Assessing the accuracy of forensic analyses

An approach to defining and quantifying (some) types of error

Constantine Gatsonis

Center for Statistical Sciences

Brown University

ASU, 4 April 2009

Topics

- **Errors and error rates:**
  - **Variety of types and sources of error**
  - **Imprecise use of language and fuzzy thinking**
- **Evaluation of accuracy of forensic analyses**
  - **Define the task!**
  - **Define measures of accuracy.**
  - **Conduct experiments.**
  - **Monitor practice.**
- **Lessons from other areas of tech. assessment**

Define the task

- **Individualization:**
  - **Can a piece of evidence be associated with a particular source?**
- **Classification:**
  - **Can a piece of evidence be associated with a particular class of sources?**
- **A few modalities have potential for individualization.**
- **More of them have potential for classification.**

Keep evaluation focused on the task

- **Individualization:**
  - **Can a piece of evidence be associated with a particular source?**
- **Classification:**
  - **Can a piece of evidence be associated with a particular class of sources?**

- **Avoid "mission creep"**

Measuring accuracy

- **Borrowing from the paradigm of diagnostic testing**
- **The well known 2x2 table for dichotomous test and truth:**

	Forensic analysis results		
Truth	**“yes”**	**“no”**	***Total***
“yes” (Target condition present)	**True Positives**	**False Negatives**	N_+
“no” (Target condition absent)	**False Positives**	**True Negatives**	N_-
Total	***Test Positives***	***Test Negatives***	N

Objective: Detection

- **Sensitivity: Probability that analysis will find the target condition, when the target condition is present.**
- **Specificity: Probability that analysis will declare target condition is not there when target condition is absent.**

Measures of error: 1-sensitivity, 1- specificity

	Hair analysis results	
Truth	**Class C**	**Not Class C**
Hair comes from individual in class C	**TP**	**FN**
Hair comes from individual not in C	**FP**	**TN**

Errors!

Objective: Prediction

- **Positive Predictive Value : Probability target condition is actually present when analysis says it is.**
- **Negative Predictive Value: Probability target condition is absent when analysis says it is not there.**

Measures of error: 1-PPV, 1- NPV

	Hair analysis results	
Truth	**Class C**	**Not Class C**
Hair comes from individual in class C	**TP**	**FN**
Hair comes from individual not in C	**FP**	**TN**

Errors!

Approach also useful for individualization studies

Hypothetical fingerprint study:

A set of pairs of prints is analyzed

	Analysis results	
Truth	**match**	**No match**
Pair of prints comes from same individual	**TP**	**FN**
Pair of prints comes from different individuals	**FP**	**TN**

Errors!

Studies of accuracy

- **Measures of accuracy can be estimated via <u>designed studies.</u>**
- **Accuracy likely to be influenced by several factors, e.g.**
  - **"Difficulty" of cases ("case mix")**
  - **Experience and training of analysts**
  - **Contextually available information**
- **Ideally, we need to know**
  - **average accuracy (across analysts, laboratories etc)**
  - **range (variability) of accuracy (across analysts, laboratories etc)**

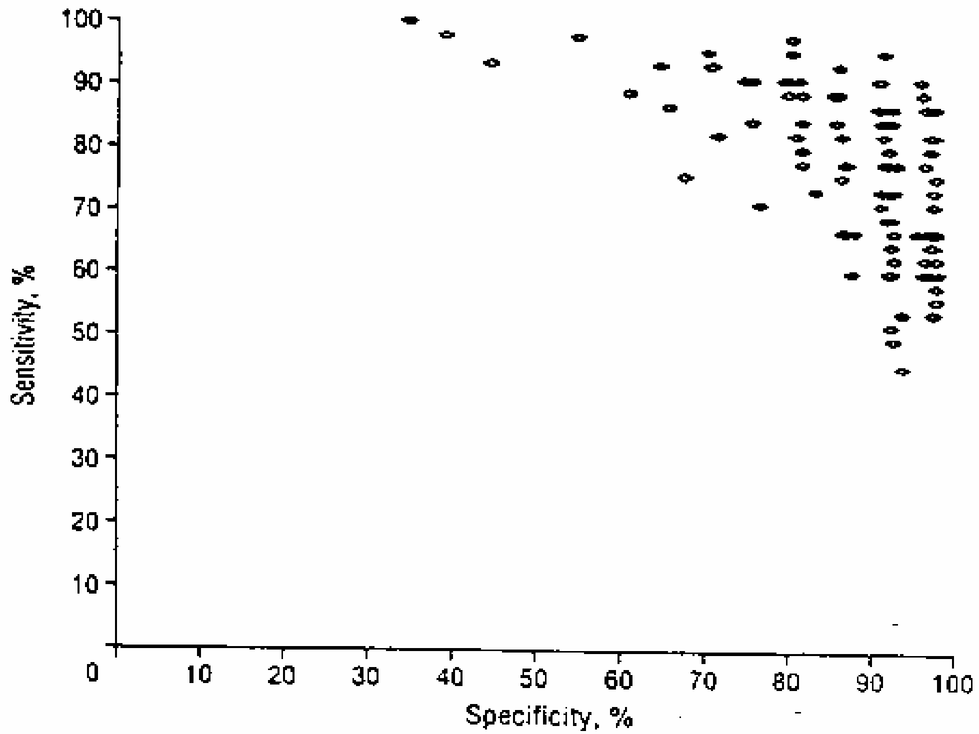
Using this accuracy paradigm

- This paradigm of accuracy assessment can be useful in many settings.
- It requires substantial research effort.
- It does not address important questions in individualization:
  - Definition of “match”
  - Estimation of random match probabilities
- Paradigm addresses performance over repeated instances of the analysis. It does not necessarily guarantee the correct answer in a specific case.

Experiences from diagnostic medicine

- **“Moving target problem”: Technology evolves, often quite rapidly.**
- **Modality performance vs reader performance**
- **Assessing/monitoring *effectiveness (*i.e. performance in everyday use) is major challenge.**
- **Do these seem familiar?**

Studies may highlight sobering realities



Performance of mammographers interpreting common set of scans. (Beam, Arch Int Med, 1996)

Expert analysts may not agree

Reader Agreement in Retrospective Interpretation of CT and MR Imaging Studies

	Multirater κ Value*		*P* Value†	
Parameter	CT	MR Imaging	CT	MR Imaging
Tumor visualization	0.16 (0.12 to 0.29)	0.32 (0.22 to 0.41)	<.001	<.001
Invasion of right parametrium	−0.04 (−0.02 to 0.13)	0.10 (0.06 to 0.27)	.961	<.001
Invasion of left parametrium	−0.05 (−0.01 to 0.11)	0.12 (0.05 to 0.29)	.981	<.001
Overall parametrial invasion‡	−0.04 (−0.02 to 0.13)	0.11 (0.05 to 0.29)	. . .	. . .
Staging§	0.26 (0.23 to 0.34)	0.44 (0.34 to 0.56)	<.001	<.001

Hricak, Gatsonis, et al Radiology 2007

High tech and new tech is not necessarily better

MRI and MRSI for localizing cancer in prostate

Radiology, on-line

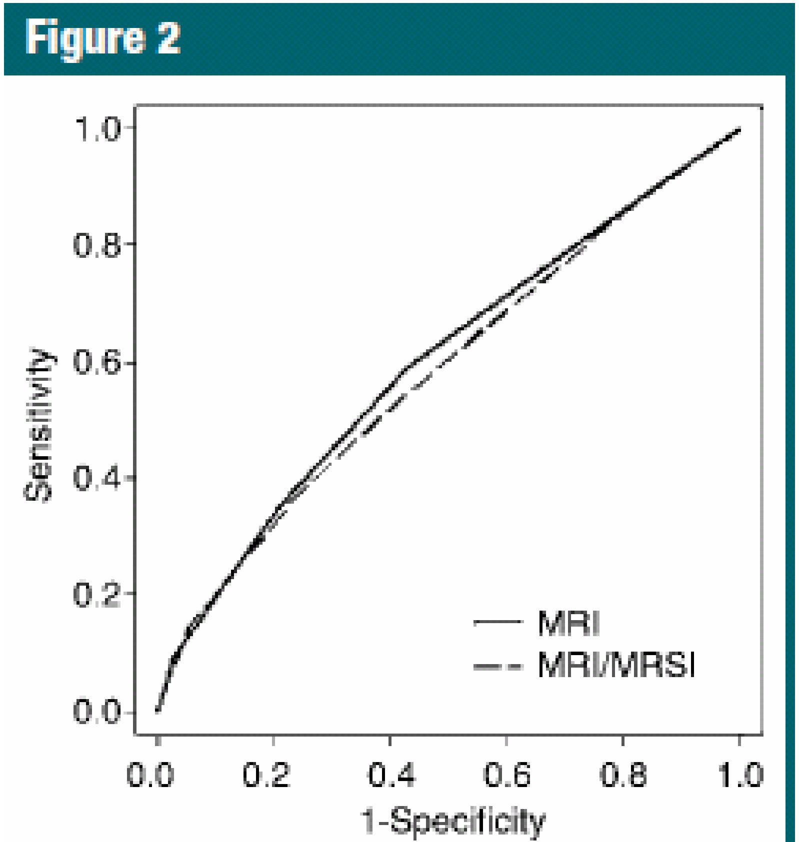
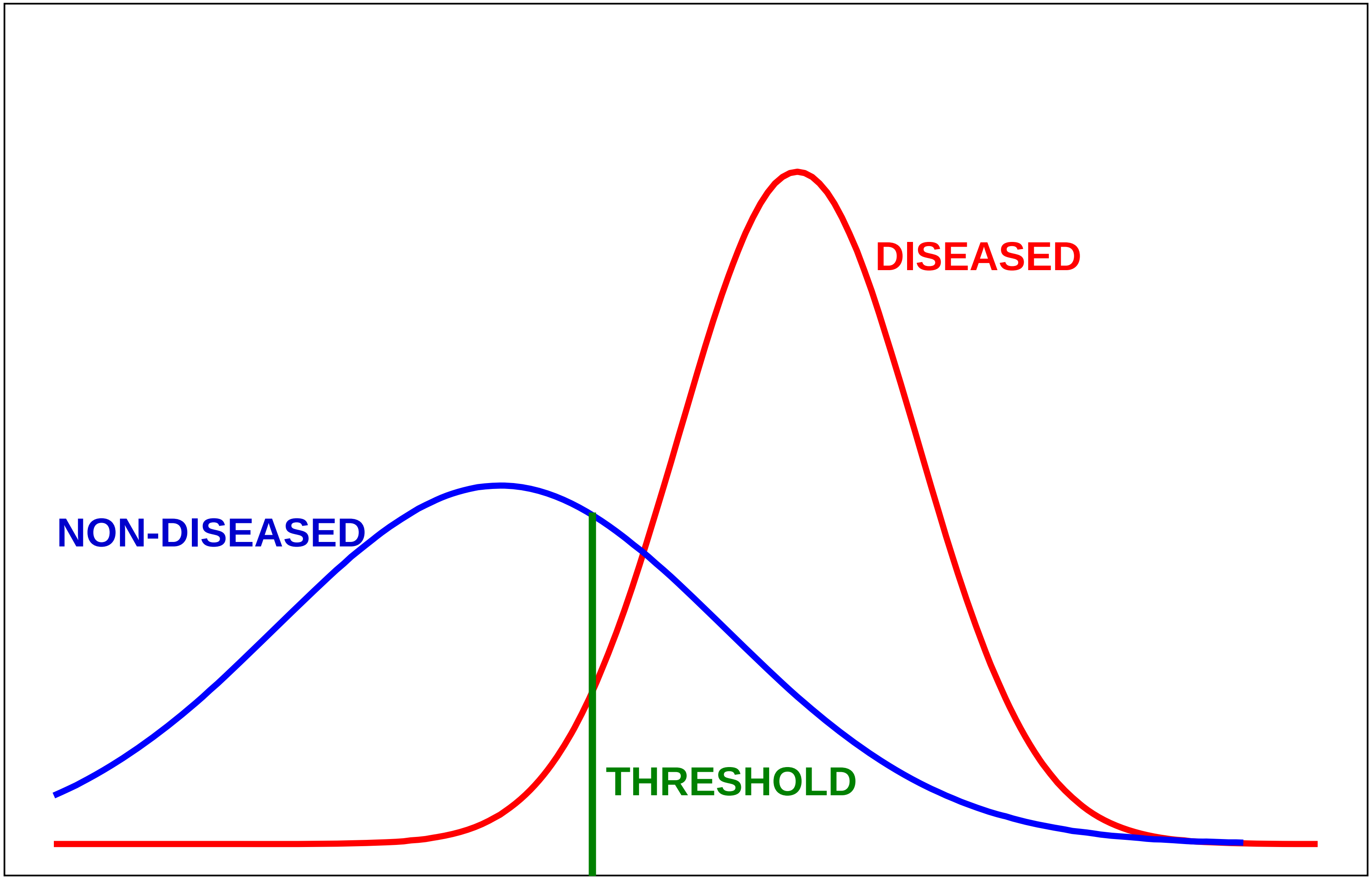


Figure 2: Receiver operating characteristic curves of MR imaging values versus combined MR imaging–MR spectroscopic imaging *(MRSI)* values for all readers.

Topics- revisited

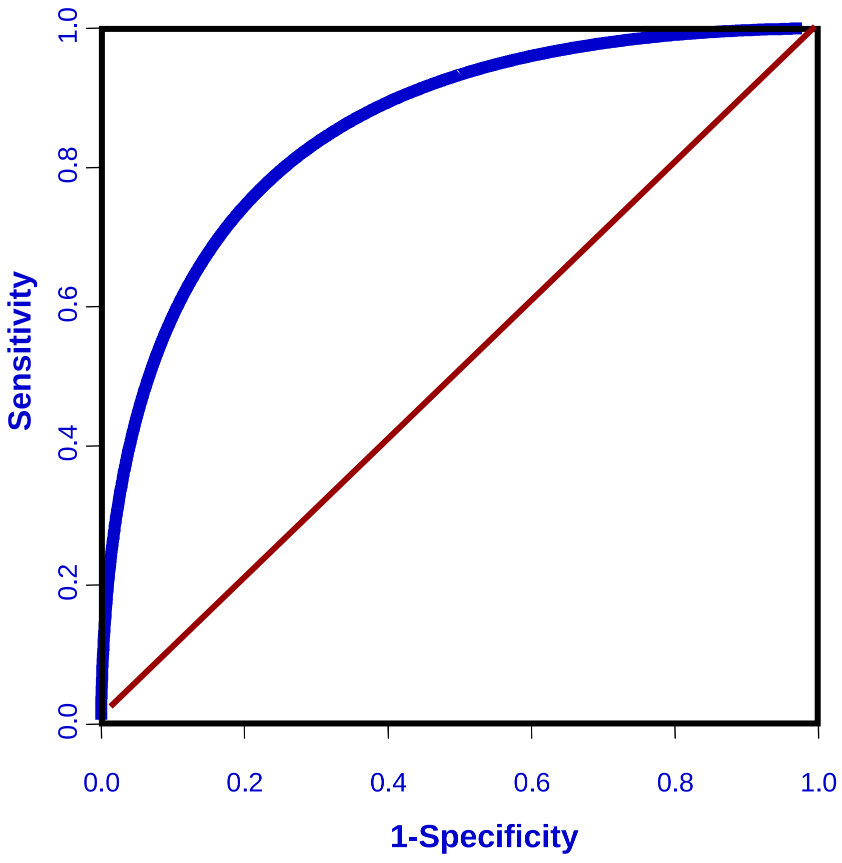
- Several types of errors and error rates are of interest.
- Evaluation of accuracy of forensic analyses
  - Define the task!
  - Define measures of accuracy.
  - Conduct experiments.
  - Monitor practice.

Fundamental conceptualization: Threshold for test positivity



ROC curves

- **Binary truth**
- **ROC curve is plot of all pairs of (1-Spec., Sens.) as positivity threshold varies**



Variability among readers in NCTC study

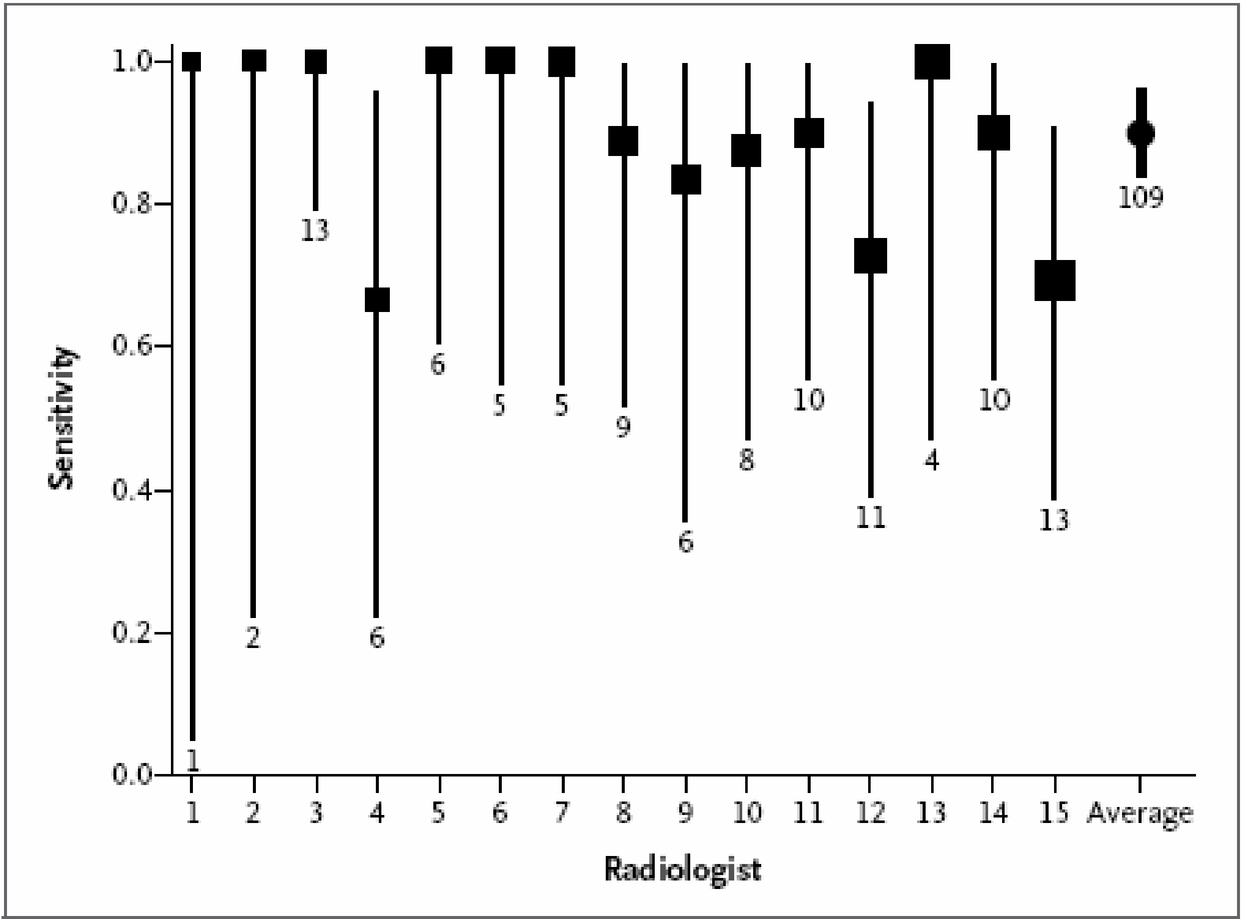


Figure 1. Individual Estimates of the Sensitivity of CT Colonography for the Detection of Adenomas or Cancers.