can substitute into the equation

Revised Thresholds Equations

- Resulting Do nothing/test threshold (excluding testing Cost)

$$p_{tt} = \frac{(1\text{-spec})\ p^* / (1\text{-}p^*)}{((1\text{-spec})\ p^* / (1\text{-}p^*)) + \text{sens}}$$

- Resulting Test / Treatment threshold (excluding testing cost)

$$p_{ttt} = \frac{\text{spec}\ p^* / (1\text{-}p^*)}{(\text{spec}\ p^* / (1\text{-}p^*)) + (1\text{-sens})}$$

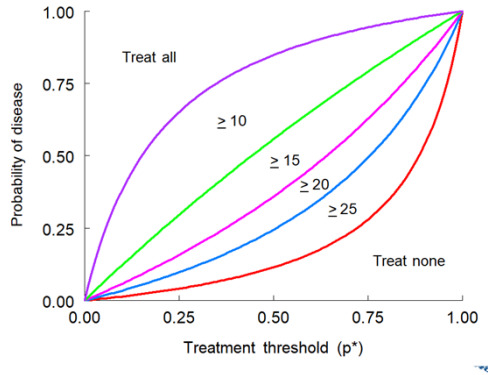
Generalized Football Threshold Equations

- Threshold (p_t) for any pair of (costless) testing strategies (including all results negative and all results positive):

$$p_t = \frac{(\text{spec}_2 - \text{spec}_1)\ \frac{p^*}{1\text{-}p^*}}{(\text{spec}_2 - \text{spec}_1)\ \frac{p^*}{1\text{-}p^*} + (\text{sens}_1 - \text{sens}_2)}$$

- Formula identifies p_t for different values of p* and p, but doesn't indicate which test falls on which side of p_t
  - Strategy with larger sensitivity superior for combinations of p* and p that fall to the left and above boundary
  - Strategy with smaller sensitivity superior for combinations of p* and p that fall to right and below boundary

WBC Net Benefit "Football"



Football Pluses and Minuses

- Pluses
  - Single graph for each test
  - Accounts for different treatment thresholds
  - Accounts for different pretest probabilities
  - Stable characteristics of test (if test cost excluded)
  - Independent of development of new therapies
- Minuses
  - ? Doesn't report a complete ranking of test strategies ?
  - Doesn't report (relative) difference in value of test strategies
- Plus and minus
  - Can incorporate test cost, but shape of of nothing and treat all regions change when test cost changes

Equation Appendix

$p^* = \Delta O_{D-} / (\Delta O_{D-} + \Delta O_{D+})$

Value of treating everyone

$p\ O_{D+Rx+} + (1-p)\ O_{D-Rx+}$

Value of treating no one

$p\ O_{D+Rx-} + (1-p)\ O_{D-Rx-}$

Set values equal and solve for p*

$p\ O_{D+Rx+} + (1-p)\ O_{D-Rx+} = p\ O_{D+Rx-} + (1-p)\ O_{D-Rx-}$

$p\ O_{D+Rx+} - p\ O_{D+Rx-} = (1-p)\ O_{D-Rx-} - (1-p)\ O_{D-Rx+}$

$p\ (O_{D+Rx+} - O_{D+Rx-}) = (1-p)\ (O_{D-Rx-} - O_{D-Rx+})$

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

$p\ \Delta O_{D+} = \Delta O_{D-} - p\ \Delta O_{D-}$

$p\ \Delta O_{D-} + p\ \Delta O_{D+} = \Delta O_{D-}$

$p\ (\Delta O_{D-} + \Delta O_{D+}) = \Delta O_{D-}$

$$p^* = \frac{\Delta O_{D-}}{\Delta O_{D-} + \Delta O_{D+}}$$

$p^*/(1-p^*) = \Delta O_{D-}/\Delta O_{D+}$

Start with definition of p^*:

$$p^* = \frac{\Delta O_{D-}}{\Delta O_{D-} + \Delta O_{D+}}$$

$$p^* (\Delta O_{D-} + \Delta O_{D+}) = \Delta O_{D-}$$

$$p^* \Delta O_{D-} + p^* \Delta O_{D+} = \Delta O_{D-}$$

$$p^* \Delta O_{D+} = \Delta O_{D-} - p^* \Delta O_{D-}$$

$$p^* \Delta O_{D+} = \Delta O_{D-} (1 - p^*)$$

$$\frac{p^*}{(1 - p^*)} = \frac{\Delta O_{D-}}{\Delta O_{D+}}$$

$DS = ((1-p)/p) \times (p^*/(1-p^*))$

$$DS = \frac{(1-p)}{p} \times \frac{\Delta O_{D-}}{\Delta O_{D+}}$$

In Appendix slide 2, showed that $(\Delta O_{D-}/\Delta O_{D+}) = (p^*/(1-p^*))$

Substitute $(p^*/(1-p^*))$ for (O_{D-}/O_{D+})

$$DS = \frac{(1-p)}{p} \times \frac{p^*}{1-p^*}$$