University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Data-driven Decision Support to Improve Breast Cancer Diagnosis and Outcomes

Elizabeth Burnside, MD, MPH, MS

Departments:
Radiology
Population Health
Biostatistics and Medical Informatics
Industrial and Systems Engineering

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Informatics & Breast Imaging

- General Overview
  - History/Motivation
- Methodological Considerations
  - Algorithms & metrics to measure performance
- Projects
  - Improving mammographic predictions
  - Improving image-guided core biopsy

# Motivation

- Information overload
  - Medical articles in pubmed-online
  - EHR information
  - Genetic risk factors
- Human decision making involves heuristics that may not scale up alone

# History

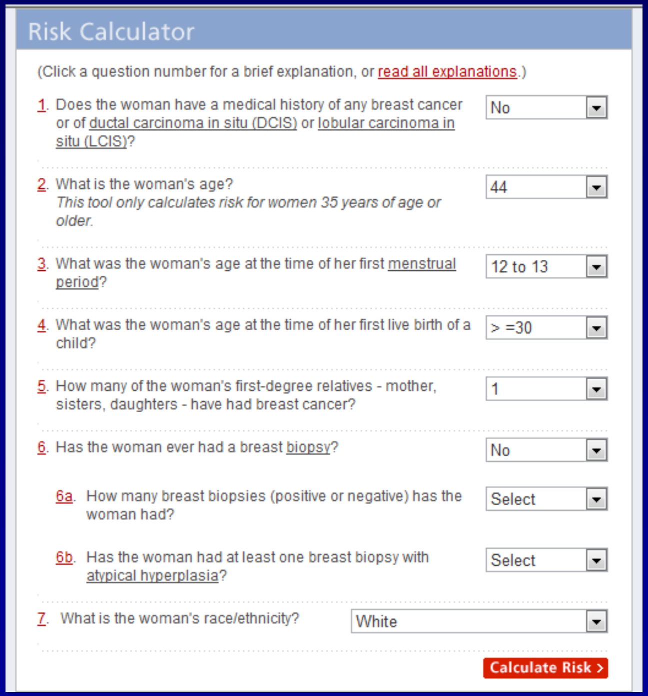
- Tools first conceived in:
  - Leeds Abdominal Pain System went operational in 1971

System = 91.8%
Physician = 79.6 %

- Barriers
  - Not integrated in clinical workflow
  - Errors

University of Wisconsin
SCHOOL OF MEDICINE AND PUBLIC HEALTH

# The Gail Model


Risk Calculator

(Click a question number for a brief explanation, or read all explanations.)

1. Does the woman have a medical history of any breast cancer or of ductal carcinoma in situ (DCIS) or lobular carcinoma in situ (LCIS)? — No

2. What is the woman's age? *This tool only calculates risk for women 35 years of age or older.* — 44

3. What was the woman's age at the time of her first menstrual period? — 12 to 13

4. What was the woman's age at the time of her first live birth of a child? — > =30

5. How many of the woman's first-degree relatives - mother, sisters, daughters - have had breast cancer? — 1

6. Has the woman ever had a breast biopsy? — No

6a. How many breast biopsies (positive or negative) has the woman had? — Select

6b. Has the woman had at least one breast biopsy with atypical hyperplasia? — Select

7. What is the woman's race/ethnicity? — White

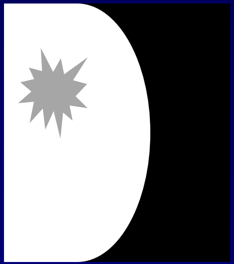
Calculate Risk >

- Uses data (BCDDP)
- Predicts Breast CA
  - Five year/lifetime risk

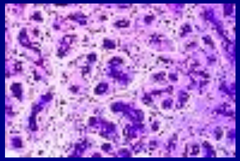
Low signal predictors

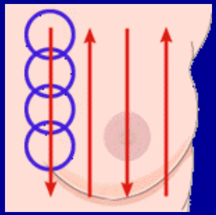
http://www.cancer.gov/bcrisktool/Default.aspx

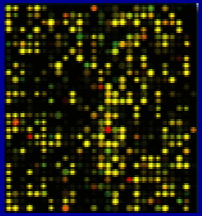
# Predictive Information



AGE



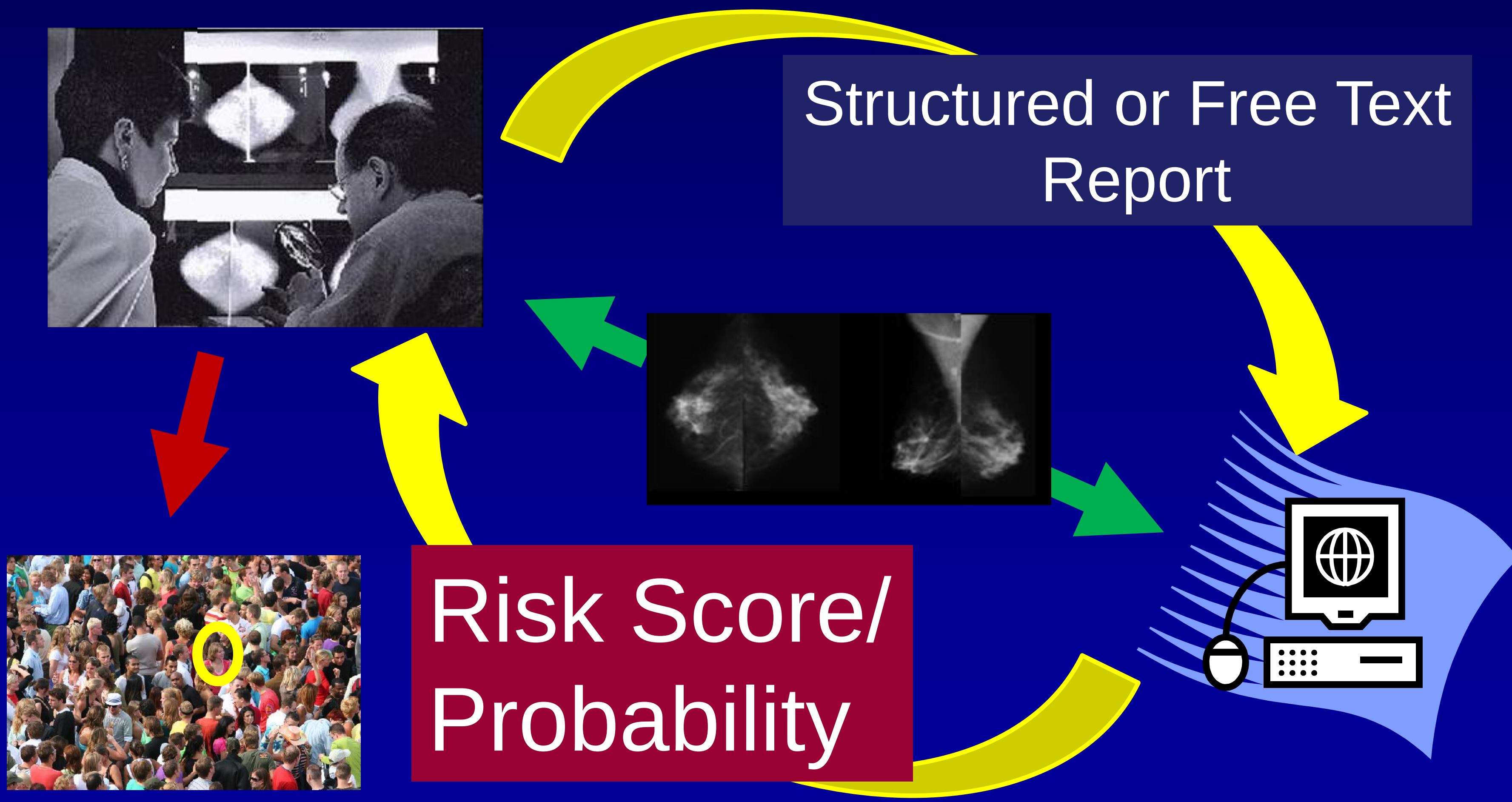





Breast Cancer

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Human Computer Interaction
# COMMUNICATION



University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

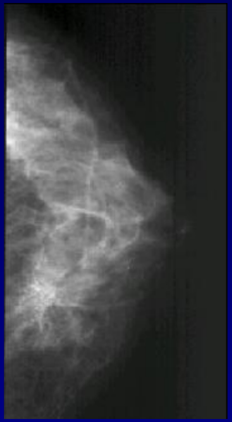
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# The Mammography Risk Prediction Project

Elizabeth Burnside, MD, MPH, MS
C. David Page, PhD
Jude Shavlik, PhD
Charles Kahn, MD (MCW)

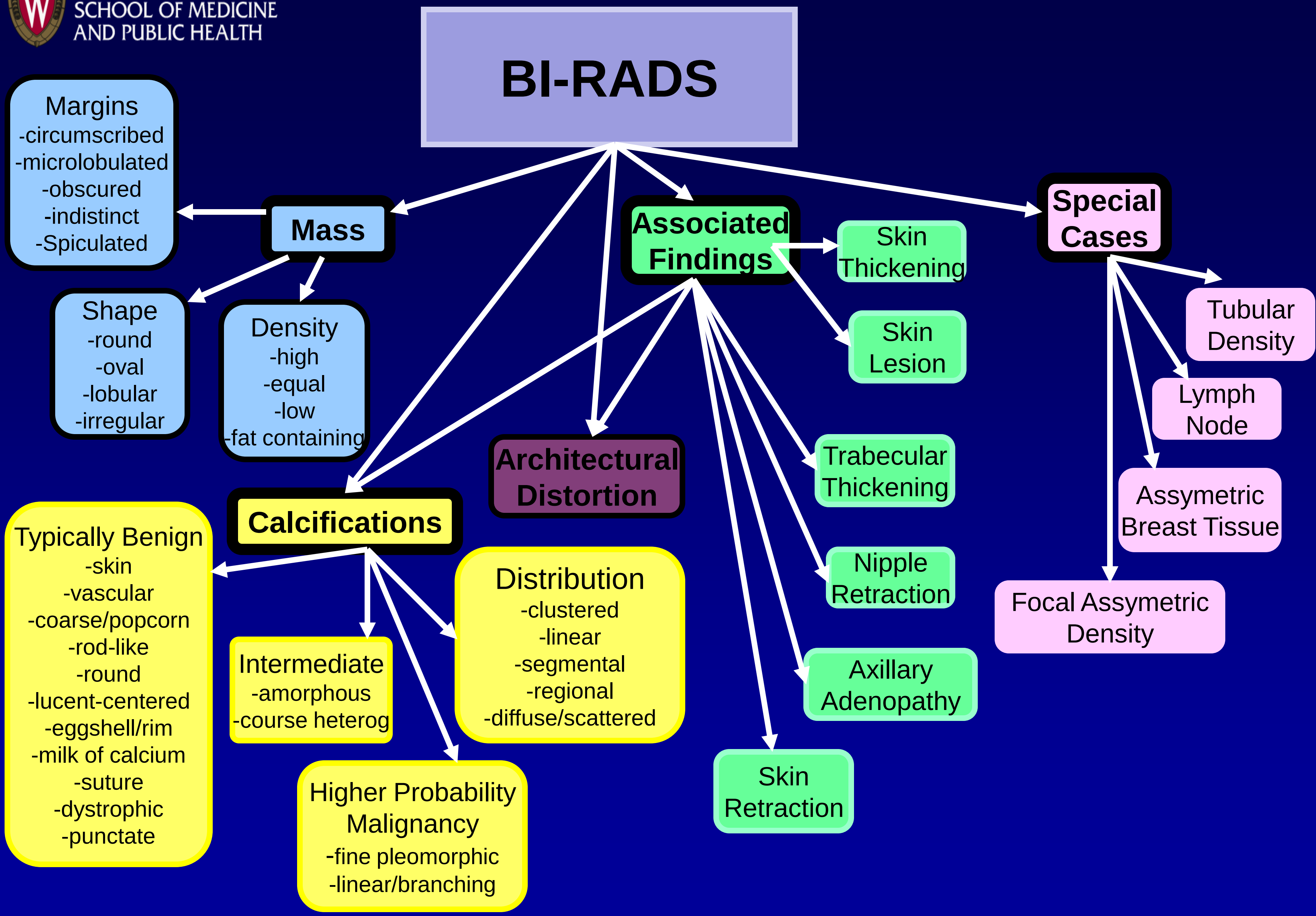
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Background-Opportunity



- 200,000 breast cancer diagnosed in US
- 20 million mammograms per year
  - False positives
    - Millions of diagnostic mammograms/US
    - Hundreds of thousands biopsies
  - False negative
    - 10-30% of breast cancers not detected on mammography
- Variability of practice impacts many women
- **Evidence-based decision support has the potential to drive substantial improvement**

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

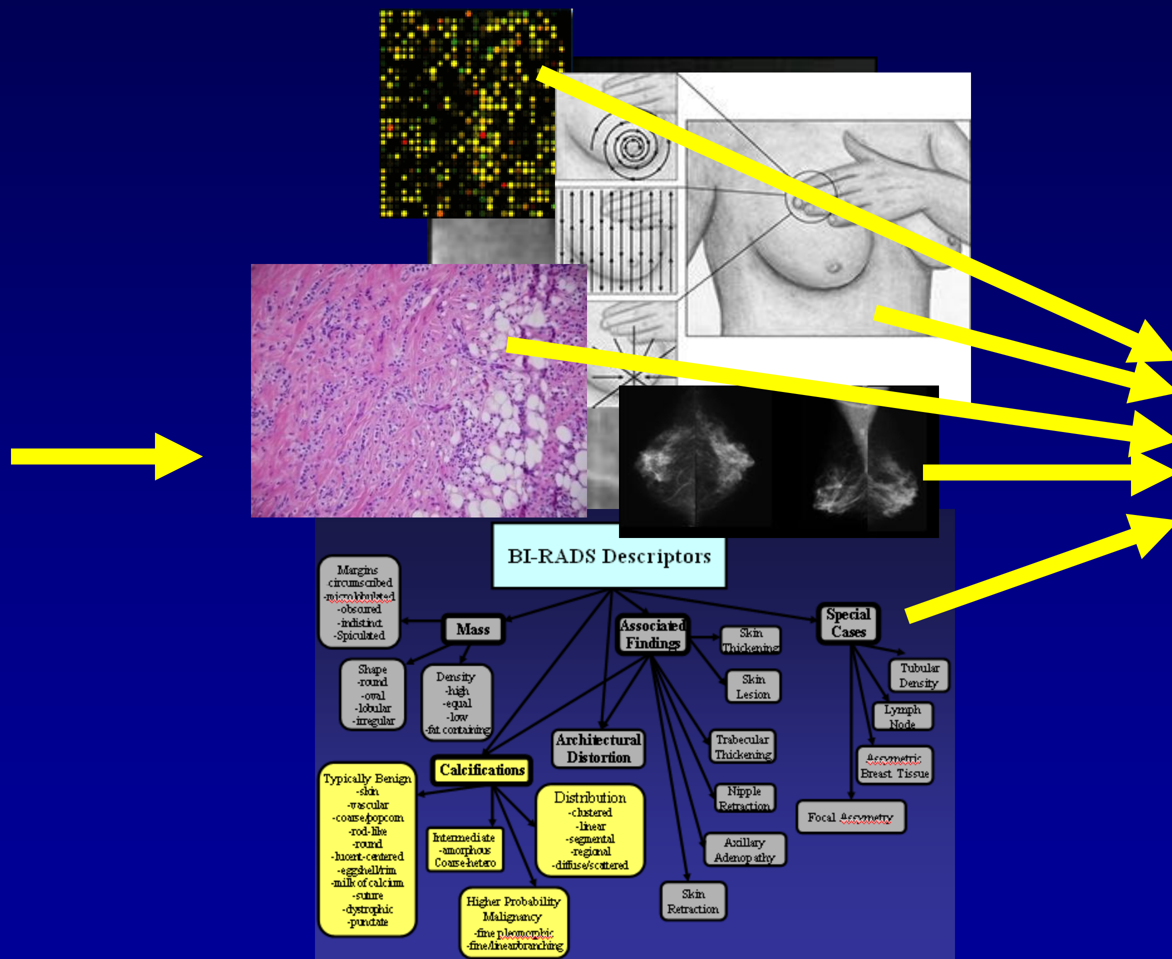


# Breast Cancer Probability Based on BI-RADS Category

| | |
|---|---|
| BI-RADS 0: | Needs Additional Imaging |
| BI-RADS 1: | Negative |
| BI-RADS 2: | Benign |
| BI-RADS 3: | Probably Benign |
| BI-RADS 4: | Suspicious for malignancy |
| BI-RADS 5: | Highly suggestive of malignancy |

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

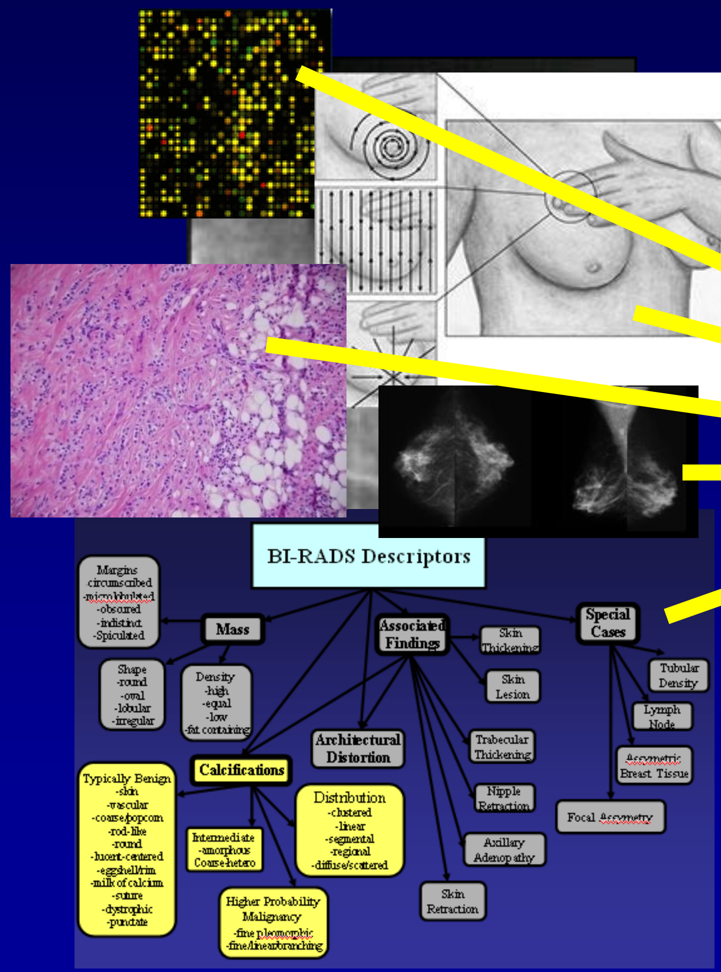
# Breast Cancer Predictors




**BI-RADS Category**

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

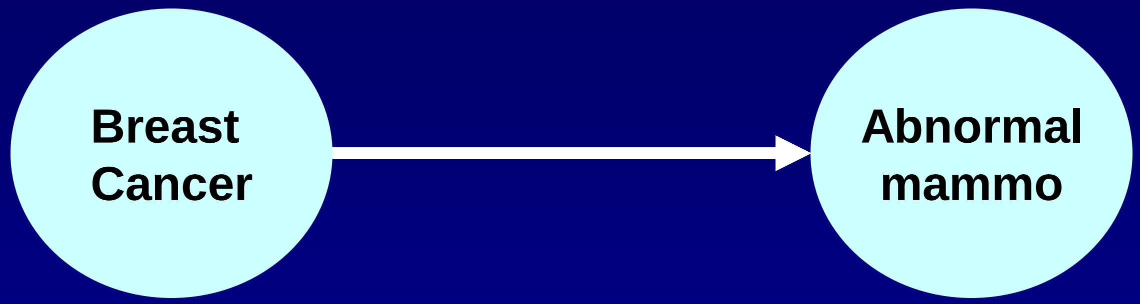
# Breast Cancer Predictors




**Risk**

**Detection** **Characterization** **Classification**

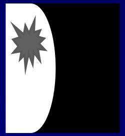
# Bayesian Networks

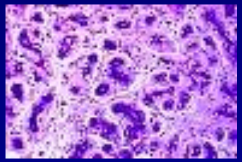


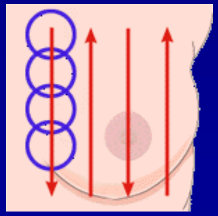
# What do we want to know?

We generally learn P(f|d)

Breast cancer



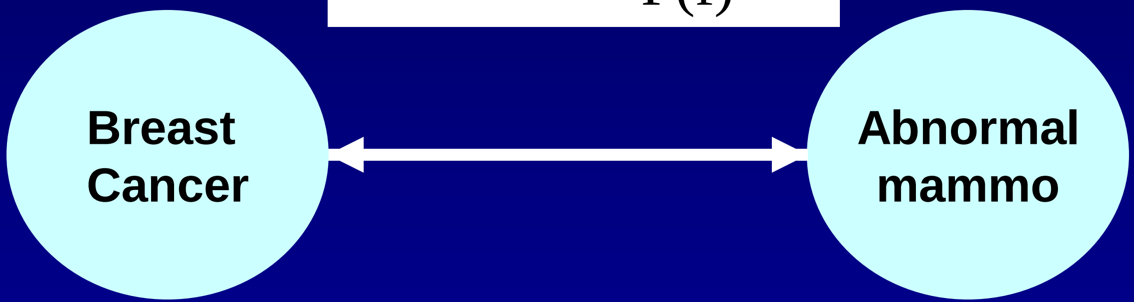




We generally want to know P(d|f)

# Bayesian Networks

$$P(d \mid f) = \frac{P(f \mid d)\, P(d)}{P(f)}$$

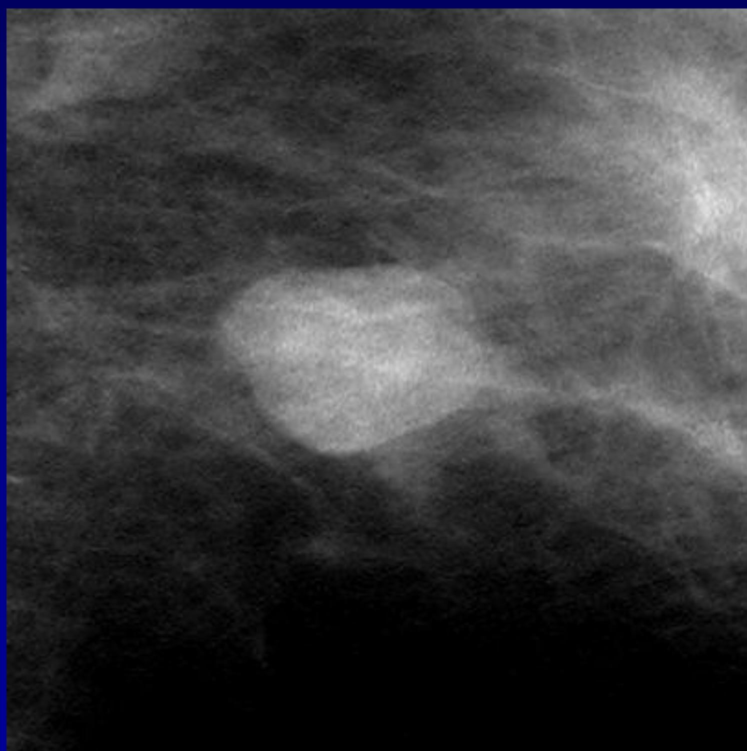


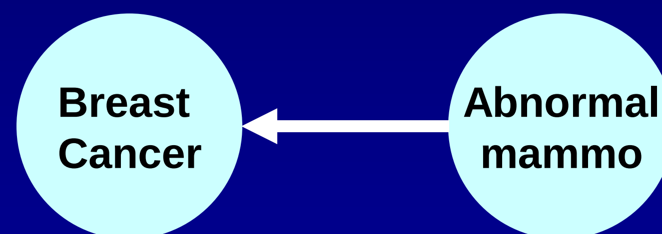

| Breast CA | Symbol | Prob |
|---|---|---|
| Yes | P(d) | 1.0% |
| No | P(d-) | 99.0% |

| Breast CA | mammo | Symbol | Prob |
|---|---|---|---|
| Yes | Yes | P(f\|d) | 90% |
| Yes | No | P(f-\|d) | 10% |
| No | Yes | P(f\|d-) | 10.1% |
| No | No | P(f-\|d-) | 89.9% |

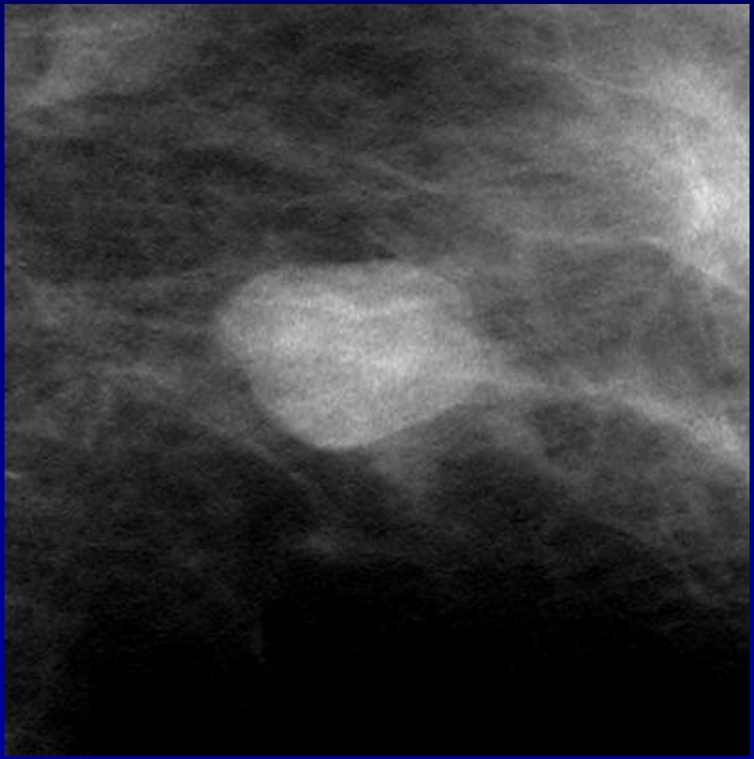
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Estimates

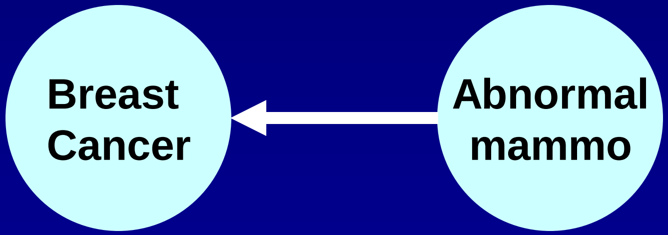




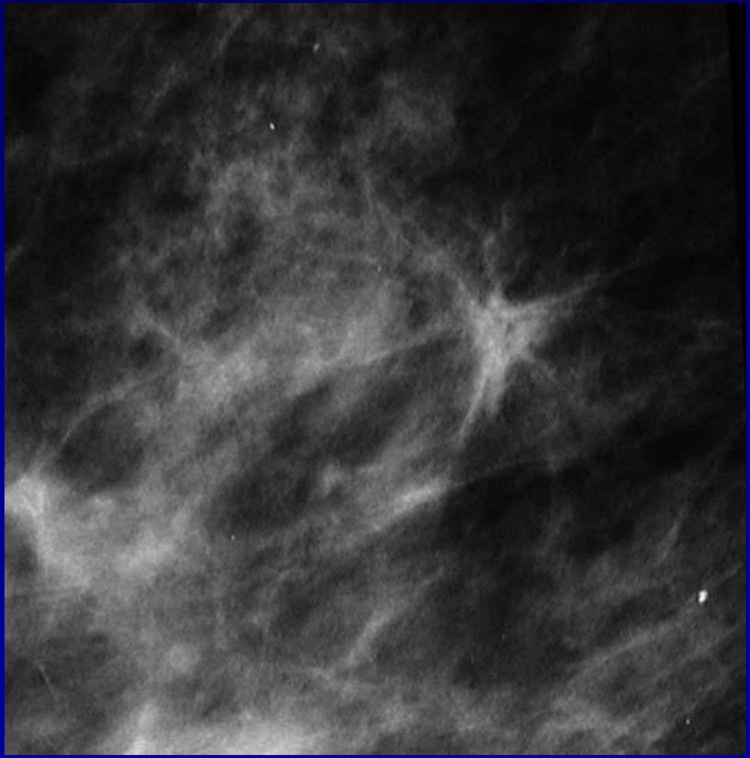
# Estimates

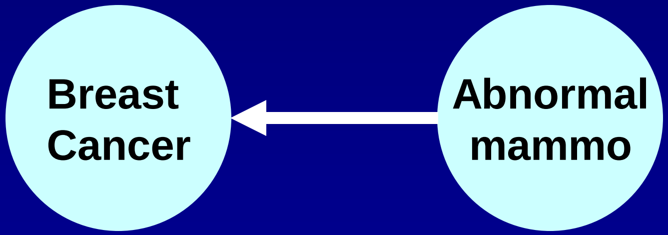


8.3%



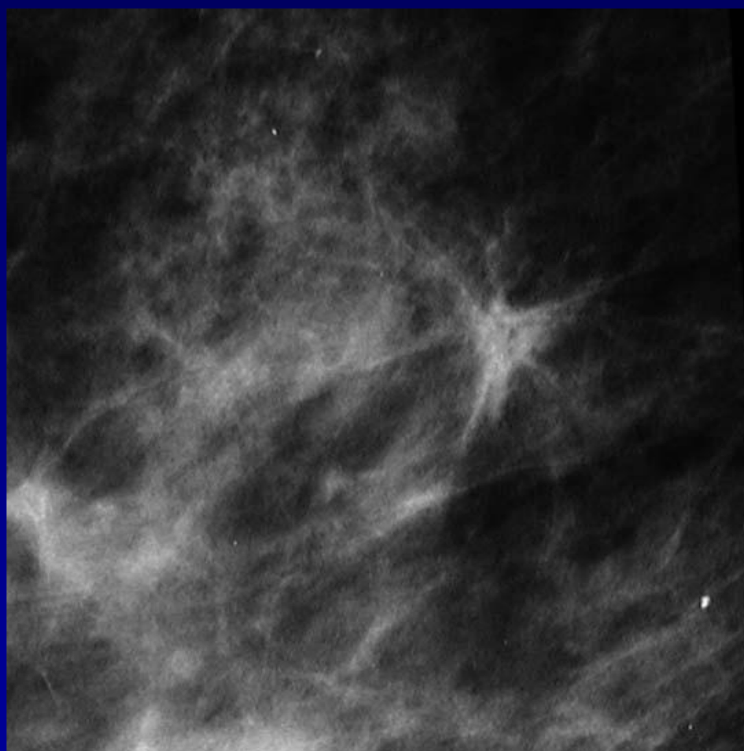
# Estimates



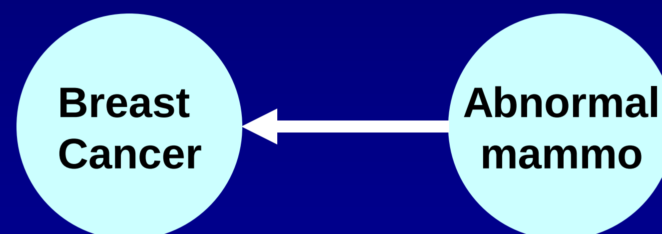


University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

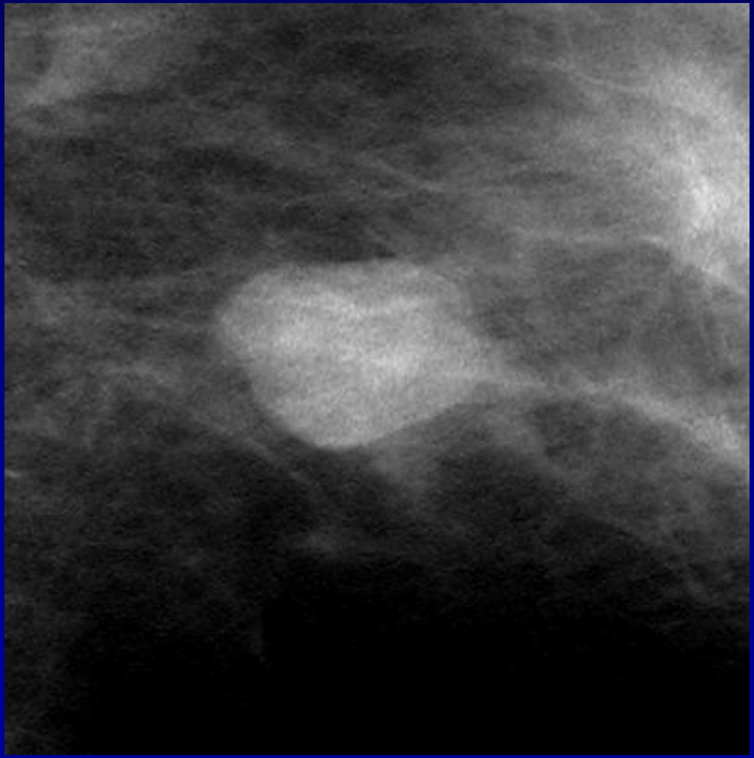
# Estimates



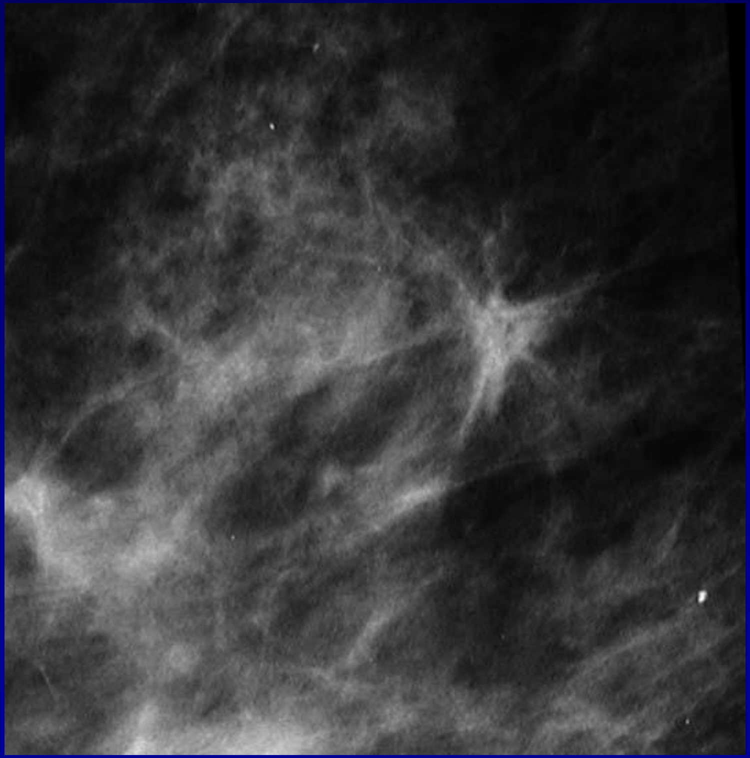
**8.3%**



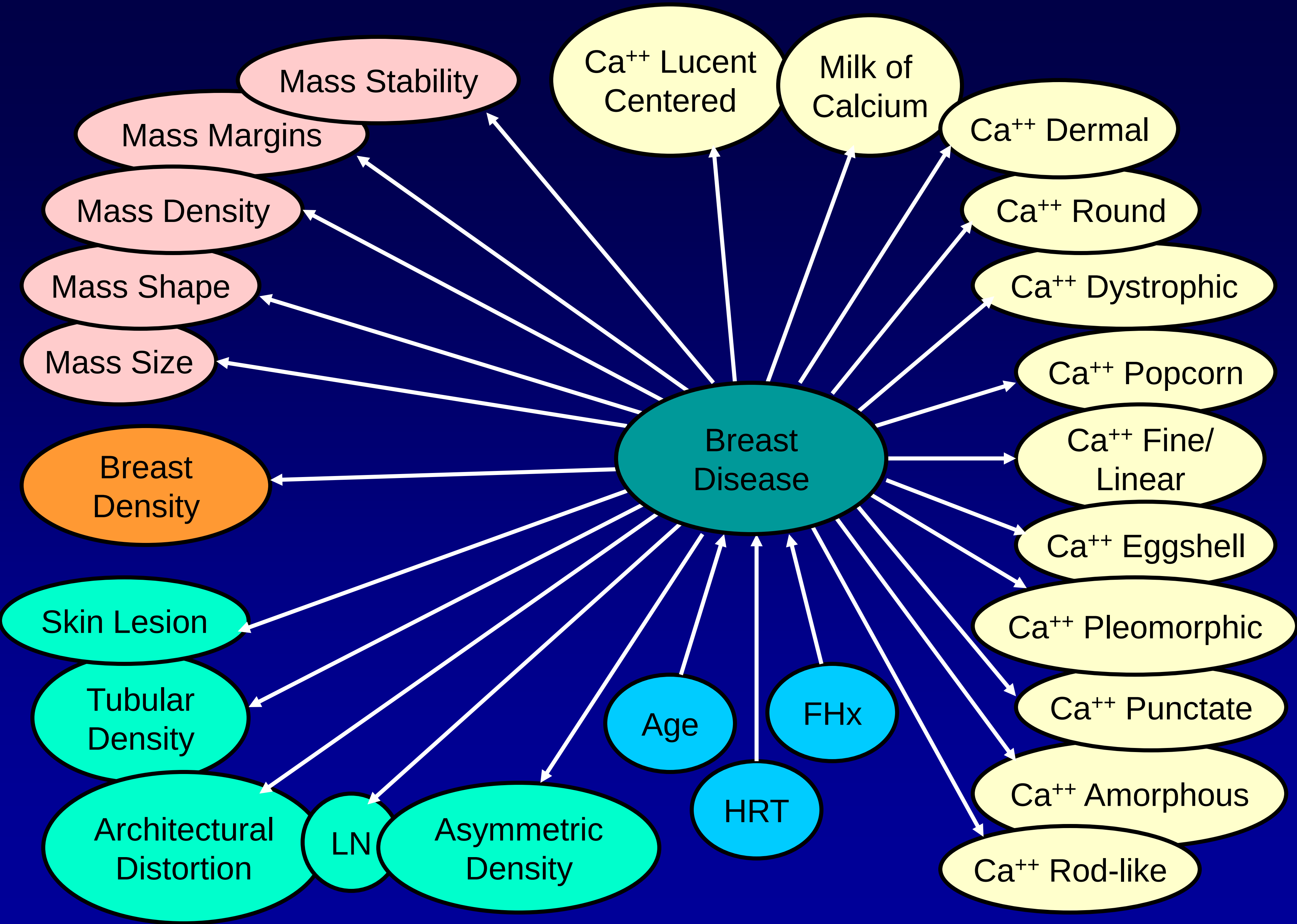
# Probability Estimates



**8.3%**

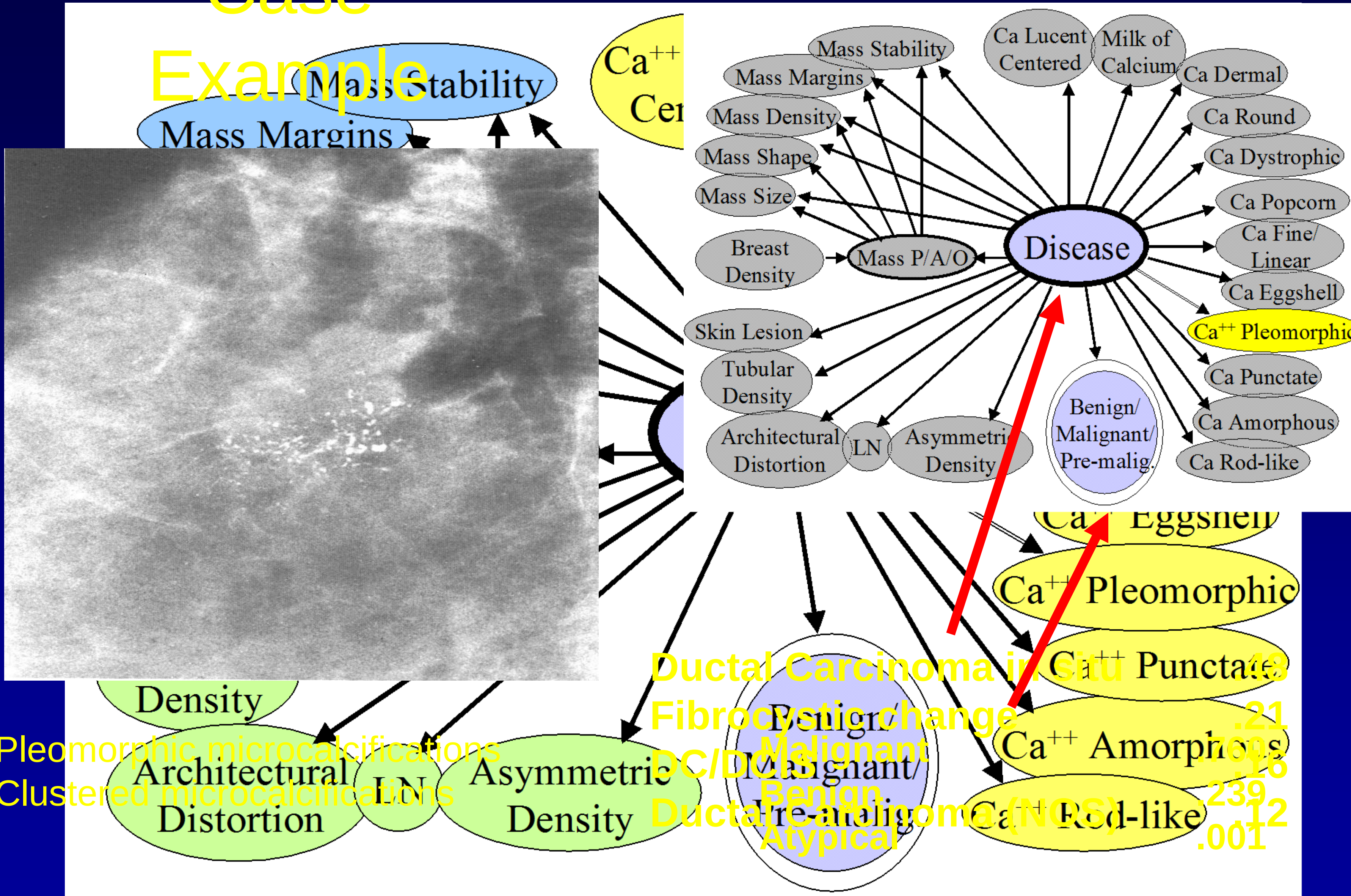


**8.3%**



Mass Stability
Mass Margins
Mass Density
Mass Shape
Mass Size
Breast Density
Skin Lesion
Tubular Density
Architectural Distortion
LN
Asymmetric Density
Age
HRT
FHx
Breast Disease
Ca++ Lucent Centered
Milk of Calcium
Ca++ Dermal
Ca++ Round
Ca++ Dystrophic
Ca++ Popcorn
Ca++ Fine/ Linear
Ca++ Eggshell
Ca++ Pleomorphic
Ca++ Punctate
Ca++ Amorphous
Ca++ Rod-like

# Case Example




Pleomorphic microcalcifications
Clustered microcalcifications

Ductal Carcinoma in Situ .48
Fibrocystic change .21
DC/DCIS .16
Ductal Carcinoma (NOS) .12

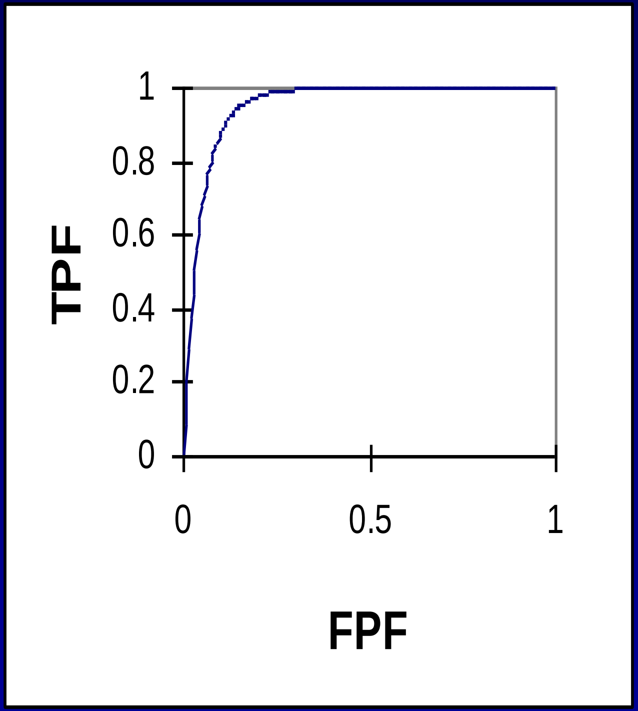
Malignant .760
Benign .239
Atypical .001

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Teaching cases

- 105 cases
- Created ROC curve
- Comparable to
  - Neural networks

Az = .953



# Training on Data

- Motivation
  - Accurate probabilities are critical
  - Some are not available in literature
  - Modeling the relevant patient population is possible with training

***Expert & Rule Based*** → ***Machine Learning***

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Idea: Data Driven Decisions

| Patient Table |
| --- |
| Age |
| Personal Hx Breast CA |
| Family Hx Breast CA |
| ….. |

| Mammogram Table |
| --- |
| Mass shape |
| Mass margins |
| Mass density |
| Mass size |
| Mass stability |
| MicroCa$^{++}$ shape |
| MicroCa$^{++}$ distribution |
| BI-RADS category |
| ….. |

| Registry Table |
| --- |
| Patient ID |
| Margin status |
| Grade |
| Prior radiation |
| ….. |

| Biopsy Table |
| --- |
| Needle size |
| Number of samples |
| Post-proc appearance |
| Accurate clip position |
| ….. |

| Pathology Table |
| --- |
| Pathology Result |
| Concordance |
| Recommendation |
| ….. |

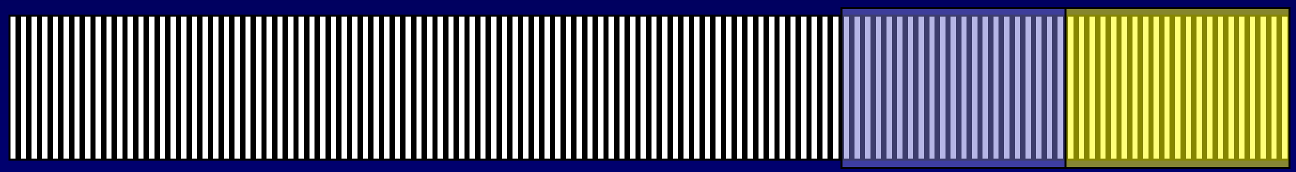
# Idea: Data Driven Decisions

| Patient | Abnormality | Date | Calc | Mass | Mass Size | location | B/M |
|---|---|---|---|---|---|---|---|
| P1 | 1 | 5/08 | N | Y | 3 mm | RUO | B |
| P1 | 2 | 5/10 | Y | Y | 5 mm | RUO | M |
| P1 | 3 | 5/10 | N | Y | 3 mm | LLI | B |
| P2 | 4 | 6/09 | N | Y | N/A | RLI | B |
| … | … | … | … | … | … | … | |

# Data

- Our dataset contains
  - 350 malignancies
  - 65,630 benign abnormalities
- Linked to cancer registry data
  - Outcomes (benign/malignant)

# Training the BN



**Training set**

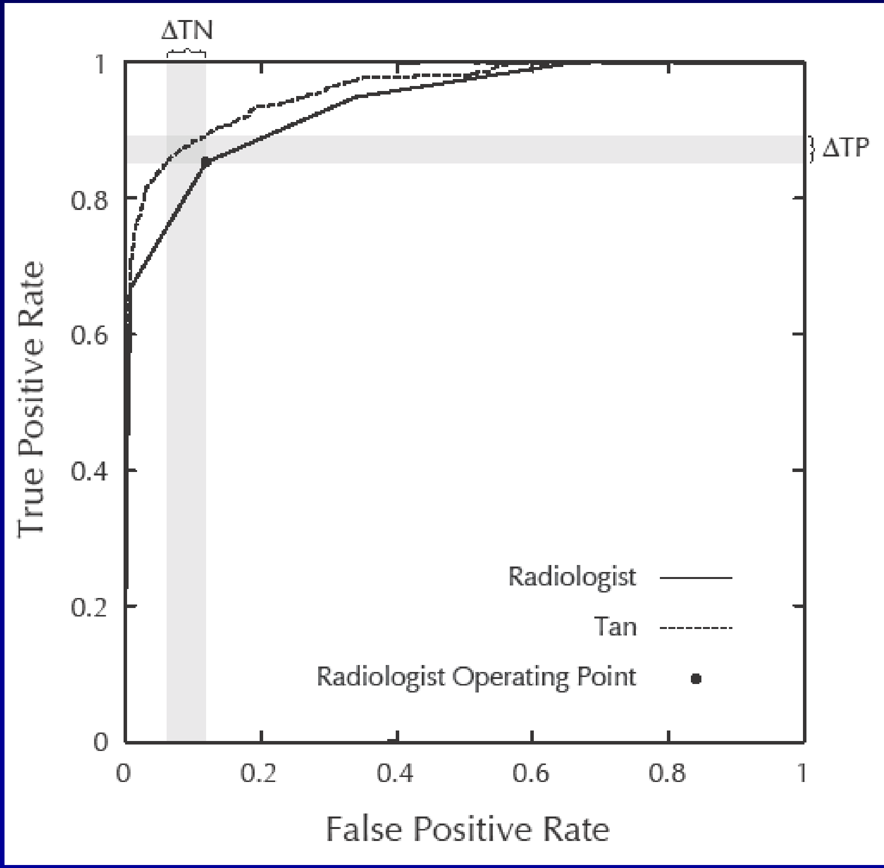
**Tuning set**

**Test set**

- Standard Machine learning
  - Use known cases to train
  - Use the tuning set for optimal training
  - Performance based on hold out test set

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Performance




- AUC 0.960 *vs*. 0.939
  - $P < 0.002$
- Sensitivity
  - 90.0% *vs*. 85.3%
  - $P < 0.001$
- Specificity
  - 93.9% *vs*. 88.1%
  - $P < 0.001$

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

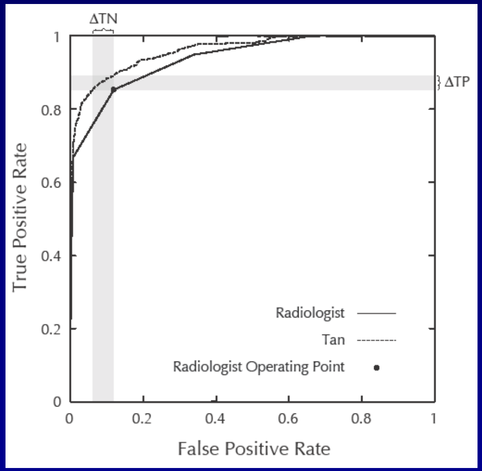
# What does that mean?

- At a specificity of 90%

  **38 conversions FN → TP**

- At a sensitivity of 85%

  **4226 conversions FP → TN**



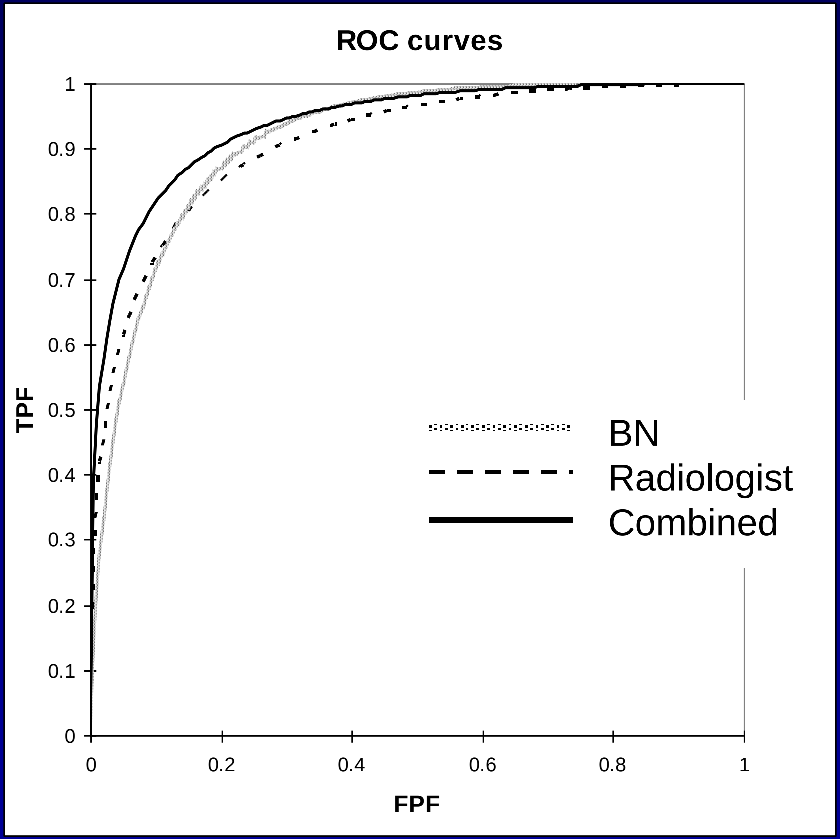
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Ultimately Decision Support Aids the Physician

- Output of the system is
  - Advisory
  - Utilized in the clinical context
  - System performance alone is not the point
  - Performance/Physician performance is the key to improvement of care

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Collaborative Experiment




Radiologist

.916

Bayes Net

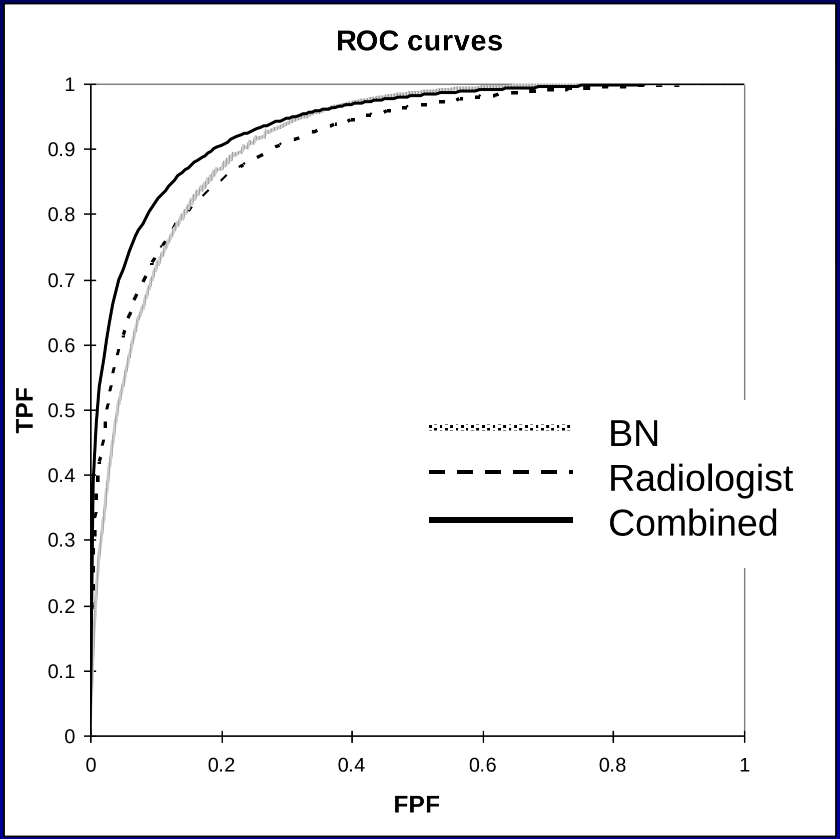
.919

Combined

.948

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Results




Radiologist
.916

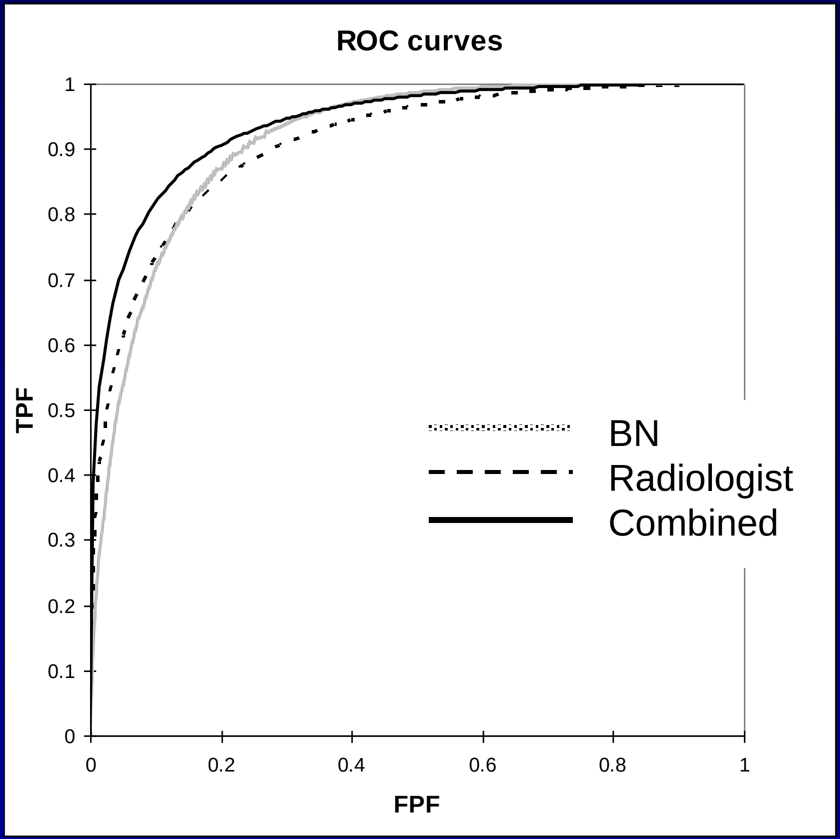
Bayes Net
.919

p=.03

Combined
.948

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Results




Radiologist

.916

Bayes Net

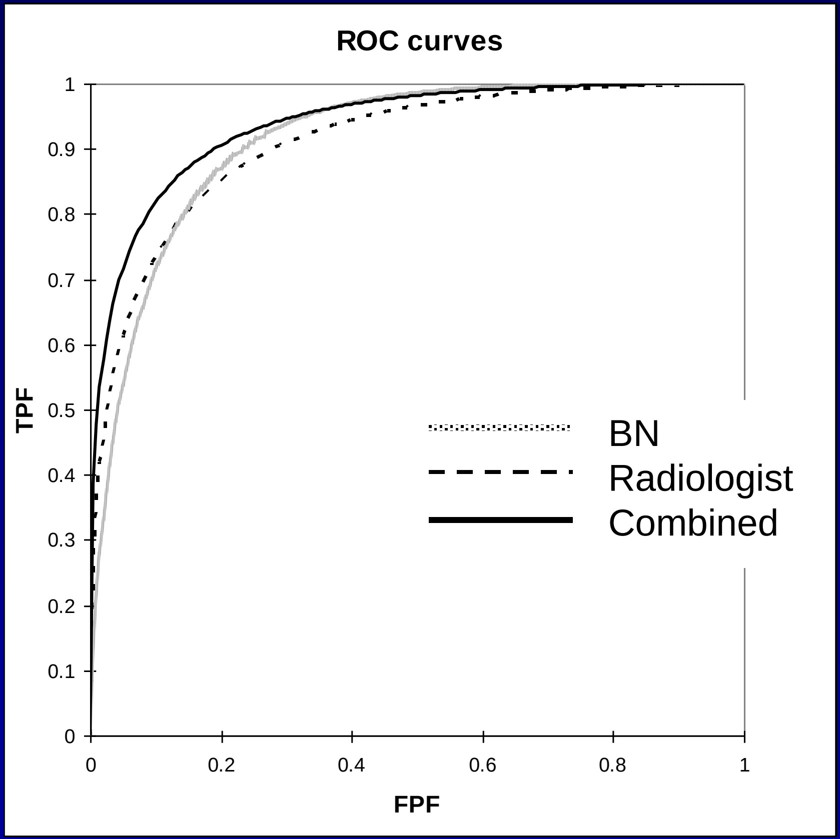
.919

p=.065

Combined

.948

# Results




Radiologist

.916

p=.99

Bayes Net

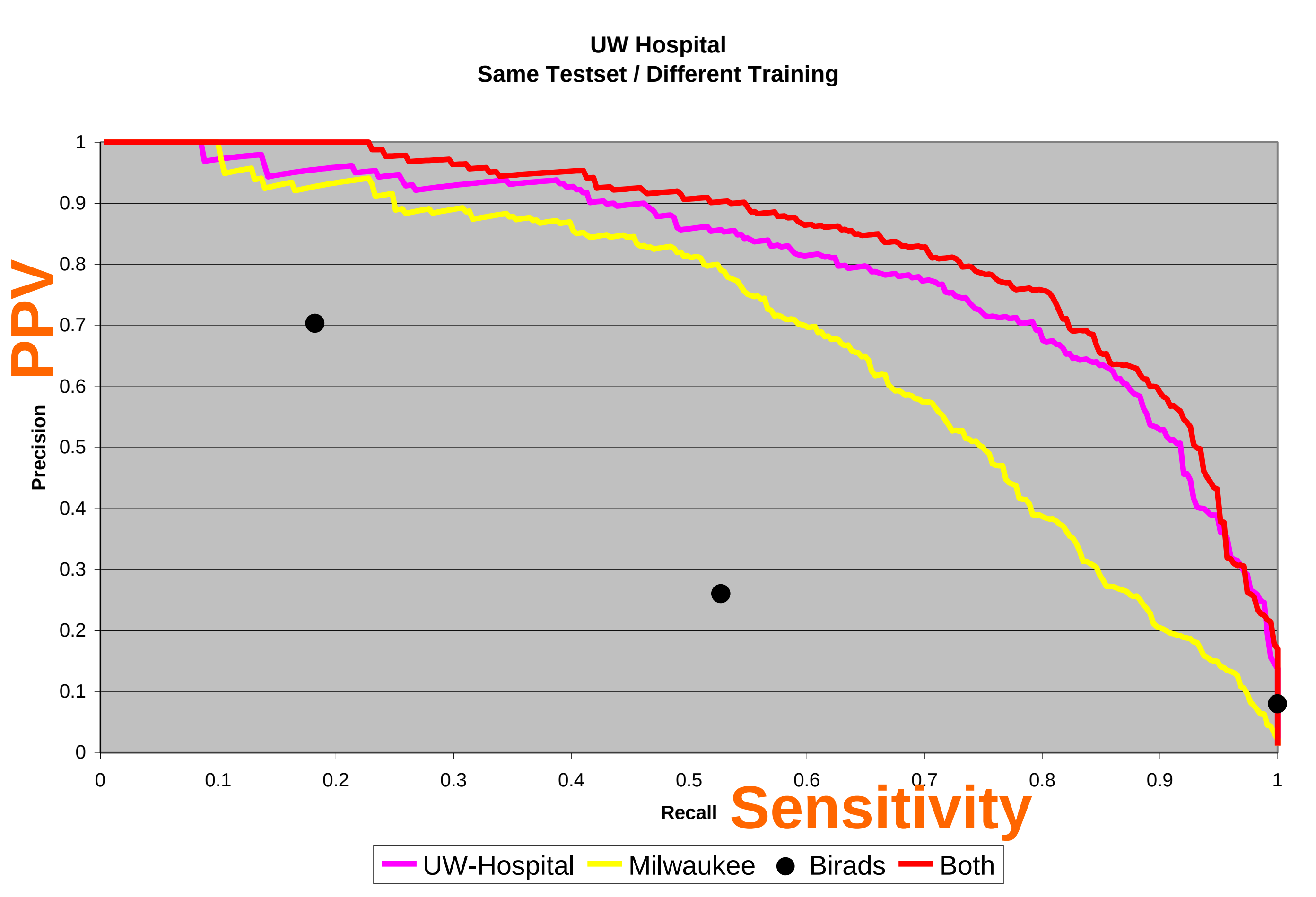
.919

Combined

.948

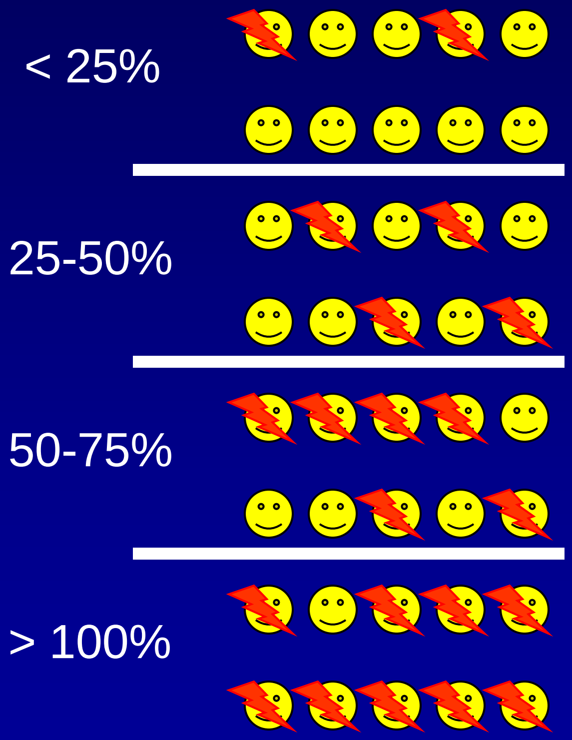
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

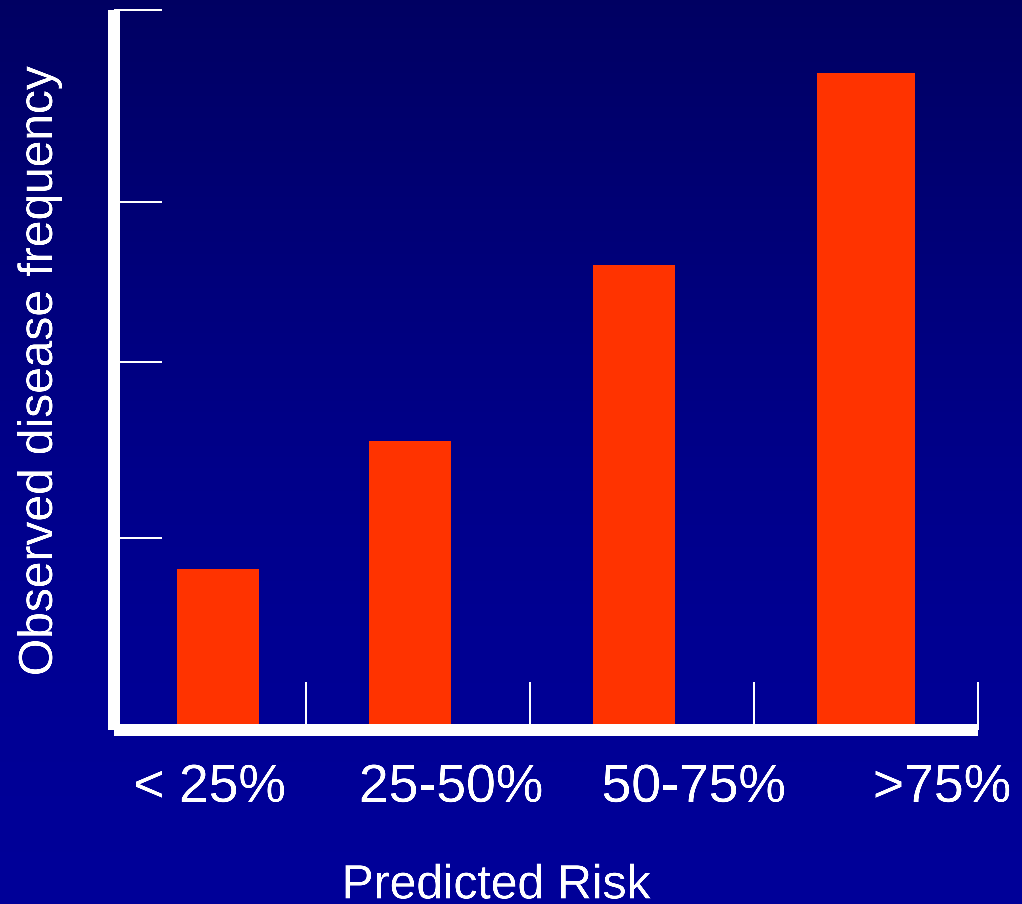
# Precision Recall Curves



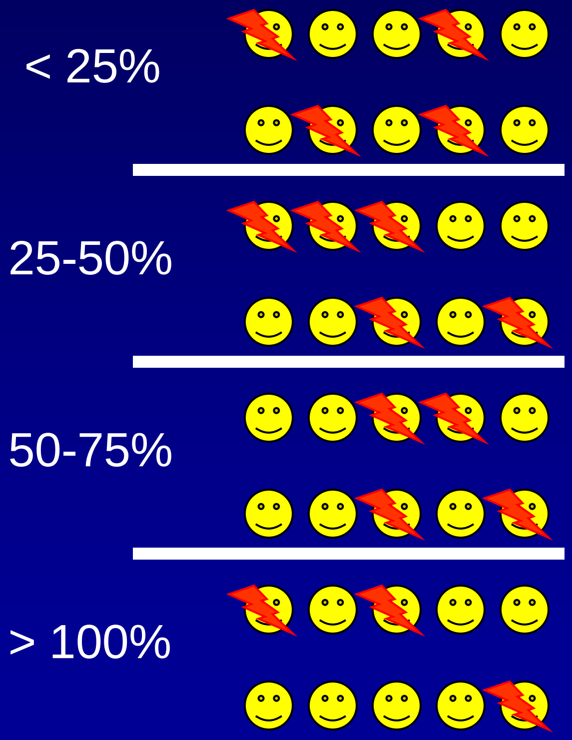
University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

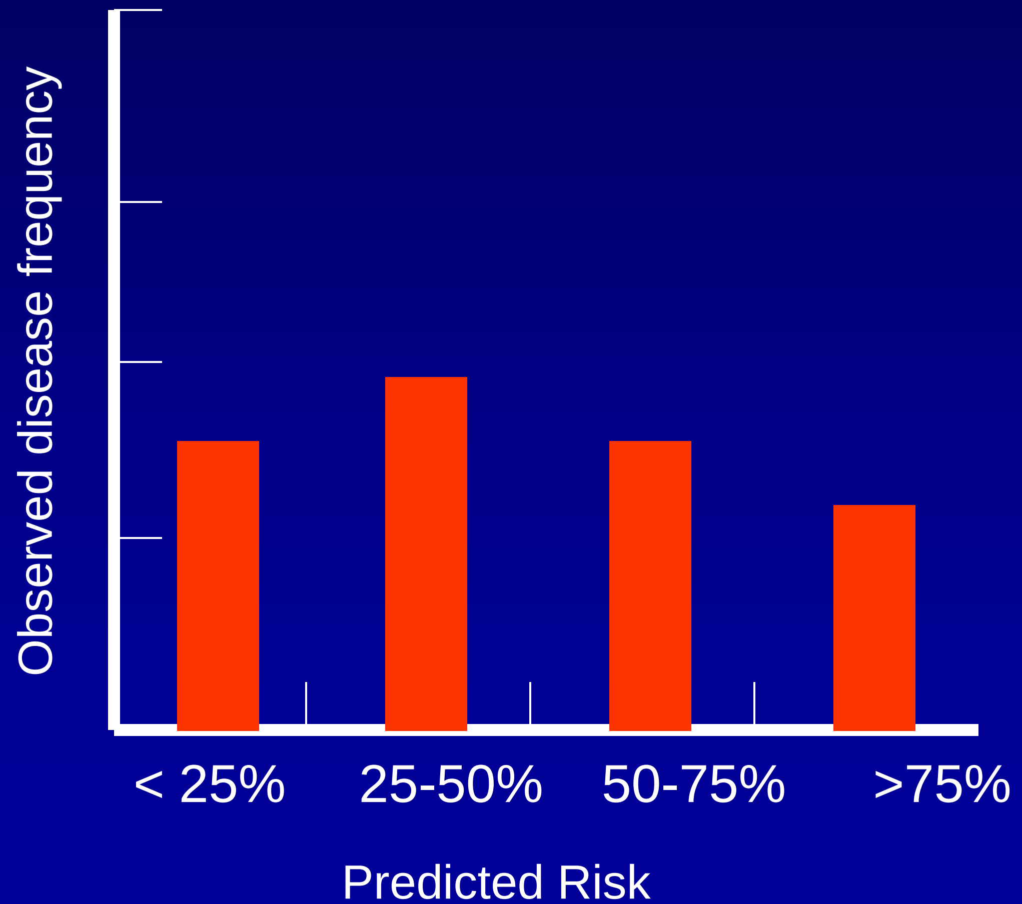
# Calibration Curves



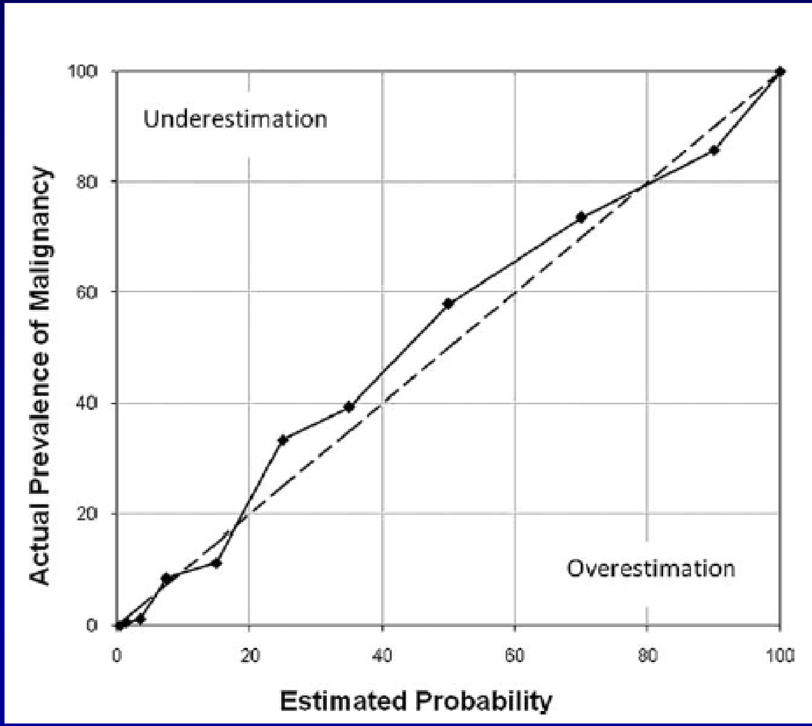
# Calibration Curves



University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Calibration




- Hosemer-Lemishow goodness of fit

Ayer, T., et al., *Breast cancer risk estimation with artificial neural networks revisited: discrimination and calibration.* Cancer, 2010. **116**(14): p. 3310-21.

# Idea: Data Driven Decisions

| **Patient Table** |
|---|
| Age |
| Personal Hx Breast CA |
| Family Hx Breast CA |
| ….. |

| **Mammogram Table** |
|---|
| Mass shape |
| Mass margins |
| Mass density |
| Mass size |
| Mass stability |
| MicroCa$^{++}$ shape |
| MicroCa$^{++}$ distribution |
| BI-RADS category |
| ….. |

| **Registry Table** |
|---|
| Patient ID |
| Margin status |
| Grade |
| Prior radiation |
| ….. |

| **Biopsy Table** |
|---|
| Needle size |
| Number of samples |
| Post-proc appearance |
| Accurate clip position |
| ….. |

| **Pathology Table** |
|---|
| Pathology Result |
| Concordance |
| Recommendation |
| ….. |

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Idea: Data Driven Decisions

| Patient | Abnormality | Date | Calc | Mass | Mass Size | location | B/M |
|---|---|---|---|---|---|---|---|
| P1 | 1 | 5/08 | N | Y | 3 mm | RUO | B |
| P1 | 2 | 5/10 | Y | Y | 5 mm | RUO | M |
| P1 | 3 | 5/10 | N | Y | 3 mm | LLI | B |
| P2 | 4 | 6/09 | N | Y | N/A | RLI | B |
| … | … | … | … | … | … | … | |

Not independent and identically distributed (IID)

# Algorithmic Opportunities

- Inductive logic programming (ILP)
- Statistical Relational Learning (SRL)
- Natural Language Processing (NLP)

# Algorithmic Opportunities

- Inductive logic programming (ILP)
- Statistical Relational Learning (SRL)
- Natural Language Processing (NLP)

# ILP

Abnormality A in
Mammogram M for
Biopsy B in
Patient P

Is malignant if:

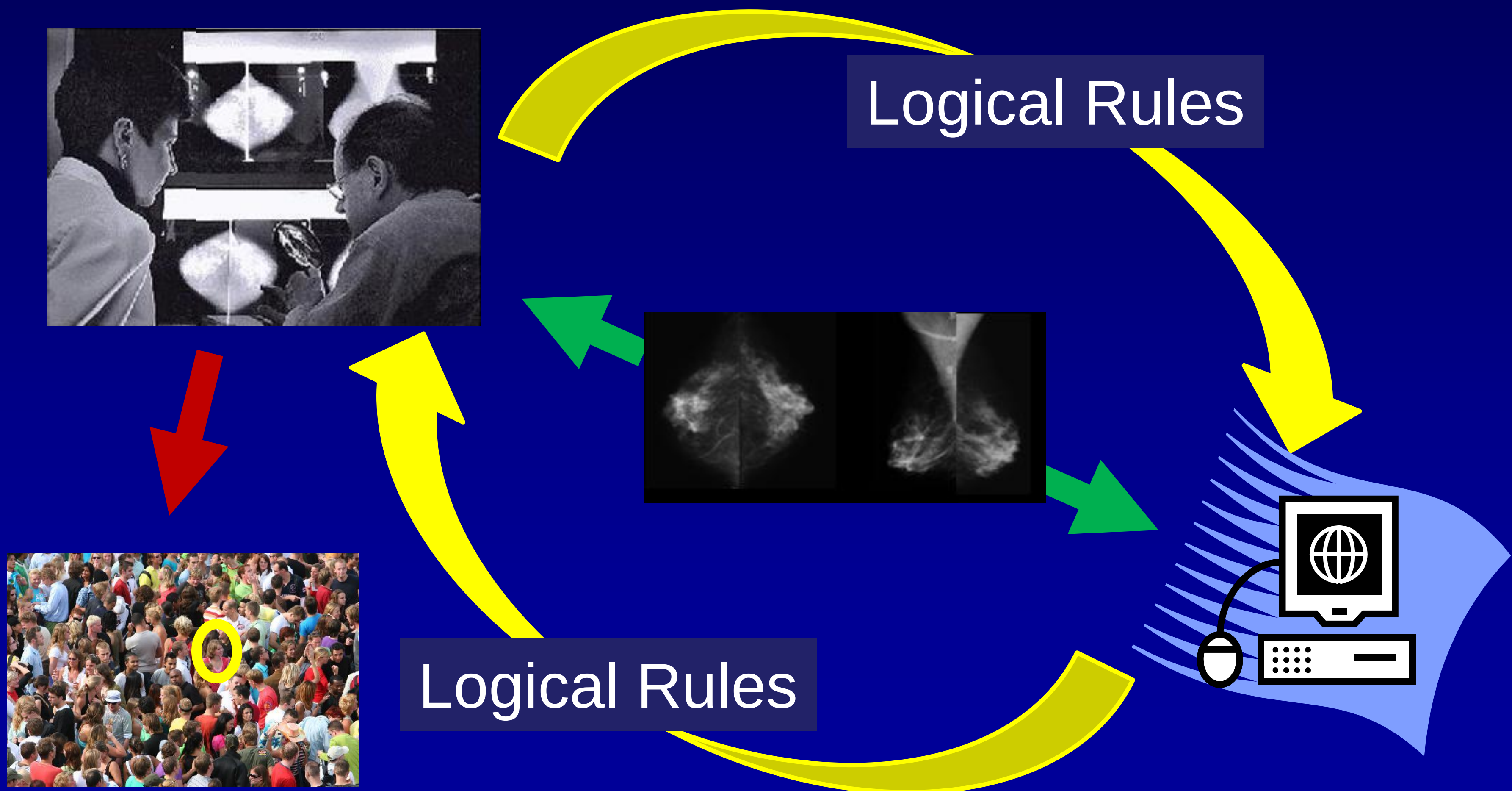
Malignant (A) IF
- A has mass present
- A has stability increasing
- P has family history of breast cancer
- B has atypia

# How does it work?

- Learn if-then rules that will become features in a predictive model
  - Inductive logic programming (ILP) to learn the rules
  - Integrated search strategy for constructing and selecting rules for classifcation algorithm

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Human Computer Interaction
# COMMUNICATION



University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

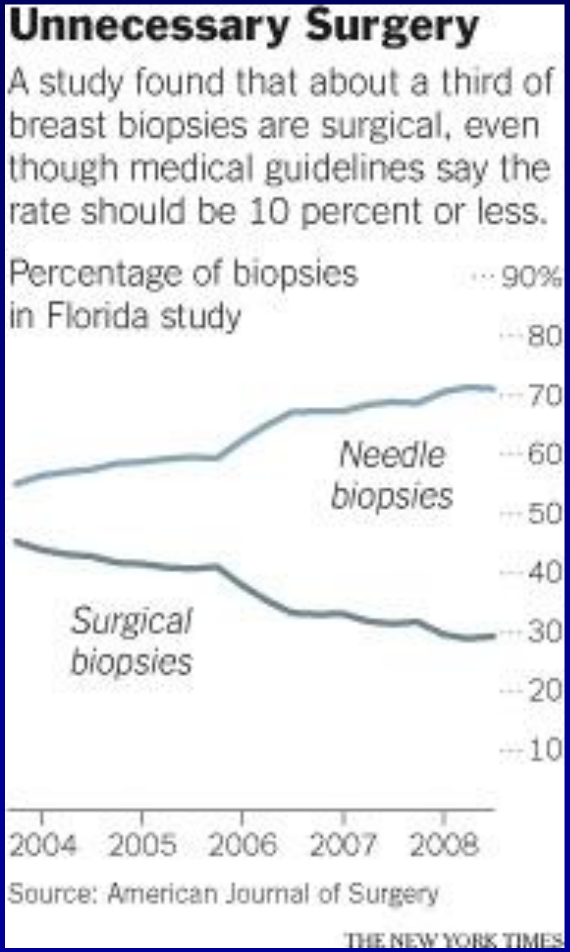
# The Breast Biopsy Project

Elizabeth Burnside, MD, MPH, MS
Heather Neuman, MD, MS
Andreas Friedl, MD
C. David Page, PhD
Jude Shavlik, PhD

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Image-guided Breast Biopsy

- **Excisional biopsy for diagnosis of findings on mammography is overutilized**

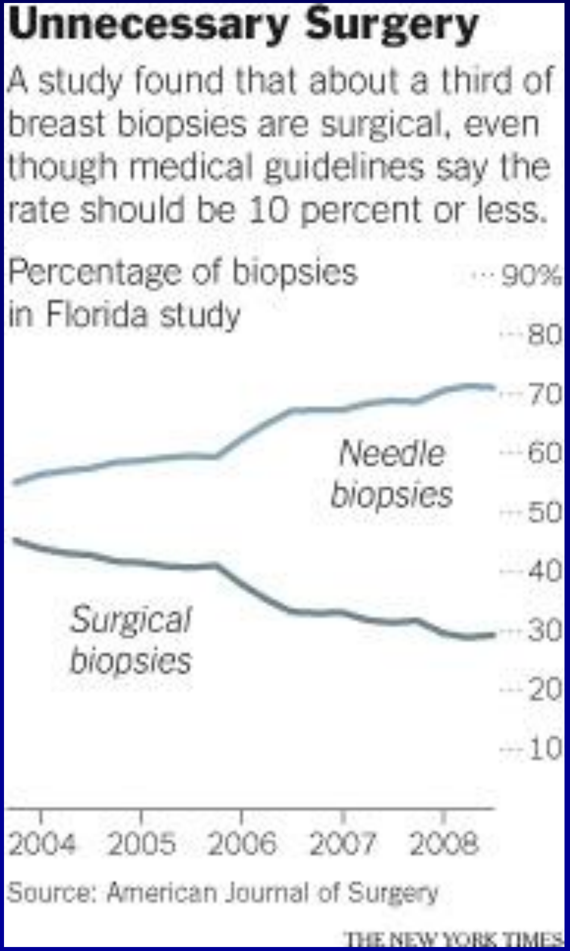



Grobmyer, SR et al. Am J Surg. 2011 Feb 2.
**Utilization of minimally invasive breast biopsy for the evaluation of suspicious breast lesions.**

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Image-guided Breast Biopsy

- **Core biopsy not perfect**
  - 10% of benign core biopsies are non-definitive
  - 10-15% of these are upgraded to cancer at excisional biopsy




Grobmyer, SR et al. Am J Surg. 2011 Feb 2.
**Utilization of minimally invasive breast biopsy for the evaluation of suspicious breast lesions.**

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Breast Biopsy

- Biopsy: single most costly component of a breast cancer screening program

- Annual breast biopsy utilization in 2010

62.6/10,000 women

➢ 700,000 women

~35,000-105,000 non-definitive

# Breast Biopsy at UW

- 5 year experience at UW
  - 1228 consecutive image-guided core biopsies
    - 890 benign
    - 94 were deemed non-definitive
    - 15 were upgraded to malignancy
- Hypothesis: ILP rules from the data and from physicians could improve the accuracy of upgrade prediction

# Biopsy data

- Example rule:

Upgrade (A) IF

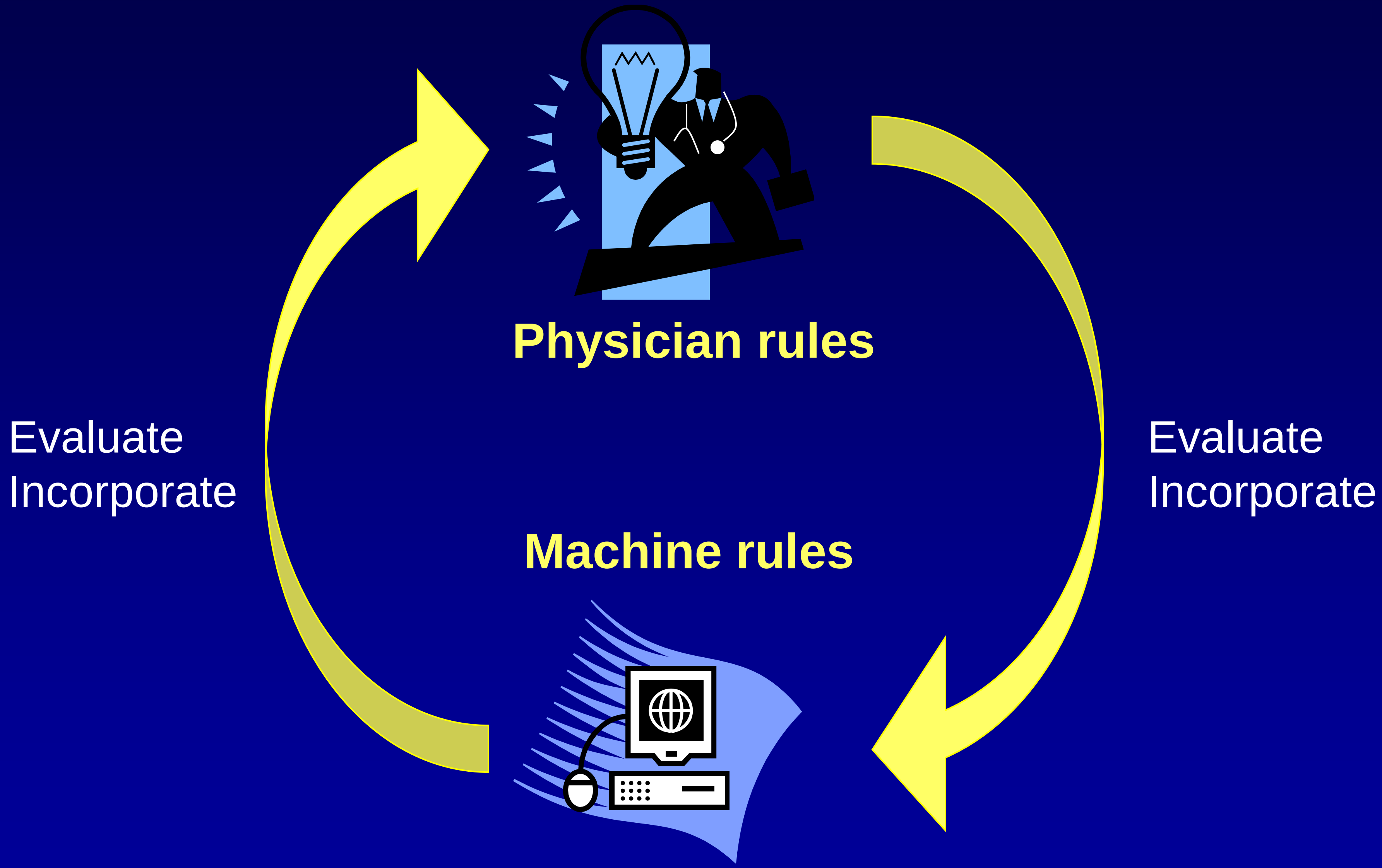
concordance (A, d),

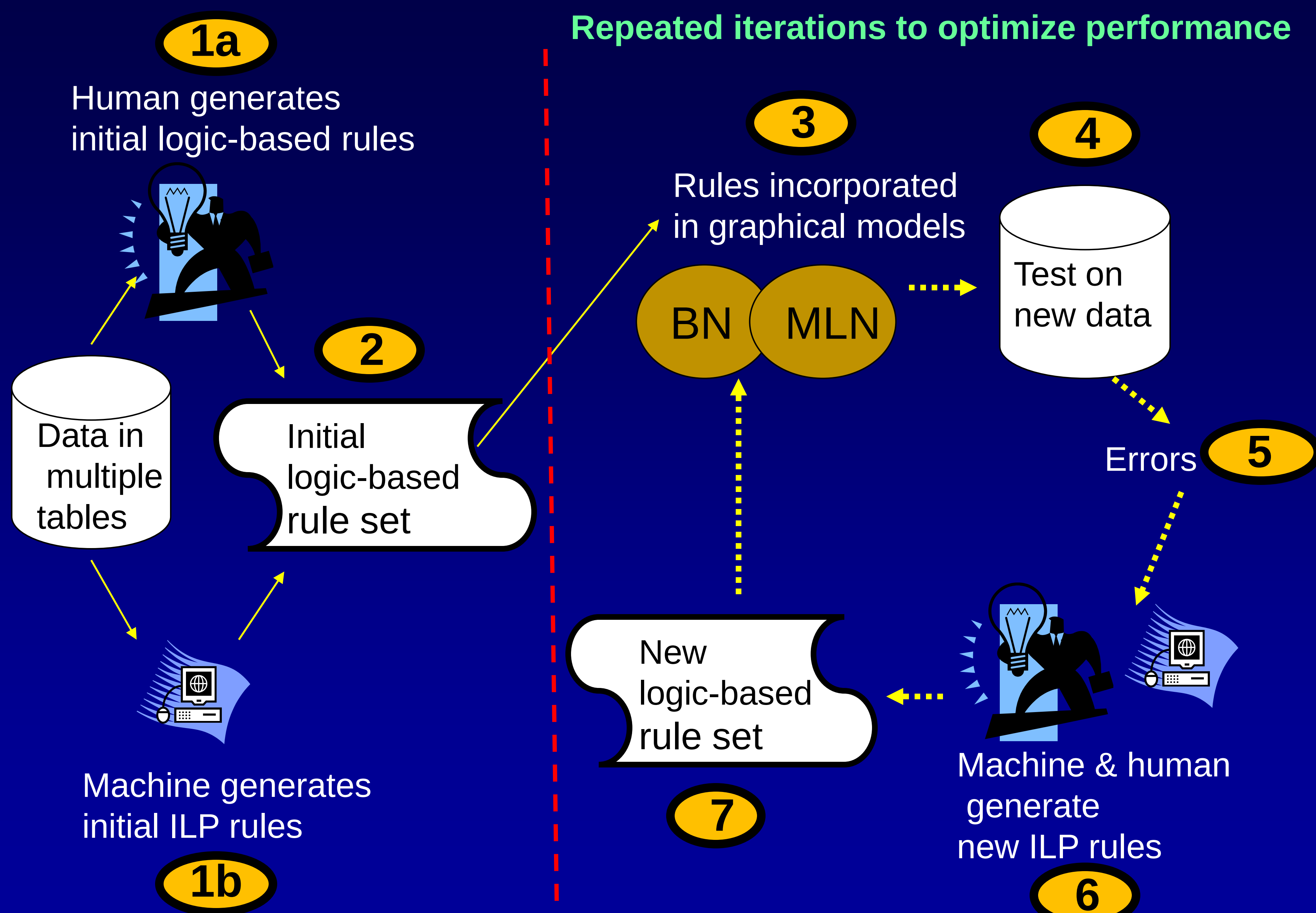
biopsyProcedure (A, US_core) and

pathDx (A, benign_breast_tissue)

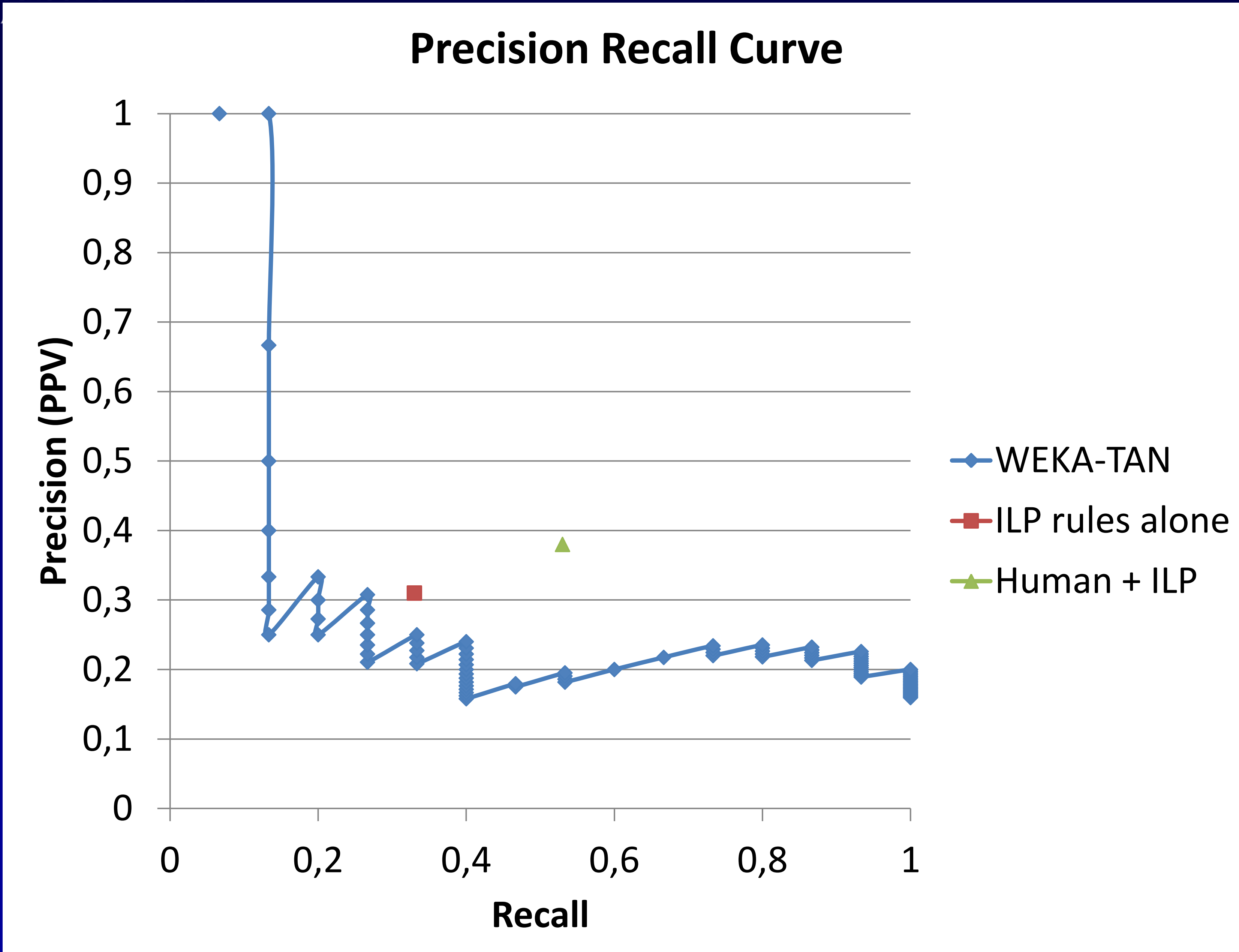
- Incorporate physician and machine rules into a Bayesian Network

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH




Repeated iterations to optimize performance
1a
Human generates
initial logic-based rules
Data in
multiple
tables
2
Initial
logic-based
rule set
1b
Machine generates
initial ILP rules
3
Rules incorporated
in graphical models
BN
MLN
4
Test on
new data
Errors
5
Machine & human
generate
new ILP rules
6
New
logic-based
rule set
7

University of Wisconsin
SCHOOL OF MEDICINE



# PPV Improvement

| | Baseline | BN with rules |
|---|---|---|
| Benign biopsy | 890 | 890 |
| Non-definitive biopsy | 94 | 75 |
| Excision avoided | 0 | 19 |
| Malignant excision | 15 | 15* |
| Benign excision | 79 | 60 |
| PPV | 16.0% | 20.0% |

***No cancers missed***

# Potential for Translation

- Translate these decision support algorithms to the clinic to improve care
- Improve evidence-based decisions
- Encourage shared decision-making

University of Wisconsin
SCHOOL OF MEDICINE
AND PUBLIC HEALTH

# Questions?


FIG. 1. Bayes' theorem in practice.