The Scientific Contributions of Charles E. Metz: ROC Analysis and Reader Study Design

Alicia Y. Toledano, Sc.D.

27 August 2012

Curriculum Vitae

- As Charlie and I discovered one day, this is *singular* – it means "the course of one's life"
  - The plural is curricul*a* vitae

wicked-trees.jpg: dublinlibrary.wordpress.com

Tree-Pic.jpg: coachingbasketballwisely.com

maple-leaf.jpg: comicfury.com

SV_–_NorthAmericaSatellite.png: althistory.wikia.com

ROC Curve: Signal and Noise

- Two groups: *with (signal)* and *without (noise)*
- Test result using a cut point on an underlying continuum (strength): *positive* or *negative*

✓ TPF = (# with positive)/(# with)

× FPF = (# without positive)/(# without)

- Trade-off:
  - Less stringent cut point ↑ TPF; also ↑ FPF
  - More stringent cut point ↓ FPF; also ↓ TPF

blipsonradar.png: thatdamnredhead.net

ROC Curve

"… imagine repeatedly changing the decision threshold and obtaining more and more points … the points representing all possible combinations of TPF and FPF must lie on a curve. This curve is called the receiver operating characteristic (ROC) curve for the diagnostic test, since it describes the inherent detection *characteristics* of the test (…) and since the *receiver* of the test information can *operate* at any point on the curve by using an appropriate decision threshold."

Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

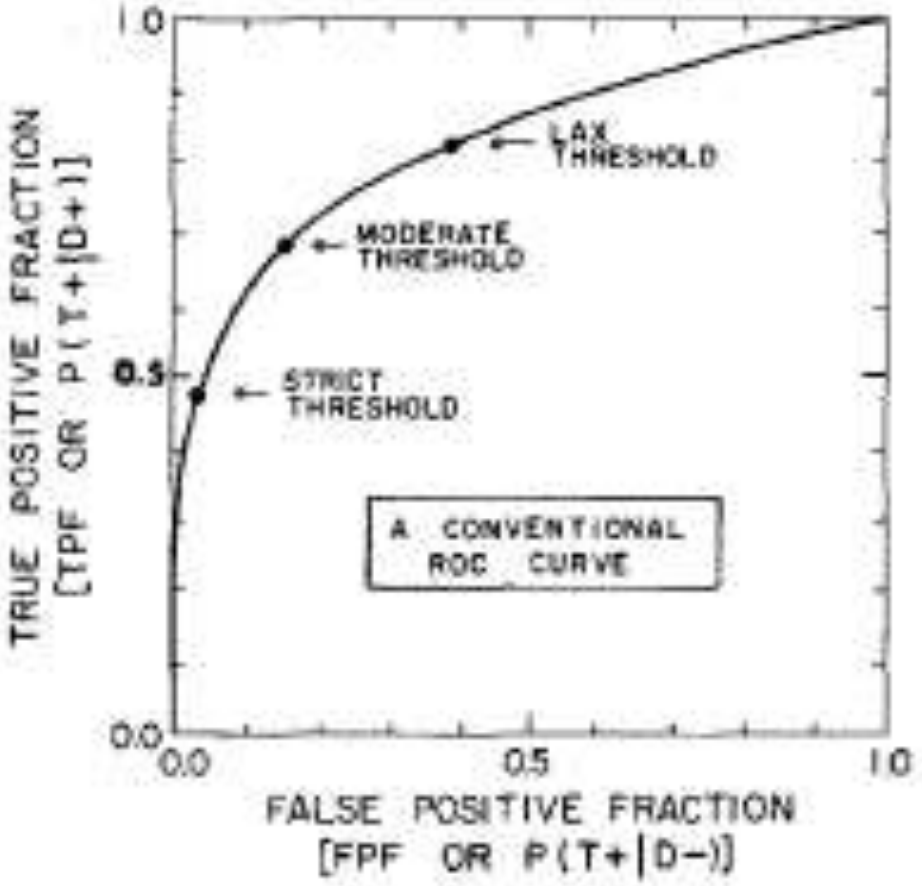


Fig. 2. A typical conventional ROC curve, showing three possible operating points.

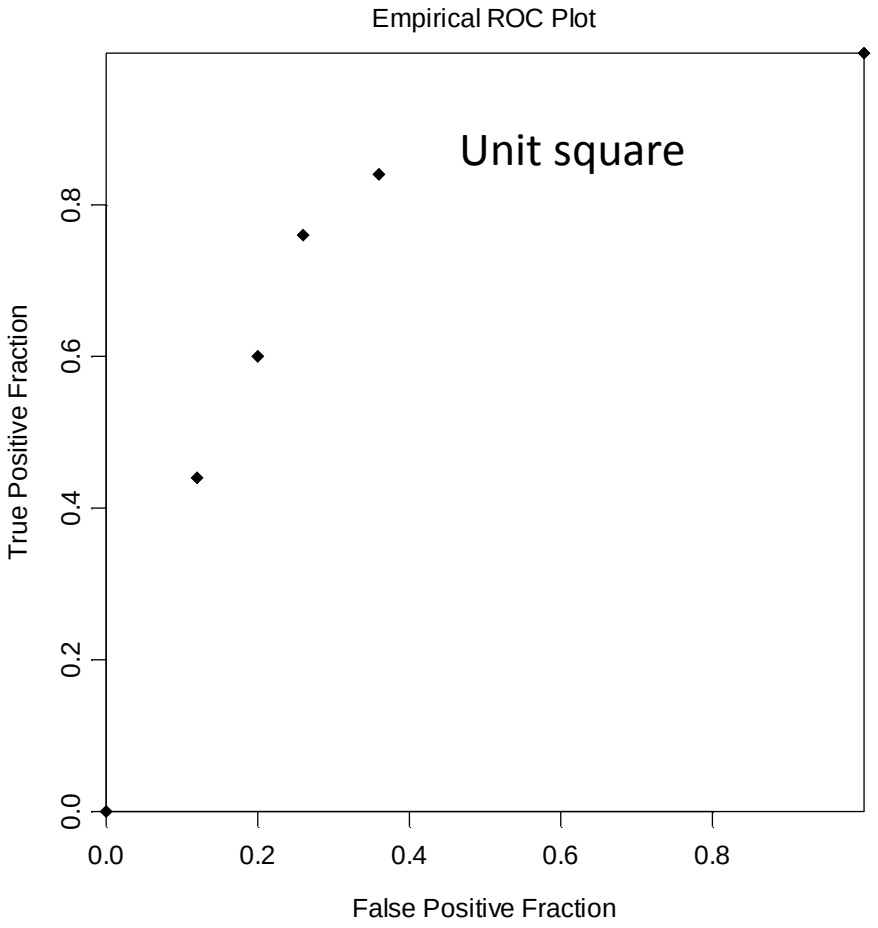
Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

Fitting ROC Curves in Practice

- Radiology studies: ordinal scales until about Y2000, then (also) continuous scales
  - Patient-reported outcomes: ordinal scales
- Even on a continuous scale, we only see a limited number of decision thresholds used in a study sample
- For now, look at ordinal

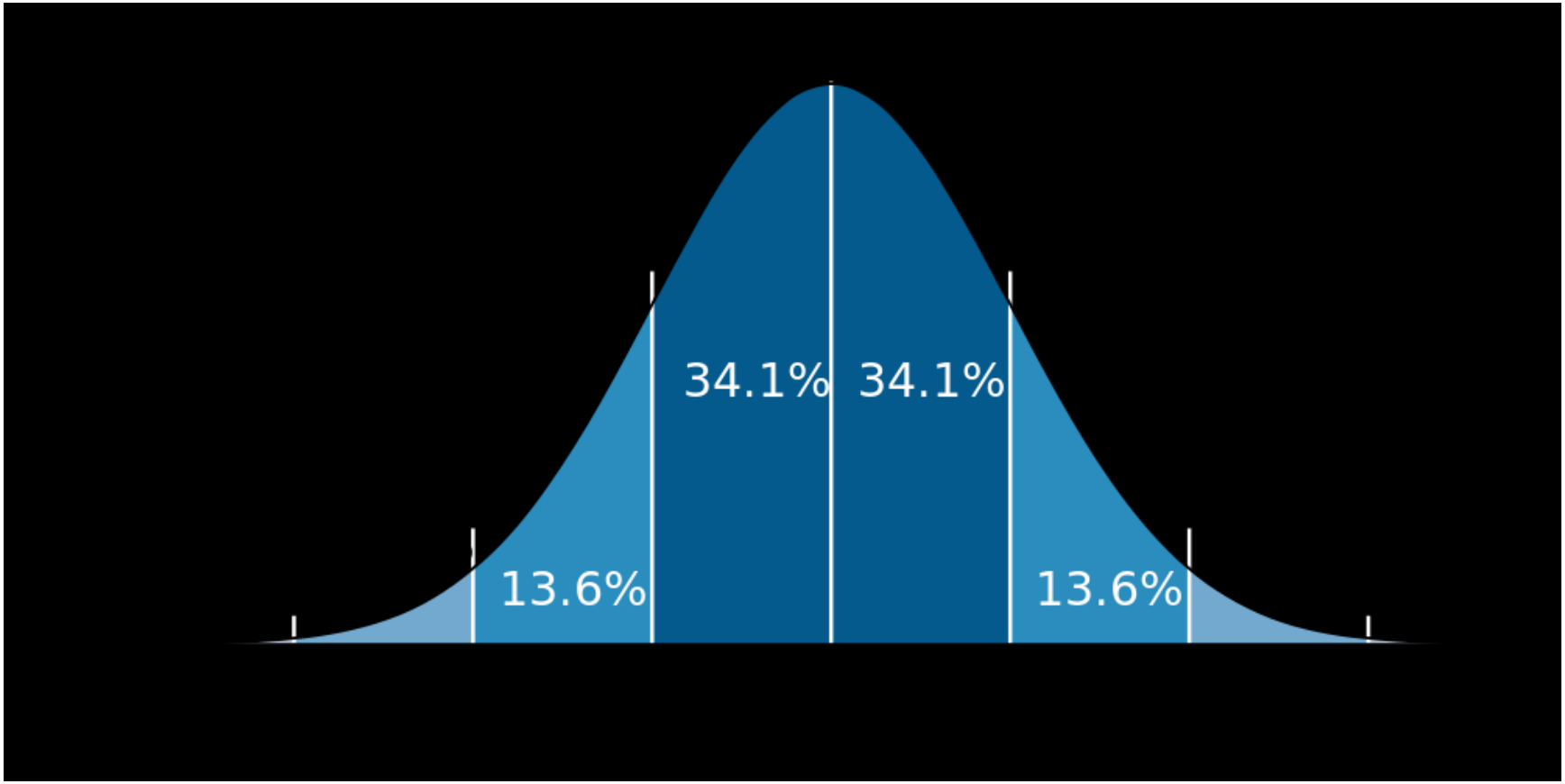
	1	2	3	4	5
Without	32	5	3	4	6
With	8	4	8	8	22

Observed Operating Points



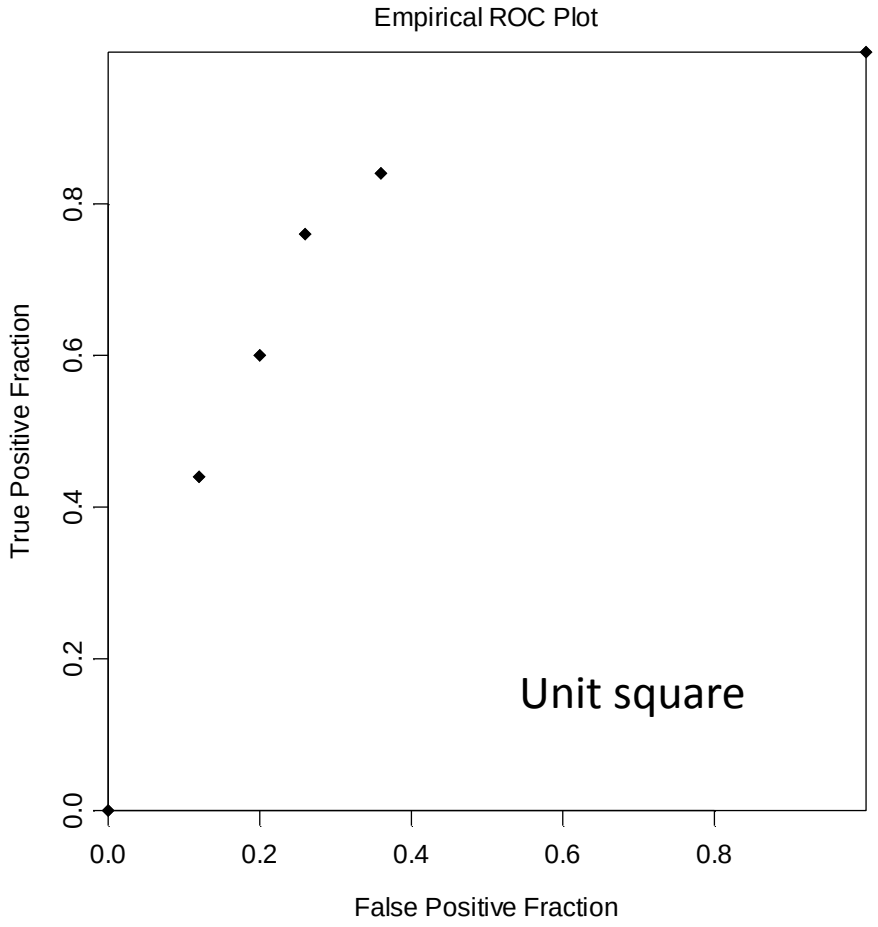
Normal Distribution

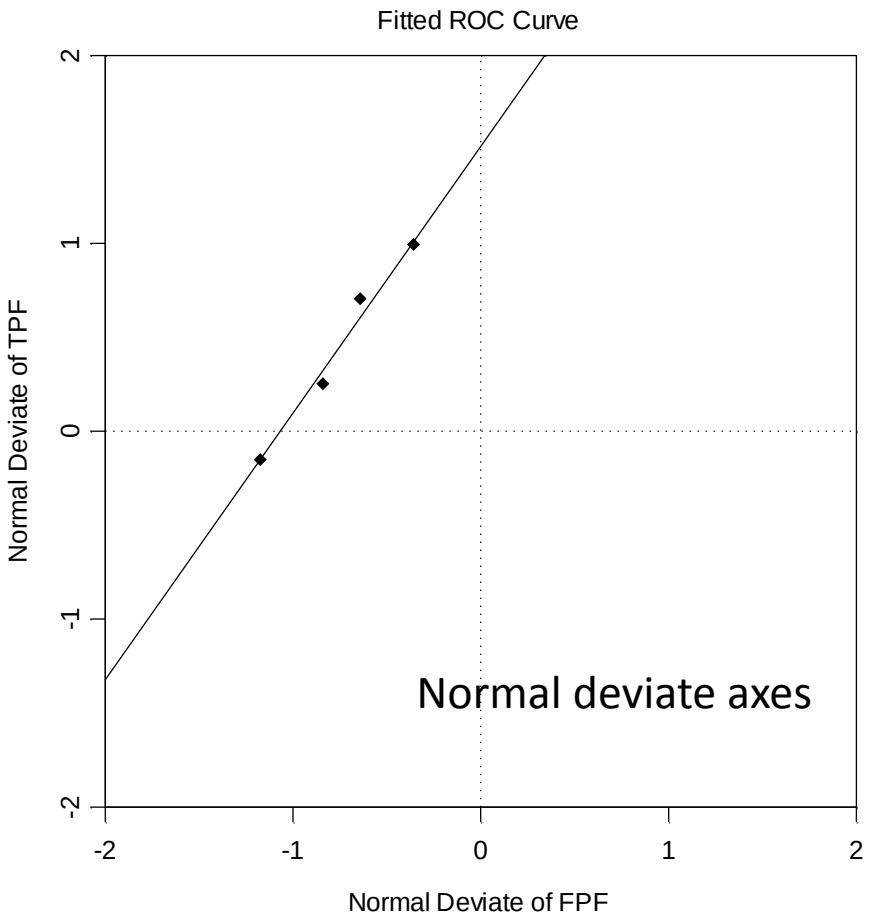
- Eventually, everything becomes normal…



Standard_deviation_diagram.svg: en.wikipedia.org/wiki/Normal_distribution

Observed Operating Points



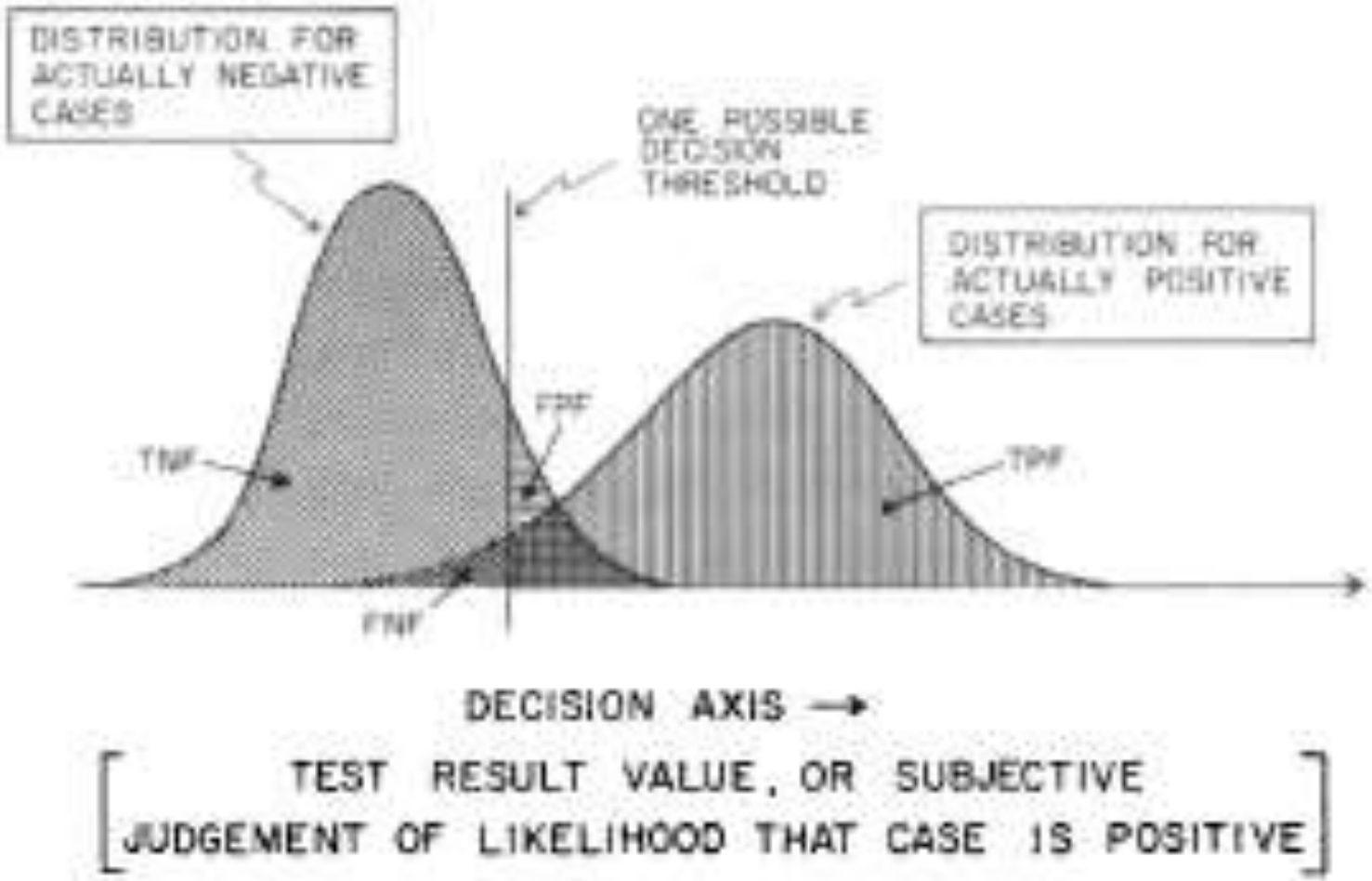


Invariance Property

- The ROC curve is invariant to monotonic transformations
  - *Monotonic* function: function between ordered sets that preserves the given order
  - If order of signal strength is preserved, get the same ROC curve
    - Multiply/divide, add/subtract, square root (NOT square), …
  - *Monotonous*: lacking in variety; tediously unvarying

The Binormal Model

- Invariance: underlying distributions specify an ROC curve, but ROC curve does not uniquely specify pair of underlying distributions
  - Many-to-one mapping
- *Binormal assumption*: ROC curve plots as a straight line on a pair of normal deviate axes
  - ROC curve has same functional form as would be generated by two normal distributions
  - Actual form of latent distributions is not specified



Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

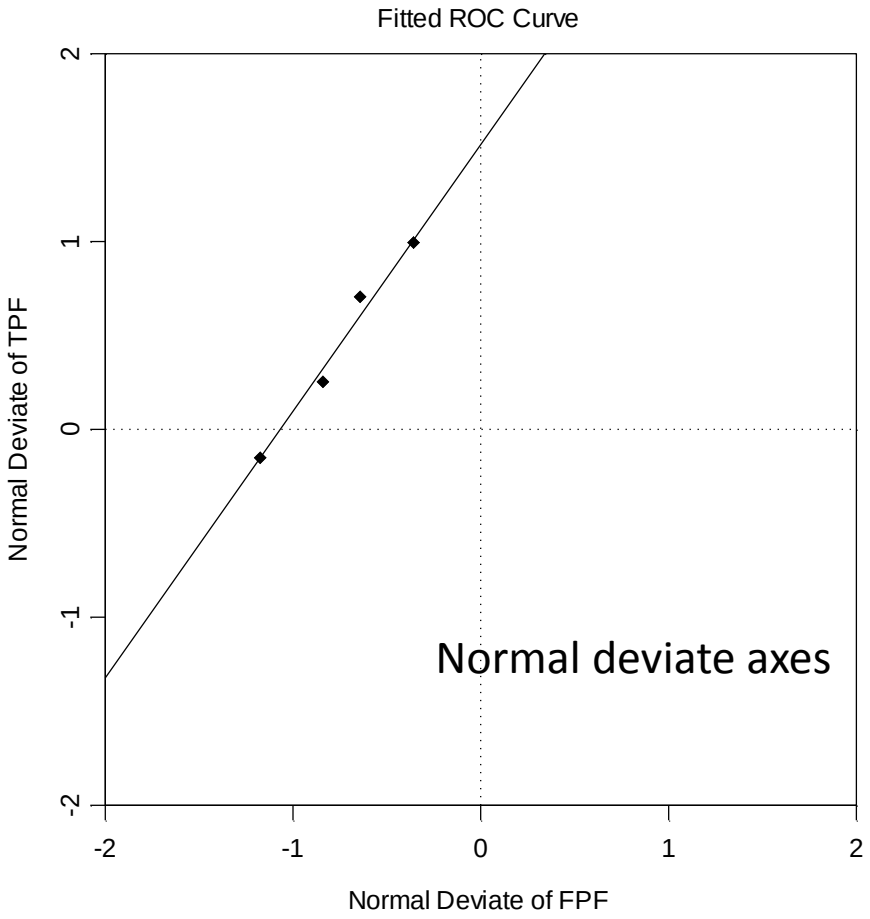
The Binormal Model

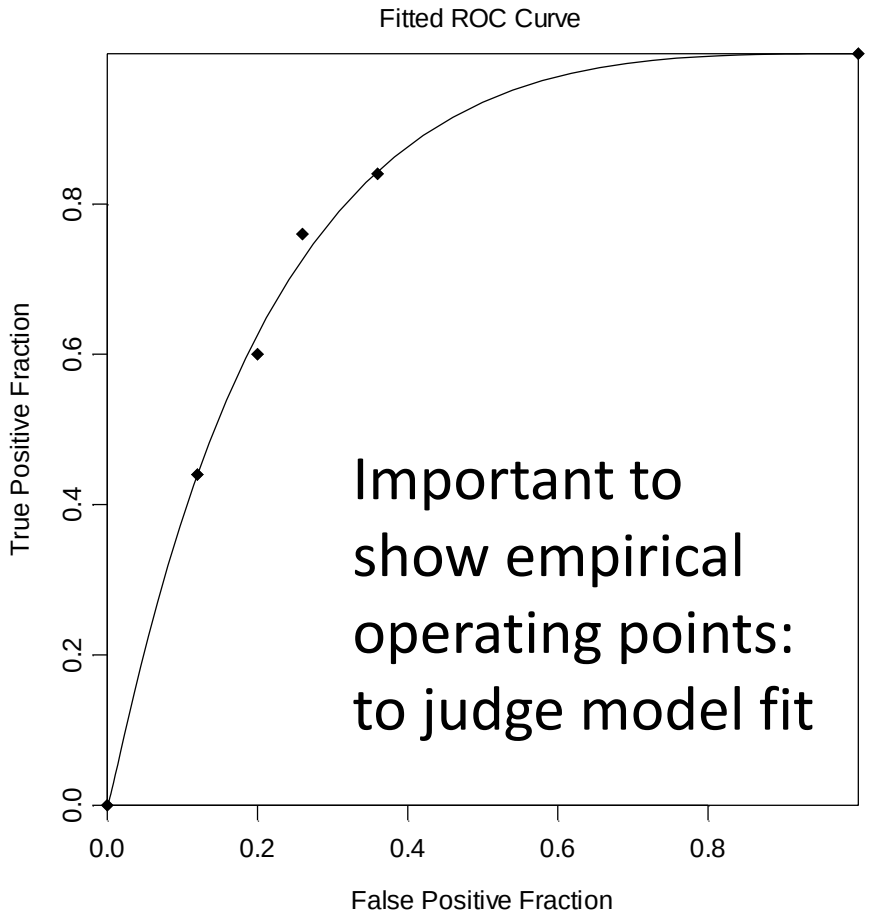
- Statistical model: intercept *a* and slope *b* on normal deviate axes

Par.	Distributions	ROC in unit square
a	Distance between centers, in units of SD_{with}	Proximity to top left corner (perfection!)
b	Ratio of widths, $SD_{without}/SD_{with}$	Symmetry around negative diagonal (b=1: symmetric)

- Complexity: cannot do usual statistical fit to straight line on normal-normal axes because *z(FPF)* estimated from sample, not "known"

Fitted ROC Curve





Anchored at Corners

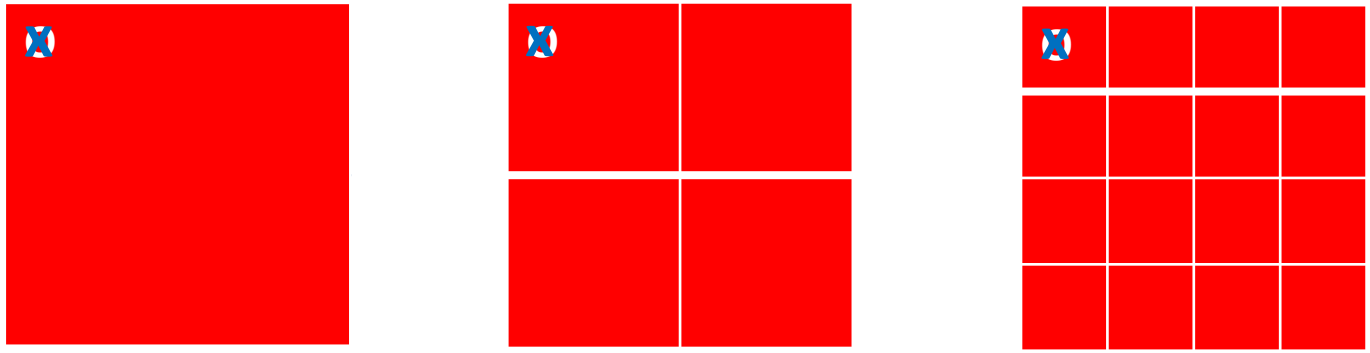
"… inevitably must pass through the lower left corner (FPF = 0, TPF = 0) of the graph because all tests can be called negative, and through the upper right corner (FPF = 1, TPF = 1) of the graph because all tests can be called positive."

Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

- Relates to tasks requiring localization: LROC does not necessarily reach (1, 1)

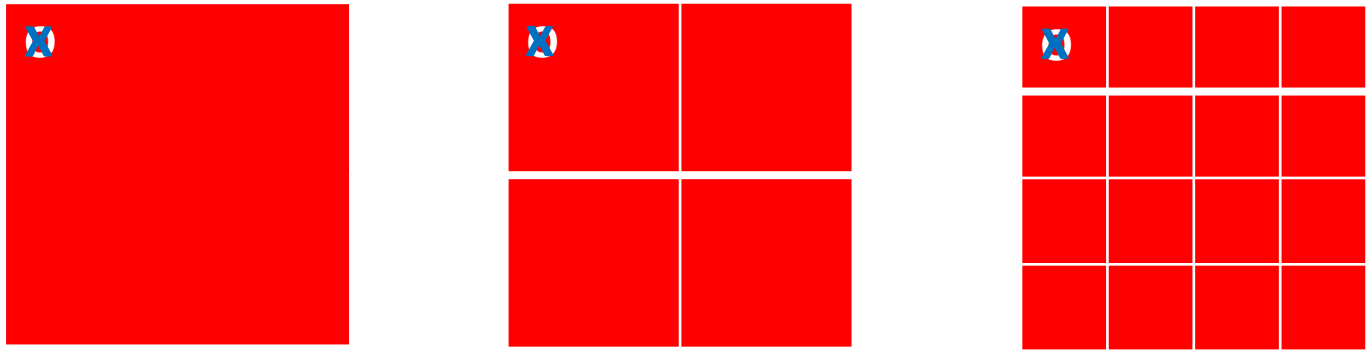
Localization: One Actual Lesion

- Combination of threshold for positive test and acceptance region for location of reader mark
  - Within mutually exclusive and exhaustive ROIs:



Localization: One Actual Lesion

- Combination of threshold for positive test and acceptance region for location of reader mark
  - Within mutually exclusive and exhaustive ROIs:

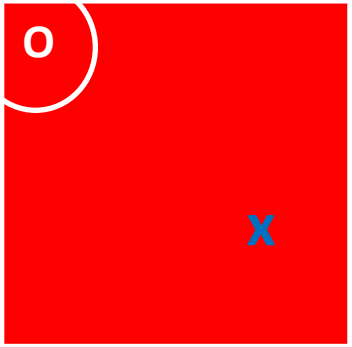


  - Credit for 1, 4, 16 correct
    - Increase statistical power

Localization: One Actual Lesion

- Within mutually exclusive and exhaustive ROIs: statistical power increases with the number of ROIs
- With an acceptance radius, we still have a single unit of statistical information

Credit for 1 correct

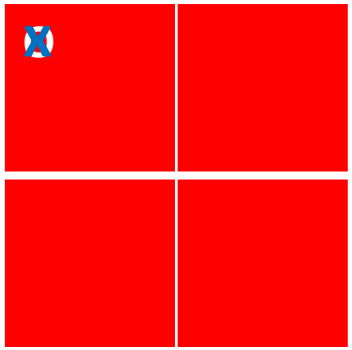


Ding for 1 miss

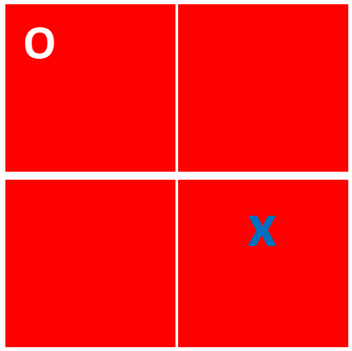
Localization: One Actual Lesion

Credit for 1 correct

Ding for 1 miss

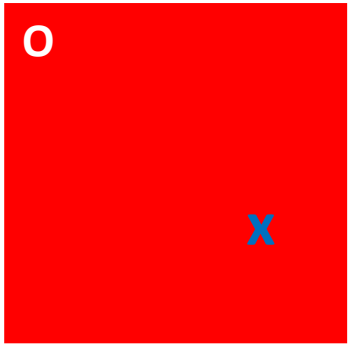


Credit for 4 correct



Credit for 2 correct; Dings for 1 miss and 1 FP

Localization: One Actual Lesion



- How should this be scored?
  - Often, FN (ding for a miss, no credit)
  - Partial credit for making a mark?

Ding for 1 miss

Localization: One or More Lesions

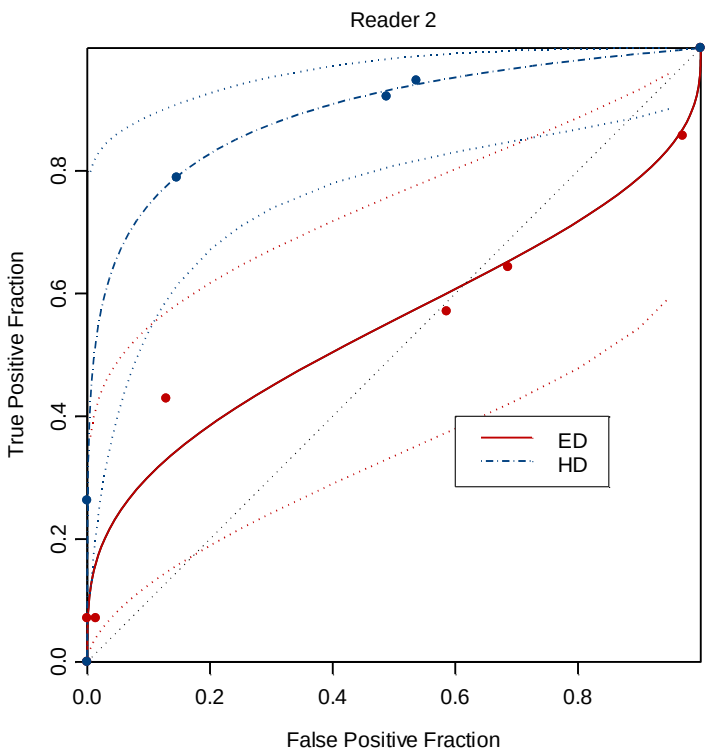
- ROI approach
  - Complexity: lesions and/or reader marks that span borders
  - Reader burden
- Acceptance radius approach
  - FROC
  - Complexity: summary statistic
    - AFROC: TPF with lesions as denominator, FPF with cases (without lesions) as denominator

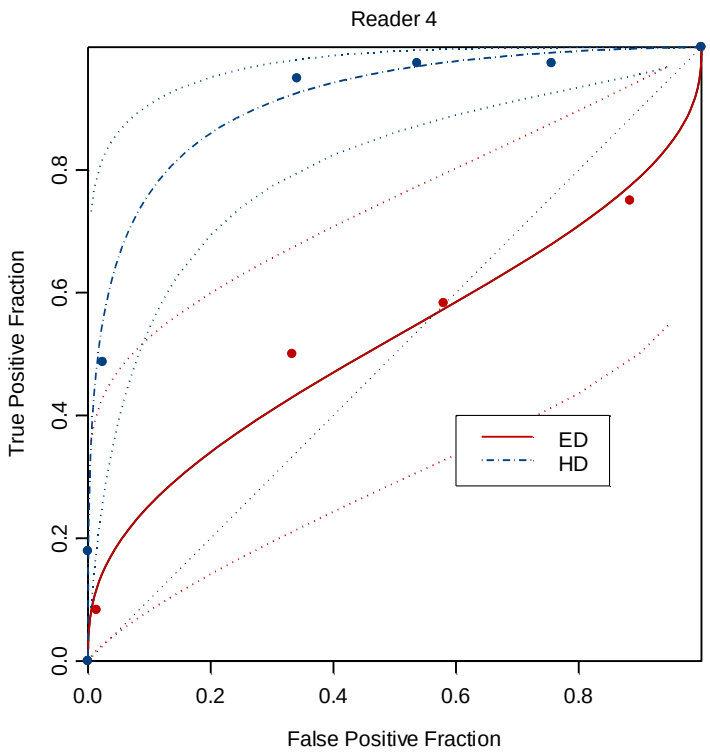
Above the Diagonal

"Also, if the test provides information to the decision maker, the intermediate points on a conventional ROC curve must be above the major (lower left to upper right) diagonal of the ROC space, because in that situation a positive decision should be more probable when a case is actually positive than when a case is actually negative, …"

Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

Above the Diagonal





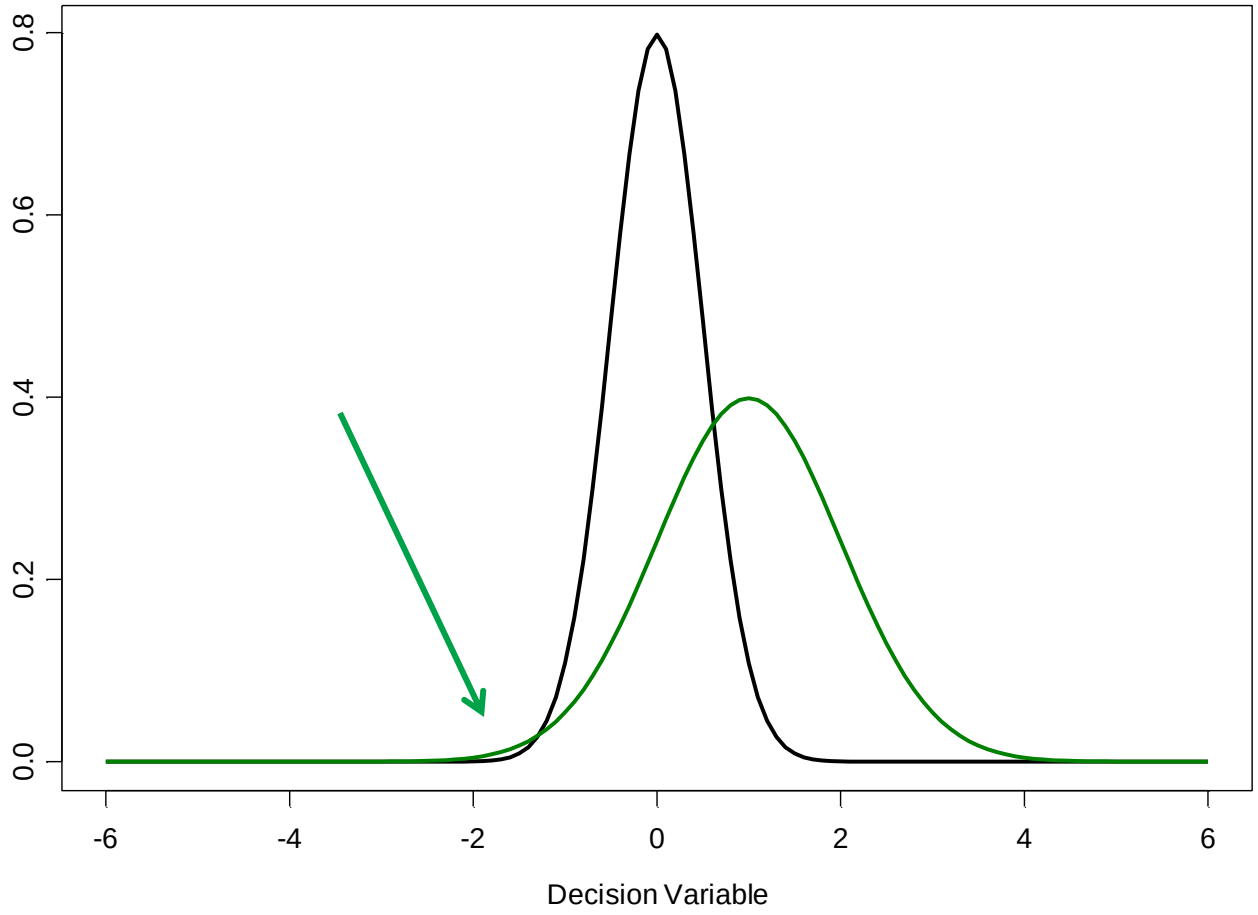
"Proper"

"…one can show on theoretical grounds that if the decision maker uses available information in a proper way, the slope of the ROC curve must steadily decrease (i.e., it must become less steep) as one moves up and to the right on the curve."

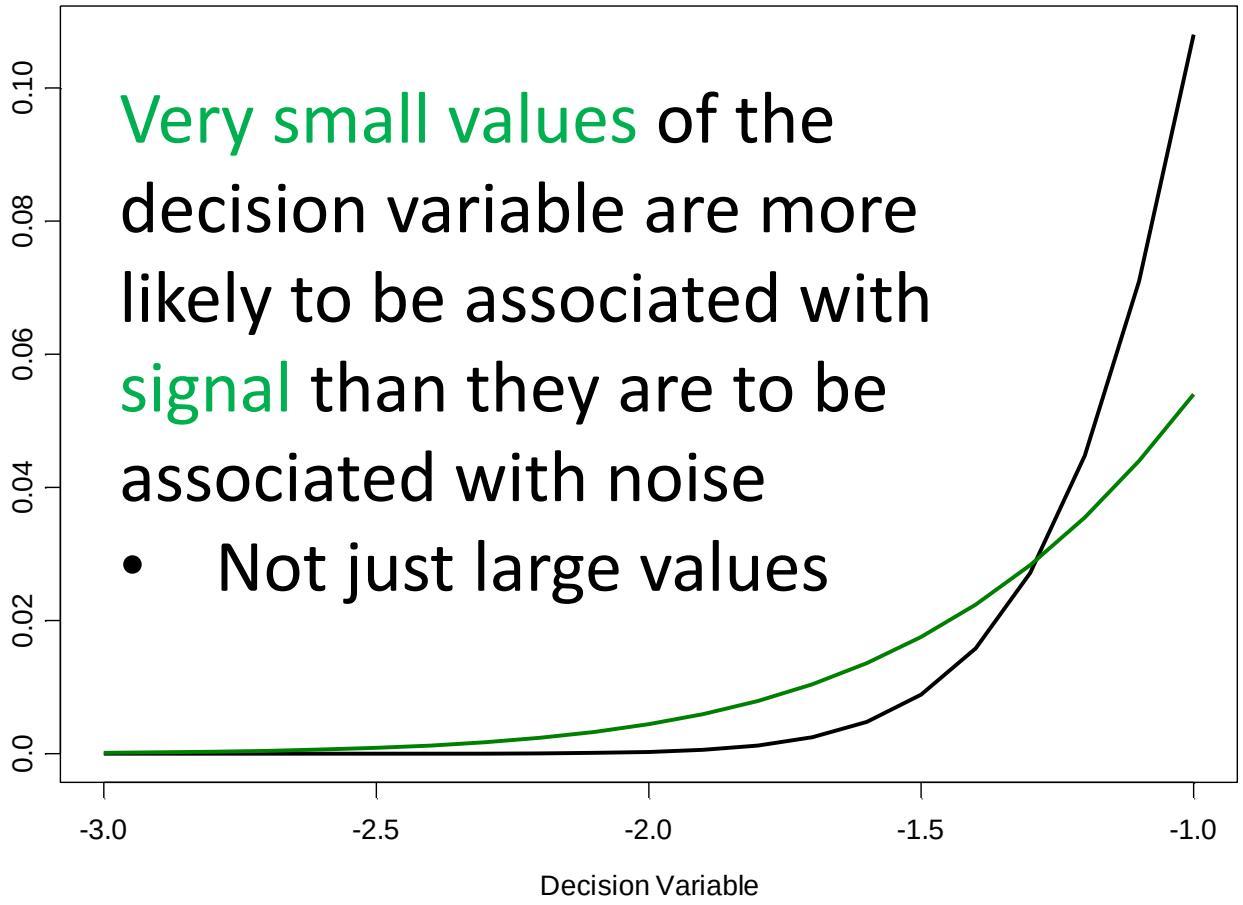
Metz, CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **8**:283-298 (1978).

- Any "hook" is an artifact of the binormal model
- "Proper" binormal model ensures monotonic slope, via relationship to the likelihood ratio

The Hook



The Hook


Very small values of the decision variable are more likely to be associated with signal than they are to be associated with noise
• Not just large values
0.0
0.02
0.04
0.06
0.08
0.10
-3.0
-2.5
-2.0
-1.5
-1.0
Decision Variable

“Proper” Binormal Model

- Likelihood ratio at x: ratio of heights of distributions, signal to noise, at x
- Invariance allows new decision variable:

$$y \equiv \ln\left(\frac{LR(x)}{b}\right) = -\frac{1}{2}\left[(bx-a)^2 - x^2\right]$$

  - Less stable numerically than conventional binormal model (“long, narrow, twisting ridge”)

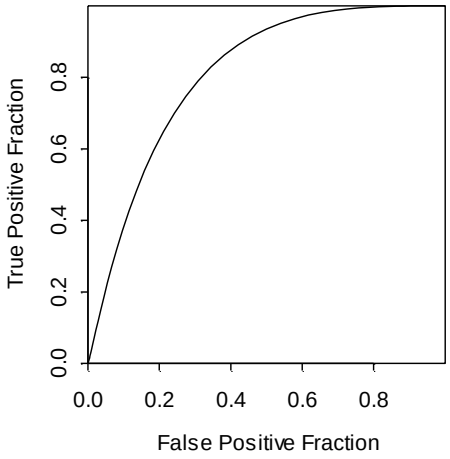
Metz, CE and Pan, X. “Proper” Binormal ROC Curves: …. *J Math Psychol* **43**:1-33 (1999).

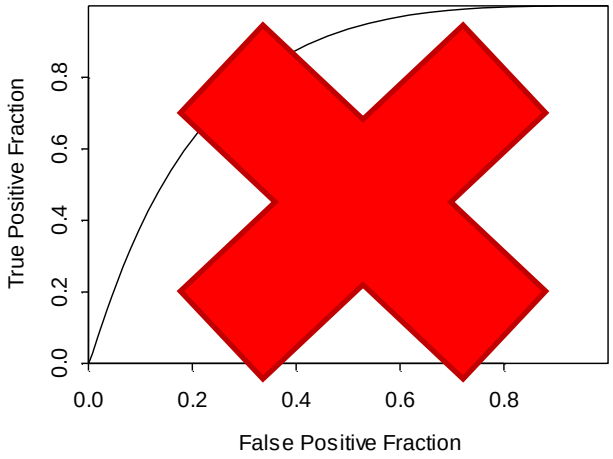
"Proper" Binormal Model

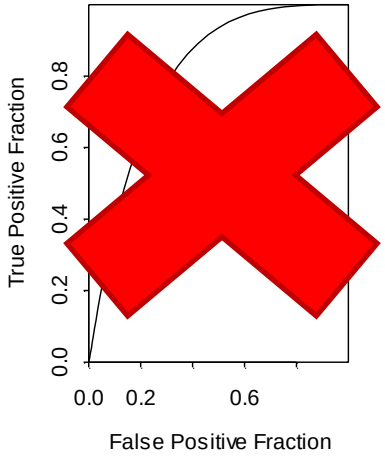
- New parameterization: (d'_e, c, v)
  - d'_e: difference between means relative to average of the SDs (rather than relative to SD_{with})
  - c: difference between the SDs relative to their sum (rather than ratio of the SDs)
  - v: monotonic transform. of y; function of (y, a, b)
- New algorithm to perform successfully in situations that caused problems earlier

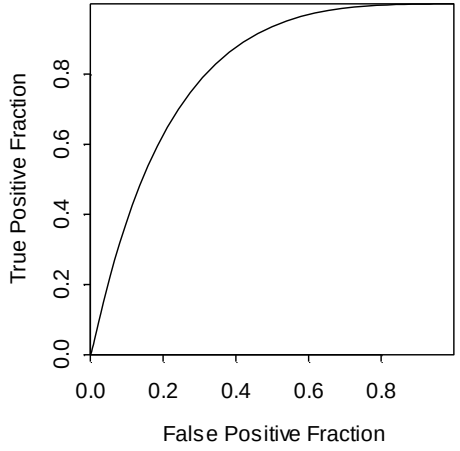
Metz, CE and Pan, X. "Proper" Binormal ROC Curves: …. *J Math Psychol* **43**:1-33 (1999).
Pesce, LL and Metz, CE. Reliable and Computationally Efficient … *Acad Radiol* **14**:814-829 (2007)

Unit Square

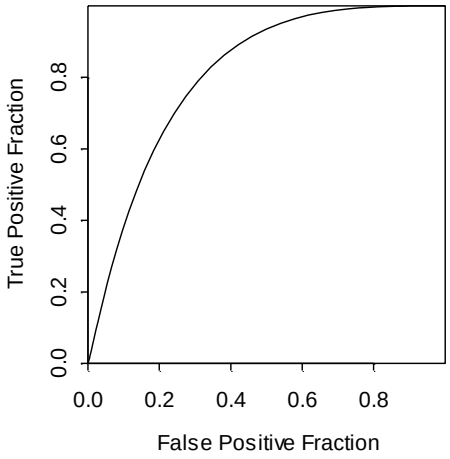

True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction

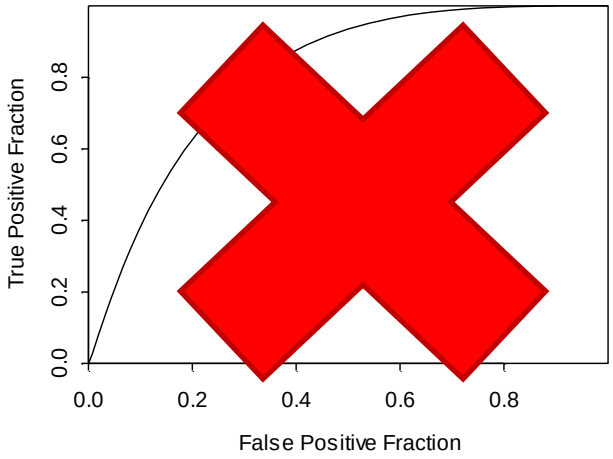

True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction

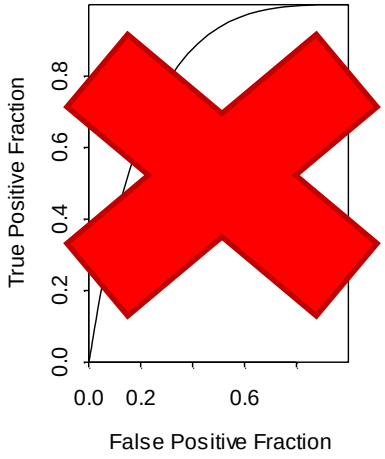

True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.6
False Positive Fraction

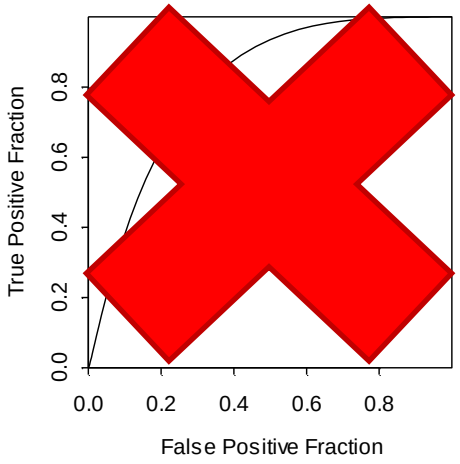

True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction

Unit Square


True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction


True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction


True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.6
False Positive Fraction


True Positive Fraction
0.0
0.2
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
False Positive Fraction

Human readers

- Not always logical
  - Don’t always use the test in an informative way
  - Not always necessarily “proper”
- Missing data: prevent if possible
  - Corrupt image files
  - Database issues
  - Flight schedules

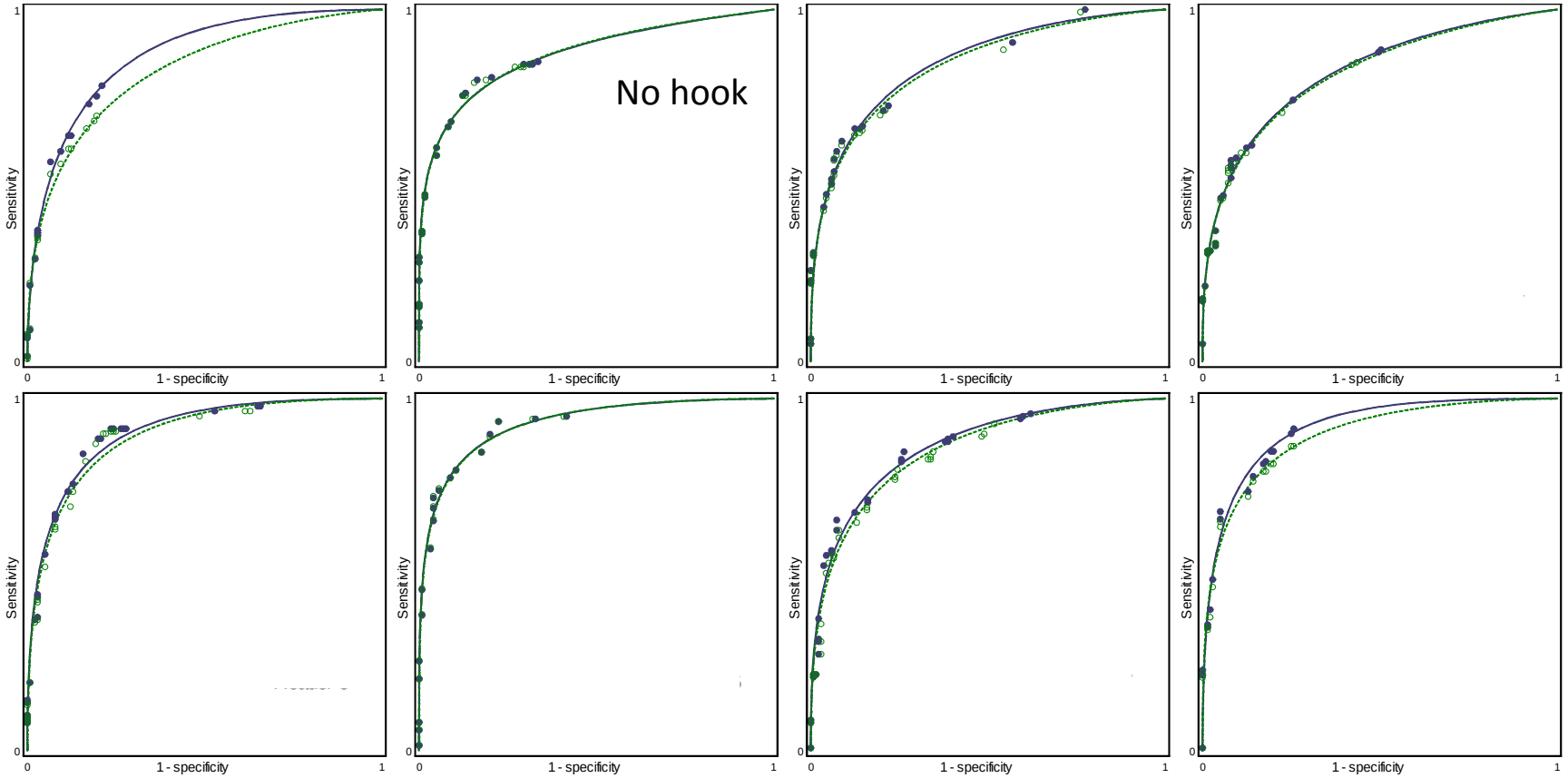
Planning MRMC Studies

- Consider case spectrum: representative of population of interest
  - Not a statistical calculation
  - Similarly for reader spectrum

✓ Pilot studies can be very helpful: practicalities

- ✓ Difference we want to detect realistic?
- ✓ Variability in ratings
- ✓ Between-reader variability in area under ROC curve, and different correlations
- × Meet with much resistance

Subset of Readers



Charles E. Metz

- "… the trickiest bit for me at the moment seems to be finding in my head an appropriate balance between my natural instinct to fight and a long-held understanding that no one ever born has failed to die."
- The better I got to know Charlie, the more I respected him, admired him, and genuinely liked him.
  - All the best, Charlie. Thanks for everything.