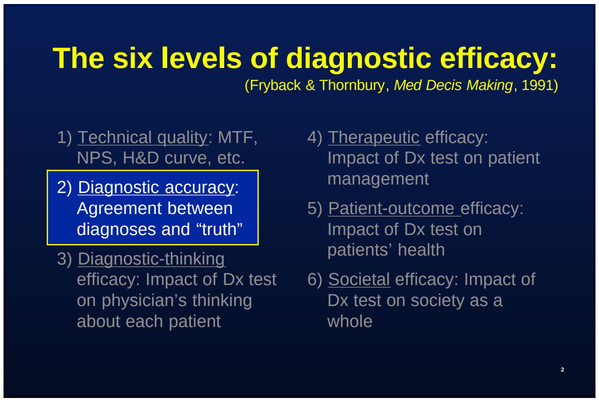
Practical Aspects of CAD Research

Assessment Methodologies for CAD

Charles E. Metz
Professor of Radiology
The University of Chicago



The six levels of diagnostic efficacy:

(Fryback & Thornbury, *Med Decis Making*, 1991)

1) Technical quality: MTF, NPS, H&D curve, etc.

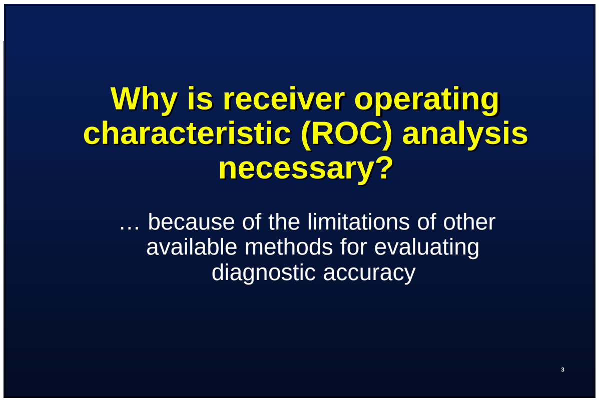
2) Diagnostic accuracy: Agreement between diagnoses and "truth"

3) Diagnostic-thinking efficacy: Impact of Dx test on physician's thinking about each patient

4) Therapeutic efficacy: Impact of Dx test on patient management

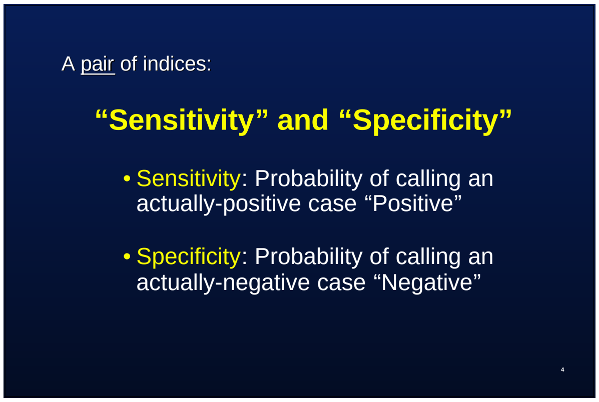
5) Patient-outcome efficacy: Impact of Dx test on patients' health

6) Societal efficacy: Impact of Dx test on society as a whole



Why is receiver operating characteristic (ROC) analysis necessary?

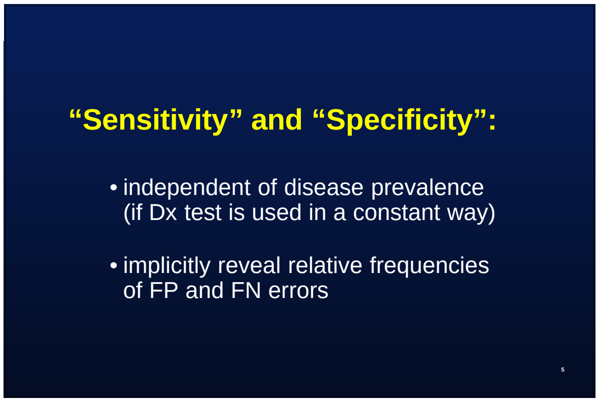
… because of the limitations of other available methods for evaluating diagnostic accuracy



A pair of indices:

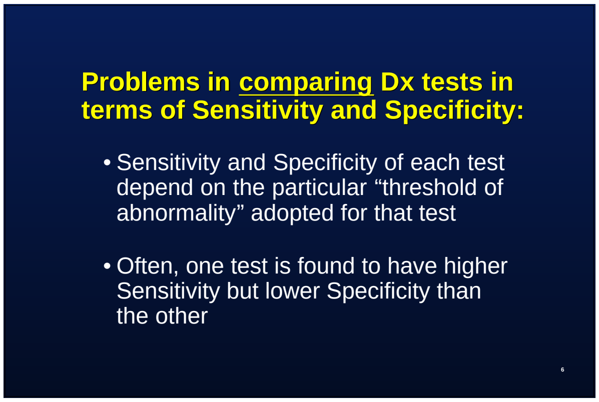
"Sensitivity" and "Specificity"

- Sensitivity: Probability of calling an actually-positive case "Positive"
- Specificity: Probability of calling an actually-negative case "Negative"



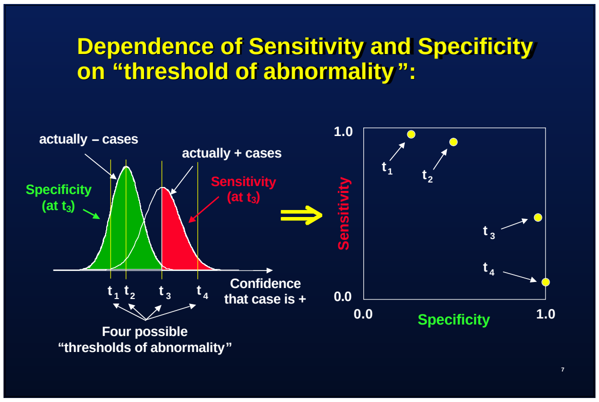
"Sensitivity" and "Specificity":

- independent of disease prevalence (if Dx test is used in a constant way)
- implicitly reveal relative frequencies of FP and FN errors

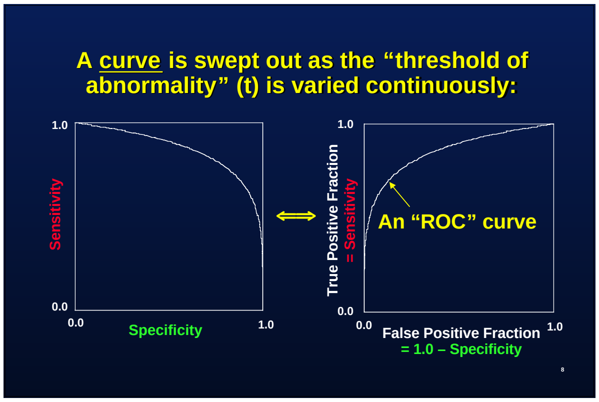


Problems in comparing Dx tests in terms of Sensitivity and Specificity:

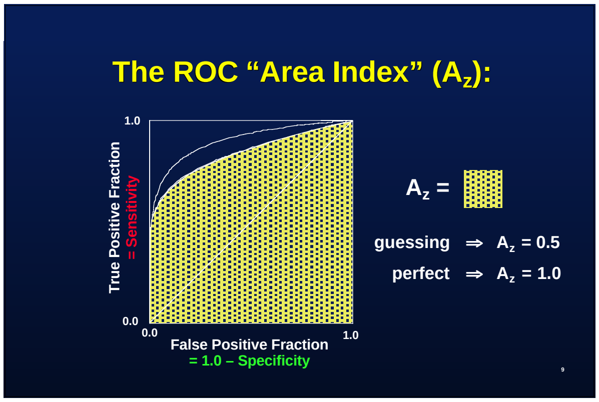
- Sensitivity and Specificity of each test depend on the particular "threshold of abnormality" adopted for that test
- Often, one test is found to have higher Sensitivity but lower Specificity than the other



Dependence of Sensitivity and Specificity on "threshold of abnormality":



A curve is swept out as the "threshold of abnormality" (t) is varied continuously:



The ROC "Area Index" (A_z):

A_z = ▨

guessing ⇒ $A_z = 0.5$

perfect ⇒ $A_z = 1.0$

Interpretations of ROC area (A_z):

- Sensitivity (TPF) averaged over all Specificities (or FPFs) — i.e., average ROC curve height
- Specificity averaged over all Sensitivities
- Probability of distinguishing correctly between a randomly selected actually-positive case and a randomly selected actually-negative case

However ...

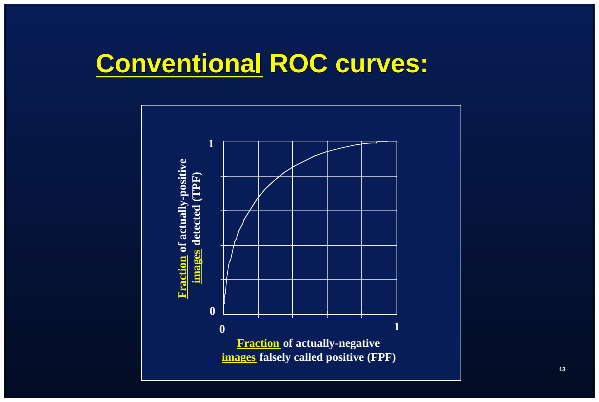
- this global index can be misleading when curves cross and/or there is only one region of interest

Other ROC-based indices of performance

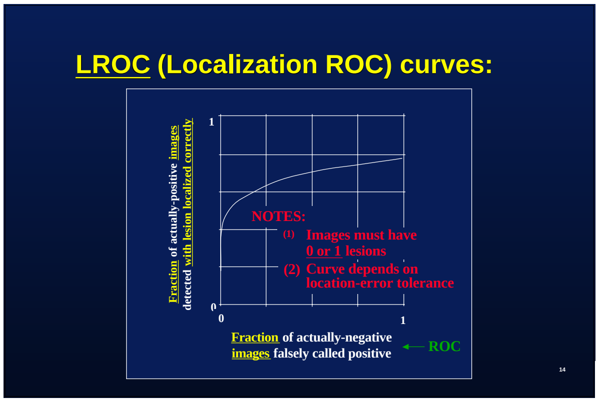
- Partial area below, or to the right of, a segment of the ROC curve (regional)
- TPF at fixed FPF or vice-versa (local)
- Expected utility at optimal operating point (local) — most meaningful but least practical

Generalized ROC analysis :

- Localization ROC (LROC) analysis
- Free-response ROC (FROC) analysis
- Alternative FROC (AFROC) analysis

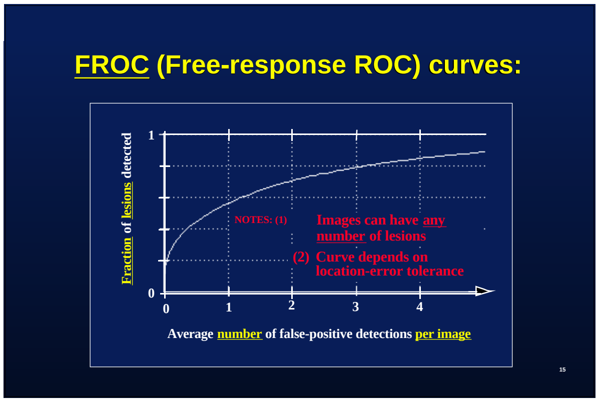


Conventional ROC curves:



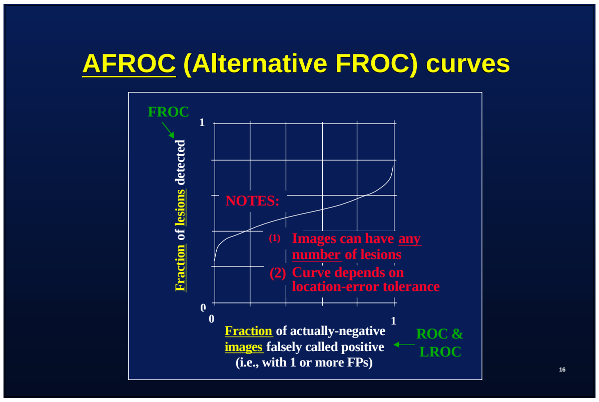
LROC (Localization ROC) curves:

NOTES:
(1) Images must have 0 or 1 lesions
(2) Curve depends on location-error tolerance



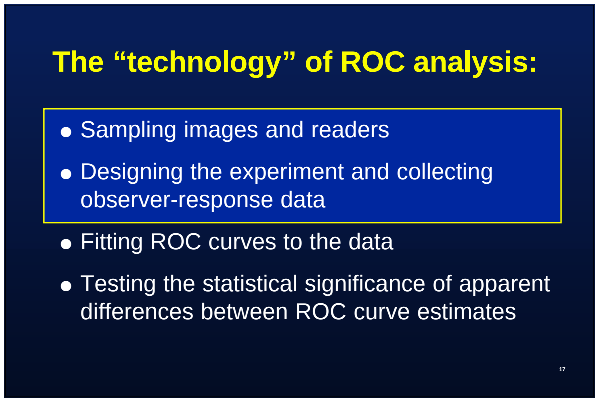
FROC (Free-response ROC) curves:

NOTES: (1) Images can have any number of lesions
(2) Curve depends on location-error tolerance



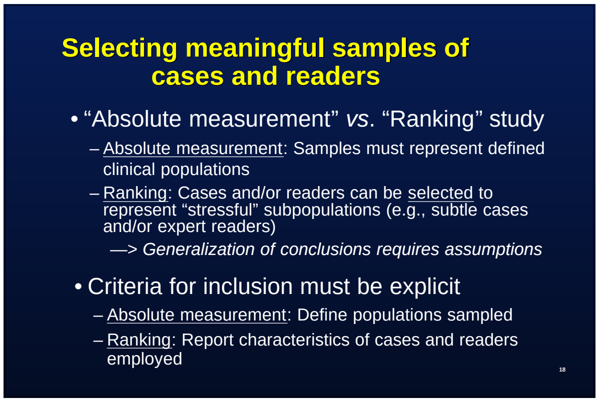
AFROC (Alternative FROC) curves

NOTES:
(1) Images can have any number of lesions
(2) Curve depends on location-error tolerance



The "technology" of ROC analysis:

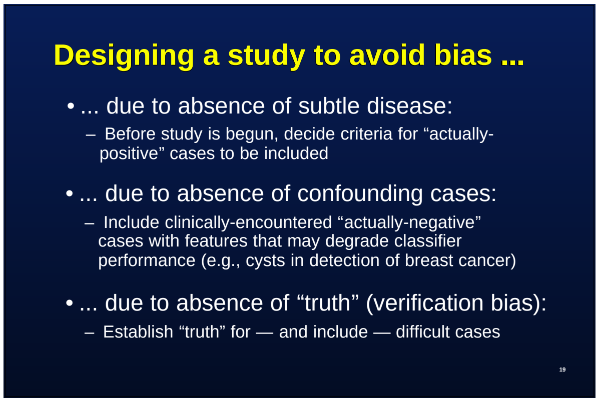
- Sampling images and readers
- Designing the experiment and collecting observer-response data
- Fitting ROC curves to the data
- Testing the statistical significance of apparent differences between ROC curve estimates



Selecting meaningful samples of cases and readers

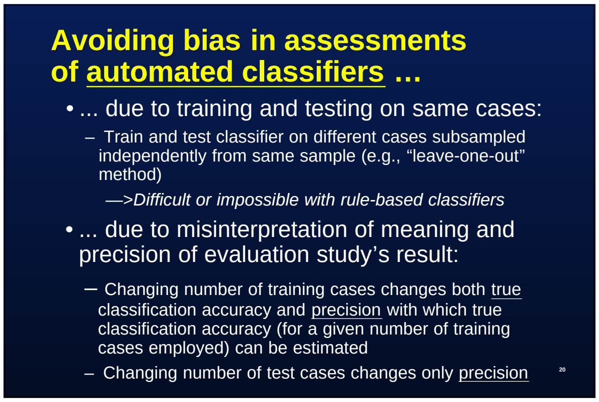
- "Absolute measurement" *vs*. "Ranking" study
  - Absolute measurement: Samples must represent defined clinical populations
  - Ranking: Cases and/or readers can be selected to represent "stressful" subpopulations (e.g., subtle cases and/or expert readers)

    —> Generalization of conclusions requires assumptions
- Criteria for inclusion must be explicit
  - Absolute measurement: Define populations sampled
  - Ranking: Report characteristics of cases and readers employed



Designing a study to avoid bias ...

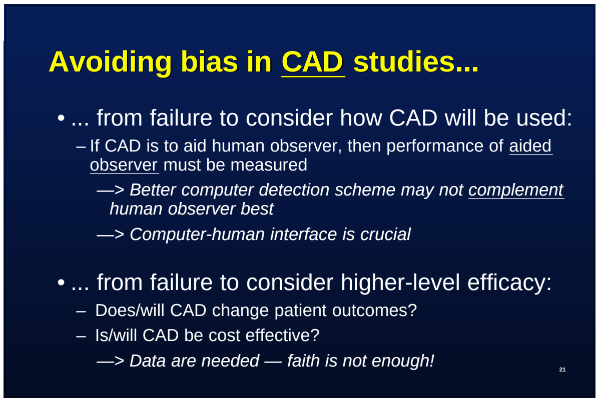
- ... due to absence of subtle disease:
  - Before study is begun, decide criteria for “actually-positive” cases to be included
- ... due to absence of confounding cases:
  - Include clinically-encountered “actually-negative” cases with features that may degrade classifier performance (e.g., cysts in detection of breast cancer)
- ... due to absence of “truth” (verification bias):
  - Establish “truth” for — and include — difficult cases



Avoiding bias in assessments of automated classifiers …

- ... due to training and testing on same cases:
  - Train and test classifier on different cases subsampled independently from same sample (e.g., “leave-one-out” method)

    —>Difficult or impossible with rule-based classifiers
- ... due to misinterpretation of meaning and precision of evaluation study’s result:
  - Changing number of training cases changes both true classification accuracy and precision with which true classification accuracy (for a given number of training cases employed) can be estimated
  - Changing number of test cases changes only precision



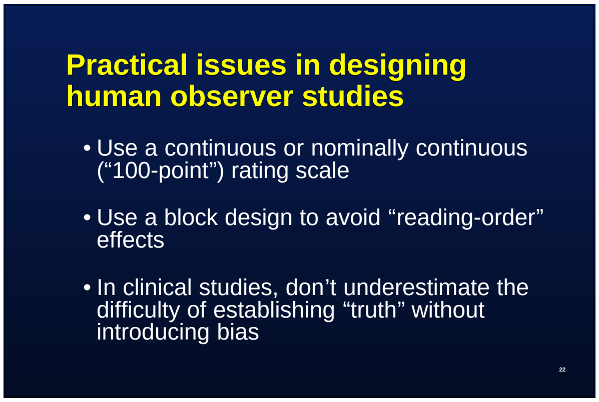
Avoiding bias in CAD studies...

- ... from failure to consider how CAD will be used:
  - If CAD is to aid human observer, then performance of aided observer must be measured

    —> Better computer detection scheme may not complement human observer best

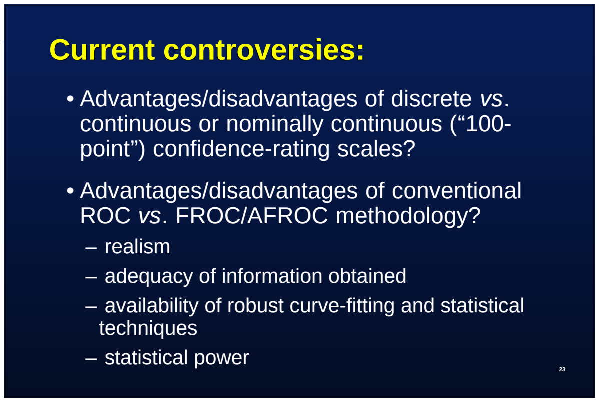
    —> Computer-human interface is crucial
- ... from failure to consider higher-level efficacy:
  - Does/will CAD change patient outcomes?
  - Is/will CAD be cost effective?

    —> Data are needed — faith is not enough!



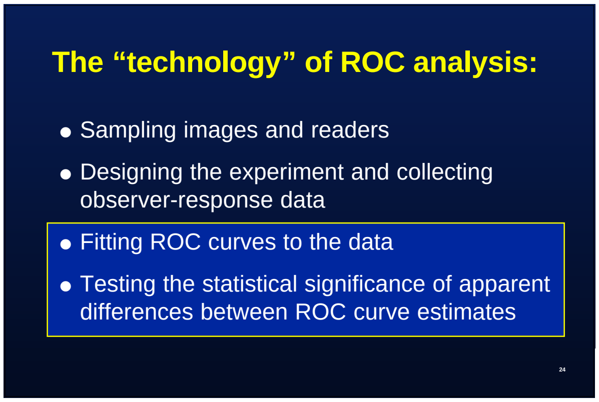
Practical issues in designing human observer studies

- Use a continuous or nominally continuous (“100-point”) rating scale
- Use a block design to avoid “reading-order” effects
- In clinical studies, don’t underestimate the difficulty of establishing “truth” without introducing bias



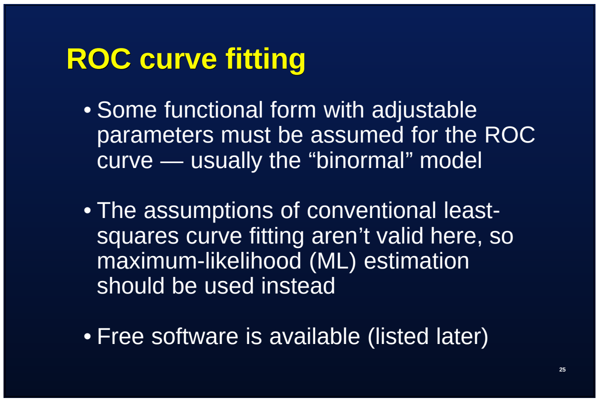
Current controversies:

- Advantages/disadvantages of discrete *vs*. continuous or nominally continuous (“100-point”) confidence-rating scales?
- Advantages/disadvantages of conventional ROC *vs*. FROC/AFROC methodology?
  - realism
  - adequacy of information obtained
  - availability of robust curve-fitting and statistical techniques
  - statistical power



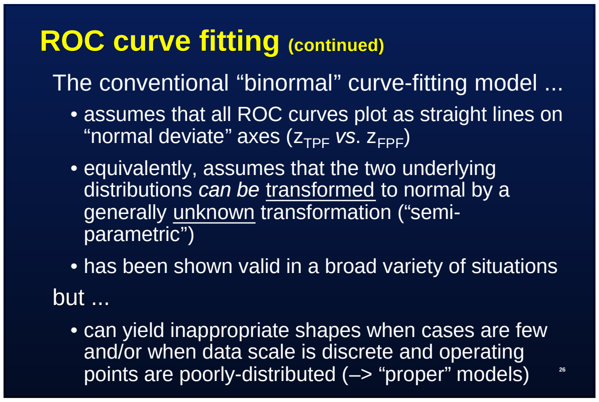
The “technology” of ROC analysis:

- Sampling images and readers
- Designing the experiment and collecting observer-response data
- Fitting ROC curves to the data
- Testing the statistical significance of apparent differences between ROC curve estimates



ROC curve fitting

- Some functional form with adjustable parameters must be assumed for the ROC curve — usually the “binormal” model
- The assumptions of conventional least-squares curve fitting aren’t valid here, so maximum-likelihood (ML) estimation should be used instead
- Free software is available (listed later)



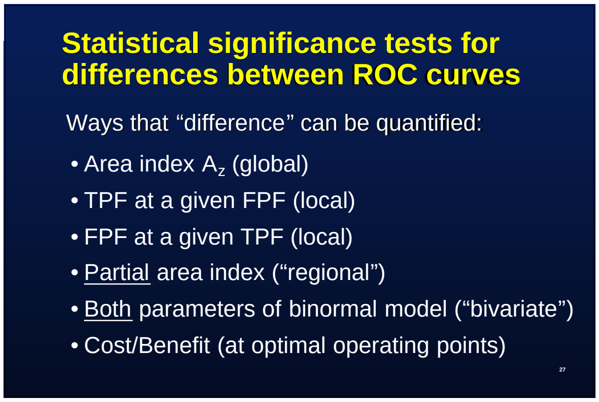
ROC curve fitting (continued)

The conventional “binormal” curve-fitting model ...

- assumes that all ROC curves plot as straight lines on “normal deviate” axes (z_{TPF} *vs.* z_{FPF})
- equivalently, assumes that the two underlying distributions *can be* transformed to normal by a generally unknown transformation (“semi-parametric”)
- has been shown valid in a broad variety of situations

but ...

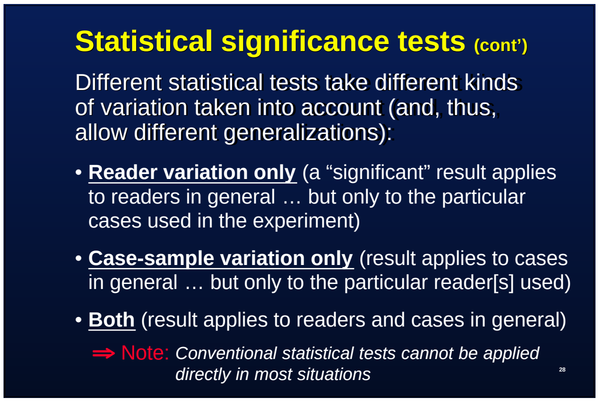
- can yield inappropriate shapes when cases are few and/or when data scale is discrete and operating points are poorly-distributed (–> “proper” models)



Statistical significance tests for differences between ROC curves

Ways that “difference” can be quantified:

- Area index A_z (global)
- TPF at a given FPF (local)
- FPF at a given TPF (local)
- Partial area index (“regional”)
- Both parameters of binormal model (“bivariate”)
- Cost/Benefit (at optimal operating points)

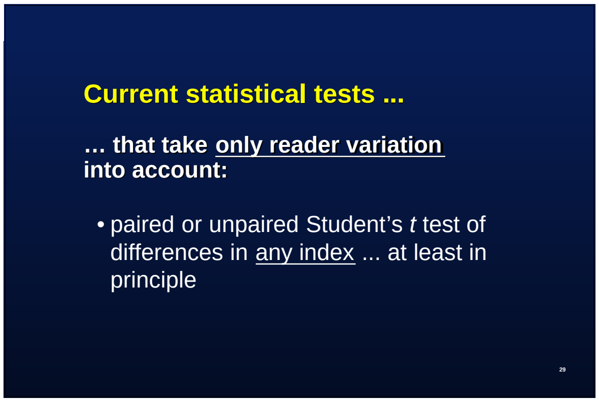


Statistical significance tests (cont’)

Different statistical tests take different kinds of variation taken into account (and, thus, allow different generalizations):

- **Reader variation only** (a “significant” result applies to readers in general … but only to the particular cases used in the experiment)
- **Case-sample variation only** (result applies to cases in general … but only to the particular reader[s] used)
- **Both** (result applies to readers and cases in general)

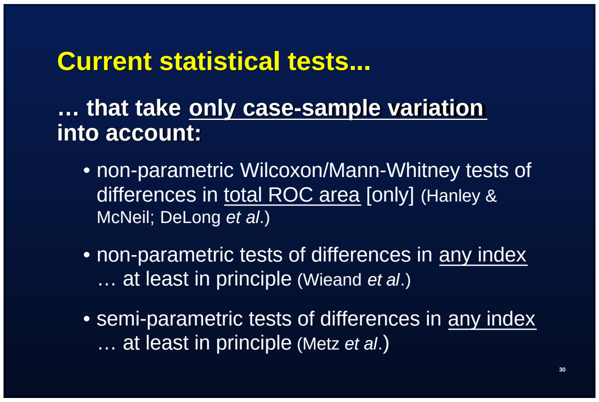
⇒ Note: *Conventional statistical tests cannot be applied directly in most situations*



Current statistical tests ...

… that take only reader variation into account:

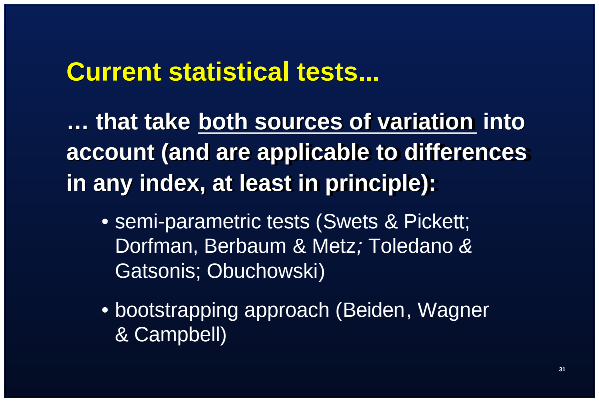
- paired or unpaired Student’s *t* test of differences in any index ... at least in principle



Current statistical tests...

… that take only case-sample variation into account:

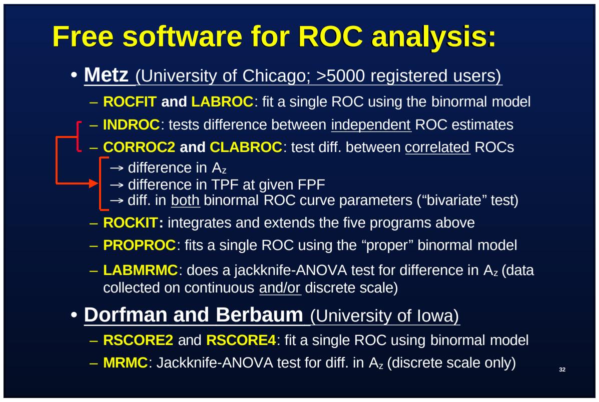
- non-parametric Wilcoxon/Mann-Whitney tests of differences in total ROC area [only] (Hanley & McNeil; DeLong *et al.*)
- non-parametric tests of differences in any index … at least in principle (Wieand *et al.*)
- semi-parametric tests of differences in any index … at least in principle (Metz *et al.*)



Current statistical tests...

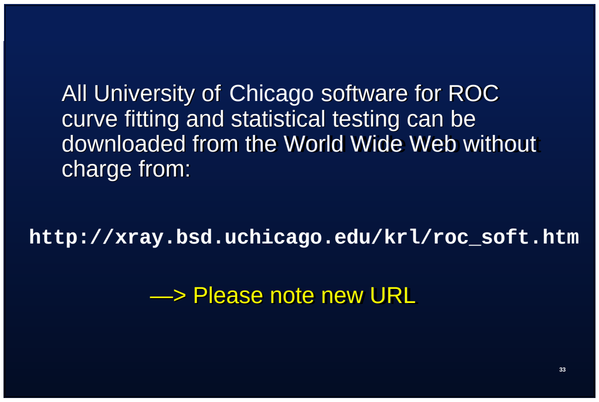
... that take both sources of variation into account (and are applicable to differences in any index, at least in principle):

- semi-parametric tests (Swets & Pickett; Dorfman, Berbaum & Metz; Toledano & Gatsonis; Obuchowski)
- bootstrapping approach (Beiden, Wagner & Campbell)



Free software for ROC analysis:

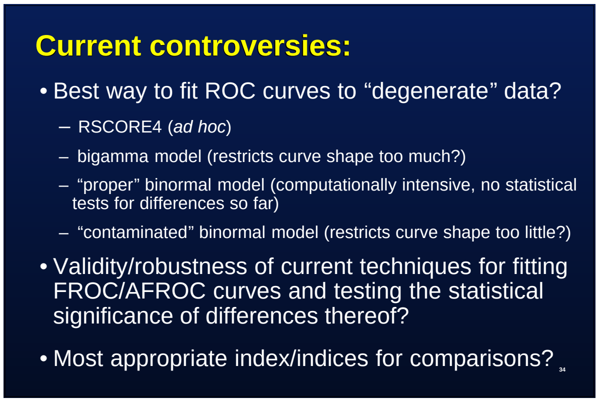
- **Metz** (University of Chicago; >5000 registered users)
  - – **ROCFIT and LABROC**: fit a single ROC using the binormal model
  - – **INDROC**: tests difference between independent ROC estimates
  - – **CORROC2 and CLABROC**: test diff. between correlated ROCs
    - → difference in A_z
    - → difference in TPF at given FPF
    - → diff. in both binormal ROC curve parameters (“bivariate” test)
  - – **ROCKIT:** integrates and extends the five programs above
  - – **PROPROC**: fits a single ROC using the “proper” binormal model
  - – **LABMRMC**: does a jackknife-ANOVA test for difference in A_z (data collected on continuous and/or discrete scale)
- **Dorfman and Berbaum** (University of Iowa)
  - – **RSCORE2** and **RSCORE4**: fit a single ROC using binormal model
  - – **MRMC**: Jackknife-ANOVA test for diff. in A_z (discrete scale only)



All University of Chicago software for ROC curve fitting and statistical testing can be downloaded from the World Wide Web without charge from:

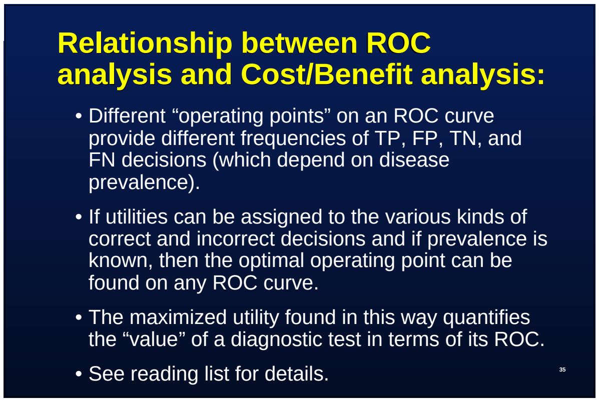
`http://xray.bsd.uchicago.edu/krl/roc_soft.htm`

—> Please note new URL



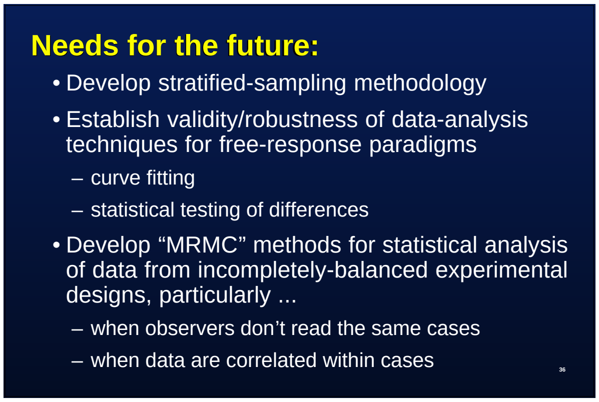
Current controversies:

- Best way to fit ROC curves to “degenerate” data?
  - – RSCORE4 (*ad hoc*)
  - – bigamma model (restricts curve shape too much?)
  - – “proper” binormal model (computationally intensive, no statistical tests for differences so far)
  - – “contaminated” binormal model (restricts curve shape too little?)
- Validity/robustness of current techniques for fitting FROC/AFROC curves and testing the statistical significance of differences thereof?
- Most appropriate index/indices for comparisons?



Relationship between ROC analysis and Cost/Benefit analysis:

- Different “operating points” on an ROC curve provide different frequencies of TP, FP, TN, and FN decisions (which depend on disease prevalence).
- If utilities can be assigned to the various kinds of correct and incorrect decisions and if prevalence is known, then the optimal operating point can be found on any ROC curve.
- The maximized utility found in this way quantifies the “value” of a diagnostic test in terms of its ROC.
- See reading list for details.



Needs for the future:

- Develop stratified-sampling methodology
- Establish validity/robustness of data-analysis techniques for free-response paradigms
  - – curve fitting
  - – statistical testing of differences
- Develop “MRMC” methods for statistical analysis of data from incompletely-balanced experimental designs, particularly ...
  - – when observers don’t read the same cases
  - – when data are correlated within cases

Needs for the future (continued):

- Develop highly efficient approaches well-suited to exploratory analyses
  - Key need is to control for decision-threshold effects
  - Other biases may be acceptable if sufficiently small
- Generalize ROC analysis to handle >2 decision alternatives
  - Must provide an appropriate compromise between complexity and practicality
  - Approaches proposed to date are *not* adequate

An incomplete list of recommended literature on ROC methodology

- BACKGROUND:
- Egan JP. Signal detection theory and ROC analysis. New York: Academic Press, 1975.
- Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991; 11: 88.
- Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures: principles and applications. Annals Int Med 1981; 94: 553.
- International Commission on Radiation Units and Measurements. Medical imaging: the assessment of image quality (ICRU Report 54). Bethesda,MD: ICRU, 1996.
- Lusted LB. Signal detectability and medical decision-making. Science 1971; 171: 1217.
- McNeil BJ, Adelstein SJ. Determining the value of diagnostic and screening tests. J Nucl Med 1976; 17: 439.
- Metz CE, Wagner RF, Doi K, Brown DG, Nishikawa RN, Myers KJ. Toward consensus on quantitative assessment of medical imaging systems. Med Phys 22: 1057-1061, 1995.
- National Council on Radiation Protection and Measurements. An introduction to efficacy in diagnostic radiology and nuclear medicine (NCRP Commentary 13). Bethesda, MD: NCRP, 1995.
- Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clinical Chemistry 1993; 39: 561. [Erratum published in Clinical Chemistry 1993; 39: 1589.]

- GENERAL:
- Hanley JA. Alternative approaches to receiver operating characteristic analysis. Radiology 1988; 168: 568.
- Hanley JA. Receiver operating characteristic (ROC) methodology: the state of the art. Critical Reviews in Diagnostic Imaging 1989; 29: 307.
- King JL, Britton CA, Gur D, Rockette HE, Davis PL. On the validity of the continuous and discrete confidence rating scales in receiver operating characteristic studies. Invest Radiol 1993; 28: 962.
- Metz CE. Basic principles of ROC analysis. Seminars in Nucl Med 1978; 8: 283.
- Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986; 21: 720.
- Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989; 24: 234.
- Metz CE. Evaluation of CAD methods. In Computer-Aided Diagnosis in Medical Imaging (K Doi, H MacMahon, ML Giger and KR Hoffmann, eds.). Amsterdam: Elsevier Science (Excerpta Medica International Congress Series, Vol. 1182), pp. 543-554, 1999.
- Metz CE. Fundamental ROC analysis. In: Handbook of Medical Imaging, Vol. 1: Physics and Psychophysics (J Beutel, H Kundel and R Van Metter, eds.). Bellingham, WA; SPIE Press, 2000, pp. 751-769.
- Metz CE, Shen J-H. Gains in accuracy from replicated readings of diagnostic images: prediction and assessment in terms of ROC analysis. Med Decis Making 1992; 12: 60.
- Rockette HE, Gur D, Metz CE. The use of continuous and discrete confidence judgments in receiver operating characteristic studies of diagnostic imaging techniques. Invest Radiol 1992; 27: 169.

- Swets JA. ROC analysis applied to the evaluation of medical imaging techniques. Invest Radiol 1979; 14: 109.
- Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. Psychol Bull 1986; 99: 100.
- Swets JA. Measuring the accuracy of diagnostic systems. Science 1988; 240: 1285.
- Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Mahwah, NJ; Lawrence Erlbaum Associates, 1996.
- Swets JA, Pickett RM. Evaluation of diagnostic systems: methods from signal detection theory. New York: Academic Press, 1982
- Wagner RF, Beiden SV, Metz CE. Continuous vs. categorical data for ROC analysis: Some quantitative considerations. Academic Radiol 2001, 8: 328, 2001.
- Wagner RF, Beiden SV, Campbell G, Metz CE, Sachs WM. Assessment of medical imaging and computer-assist systems: lessons from recent experience. Academic Radiol 2002; 8: 1264.

BIAS:

- Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983; 39: 207.
- Begg CB, McNeil BJ. Assessment of radiologic tests: control of bias and other design considerations. Radiology 1988; 167: 565.
- Gray R, Begg CB, Greenes RA. Construction of receiver operating characteristic curves when disease verification is subject to selection bias. Med Decis Making 1984; 4: 151.
- Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. New Engl J Med 1978; 299: 926.

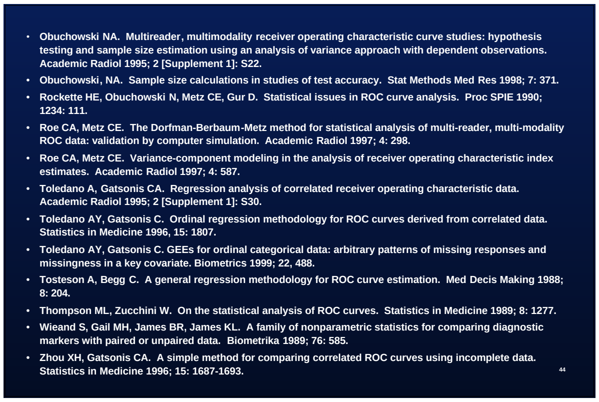
CURVE FITTING:

- Dorfman DD, Alf E. Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals — rating method data. J Math Psych 1969; 6: 487.
- Dorfman DD, Berbaum KS, Metz CE, Lenth RV, Hanley JA, Dagga HA. Proper ROC analysis: the bigamma model. Academic Radiol 1997; 4: 138.
- Grey DR, Morgan BJT. Some aspects of ROC curve-fitting: normal and logistic models. J Math Psych 1972; 9: 128.
- Hanley JA. The robustness of the "binormal" assumptions used in fitting ROC curves. Med Decis Making 1988; 8: 197.
- Metz CE, Herman BA, Shen J-H. Maximum-likelihood estimation of ROC curves from continuously-distributed data. Stat Med 1998; 17: 1033.
- Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. J Math Psych 1999; 43: 1.
- Pan X, Metz CE. The "proper" binormal model: parametric ROC curve estimation with degenerate data. Academic Radiol 1997; 4: 380.
- Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. Med Phys 1996; 23: 1709.
- Swets JA. Form of empirical ROCs in discrimination and diagnostic tasks: implications for theory and measurement of performance. Psychol Bull 1986; 99: 181.

STATISTICS:

- Agresti A. A survey of models for repeated ordered categorical response data. Statistics in Medicine 1989; 8; 1209.
- Bamber D. The area above the ordinal dominance graph and the area below the receiver operating graph. J Math Psych 1975; 12: 387.
- Beiden SV, Wagner RF, Campbell G. Components-of-variance models and multiple-bootstrap experiments: and alternative method for random-effects, receiver operating characteristic analysis. Academic Radiol. 2000; 7: 341.
- Beiden SV, Wagner RF, Campbell G, Metz CE, Jiang Y. Components-of-variance models for random-effects ROC analysis: The case of unequal variance structures across modalities. Academic Radiol. 2001; 8: 605.
- Beiden SV, Wagner RF, Campbell G, Chan H-P. Analysis of uncertainties in estimates of components of variance in multivariate ROC analysis. Academic Radiol. 2001; 8: 616.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837.
- Dorfman DD, Berbaum KS, Metz CE. ROC rating analysis: generalization to the population of readers and cases with the jackknife method. Invest Radiol 1992; 27: 723.
- Dorfman DD, Berbaum KS, Lenth RV, Chen Y-F, Donaghy BA. Monte Carlo validation of a multireader method for receiver operating characteristic discrtet rating data: factorial experimental design. Academic Radiol 1998; 5: 591.
- Dorfman DD, Metz CE. Multi-reader multi-case ROC analysis: comments on Begg’s commentary. Academic Radiol 1995; 2 (Supplement 1): S76.

- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143: 29.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839.
- Jiang Y, Metz CE, Nishikawa RM. A receiver operating characterisitc partial area index for highly sensitive diagnostic tests. Radiology 1996; 201: 745.
- Ma G, Hall WJ. Confidence bands for receiver operating characteristic curves. Med Decis Making 1993; 13: 191.
- McClish DK. Analyzing a portion of the ROC curve. Med Decis Making 1989; 9: 190.
- McClish DK. Determining a range of false-positive rates for which ROC curves differ. Med Decis Making 1990; 10: 283.
- McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med Decis Making 1984; 4: 137.
- Metz CE. Statistical analysis of ROC data in evaluating diagnostic performance. In: Multiple regression analysis: applications in the health sciences (D Herbert and R Myers, eds.). New York: American Institute of Physics, 1986, pp. 365.
- Metz CE. Quantification of failure to demonstrate statistical significance: the usefulness of confidence intervals. Invest Radiol 1993; 28: 59.
- Metz CE, Herman BA, Roe CA. Statistical comparison of two ROC curve estimates obtained from partially-paired datasets.   Med Decis Making 1998; 18: 110.
- Metz CE, Kronman HB. Statistical significance tests for binormal ROC curves. J Math Psych 1980; 22: 218.
- Metz CE, Wang P-L, Kronman HB. A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Information processing in medical imaging (F Deconinck, ed.). The Hague: Nijhoff, 1984, p. 432.



- Obuchowski NA. Multireader, multimodality receiver operating characteristic curve studies: hypothesis testing and sample size estimation using an analysis of variance approach with dependent observations. Academic Radiol 1995; 2 [Supplement 1]: S22.
- Obuchowski, NA. Sample size calculations in studies of test accuracy. Stat Methods Med Res 1998; 7: 371.
- Rockette HE, Obuchowski N, Metz CE, Gur D. Statistical issues in ROC curve analysis. Proc SPIE 1990; 1234: 111.
- Roe CA, Metz CE. The Dorfman-Berbaum-Metz method for statistical analysis of multi-reader, multi-modality ROC data: validation by computer simulation. Academic Radiol 1997; 4: 298.
- Roe CA, Metz CE. Variance-component modeling in the analysis of receiver operating characteristic index estimates. Academic Radiol 1997; 4: 587.
- Toledano A, Gatsonis CA. Regression analysis of correlated receiver operating characteristic data. Academic Radiol 1995; 2 [Supplement 1]: S30.
- Toledano AY, Gatsonis C. Ordinal regression methodology for ROC curves derived from correlated data. Statistics in Medicine 1996, 15: 1807.
- Toledano AY, Gatsonis C. GEEs for ordinal categorical data: arbitrary patterns of missing responses and missingness in a key covariate. Biometrics 1999; 22, 488.
- Tosteson A, Begg C. A general regression methodology for ROC curve estimation. Med Decis Making 1988; 8: 204.
- Thompson ML, Zucchini W. On the statistical analysis of ROC curves. Statistics in Medicine 1989; 8: 1277.
- Wieand S, Gail MH, James BR, James KL. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. Biometrika 1989; 76: 585.
- Zhou XH, Gatsonis CA. A simple method for comparing correlated ROC curves using incomplete data. Statistics in Medicine 1996; 15: 1687-1693.

RELATIONSHIPS WITH COST/BENEFIT ANALYSIS:

- Halpern EJ, Alpert M, Krieger AM, Metz CE, Maidment AD. Comparisons of ROC curves on the basis of optimal operating points. Academic Radiology 1996; 3: 245-253.
- Metz CE. Basic principles of ROC analysis. Seminars in Nucl Med 1978; 8: 283-298.
- Metz CE, Starr SJ, Lusted LB, Rossmann K. Progress in evaluation of human observer visual detection performance using the ROC curve approach. In: Information Processing in Scintigraphy (C Raynaud and AE Todd-Pokropek, eds.). Orsay, France: Commissariat à l'Energie Atomique, Département de Biologie, Service Hospitalier Frédéric Joliot, 1975, p. 420.
- Phelps CE, Mushlin AI. Focusing technology assessment. Med Decis Making 1988; 8: 279.
- Sainfort F. Evaluation of medical technologies: a generalized ROC analysis. Med Decis Making 1991; 11: 208.

- GENERALIZATIONS:

- Anastasio MA, Kupinski MA, Nishikawa RN. Optimization and FROC analysis of rule-based detection schemes using a multiobjective approach. IEEE Trans Med Imaging 1998; 17: 1089.
- Bunch PC, Hamilton JF, Sanderson GK, Simmons AH. A free response approach to the measurement and characterization of radiographic observer performance. Proc SPIE 1977; 127: 124.
- Chakraborty DP. Maximum likelihood analysis of free-response receiver operating characteristic (FROC) data. Med Phys 1989; 16: 561.
- Chakraborty DP, Winter LHL. Free-response methodology: alternate analysis and a new observer-performance experiment. Radiology 1990; 174: 873.
- Edwards DC, Kupinski MA, Metz CE, Nishikawa RN. Maximum-likelihood fitting of FROC curves under an initial-detection-and-candidate-analysis model. Med Phys 2002; 29: 2861.
- Egan JP, Greenberg GZ, Schulman AI. Operating characteristics, signal detection, and the method of free response. J Acoust Soc Am 1961; 33: 993.
- Metz CE, Starr SJ, Lusted LB. Observer performance in detecting multiple radiographic signals: prediction and analysis using a generalized ROC approach. Radiology 1976; 121: 337.
- Starr SJ, Metz CE, Lusted LB, Goodenough DJ. Visual detection and localization of radiographic images. Radiology 1975; 116: 533.
- Swensson RG. Unified measurement of observer performance in detecting and localizing target objects on images. Med Phys 1996; 23: 1709.