What kind of tests are these?

- Devices
  - CT scans, ultrasound, thermography, FOBT, colonoscopy, oral rinse
- Biomarkers
  - PSA, BMI, GTT, d-Dimer, pregnancy test
- Genes
  - Huntington’s, BrCA, Gene expression
- Clinical hx, signs and symptoms!

For us to care, this information must be actionable.
### Differences from explanatory or etiologic epi

Subject → Test setting → Test → Result → Action → Consequences

### We will care about....
- Who is being tested
- Where they are being tested
- Why they are being tested
- How they are being tested
- The consequences of testing....
  - to the one tested
  - to society

### We will care about...

Individual classification, not group differences

N=70  t= 4  P=0.001  OR = 2.5  AUC = 0.76

N=40  P=0.15

We will care about...

N=1000  t= 15  P<0.0001  OR = 2.7  AUC = 0.76

We will care about...

Individual classification, not group differences
We will care about...

Absolute, not relative risks.

Diagnostic study design

Courtesy of
Milo A. Puhan, MD, PhD
JHU Dept. of Epidemiology

Key messages

- Diagnostic research goes far beyond sensitivity and specificity
- Diagnostic test evaluation ranges from determination of test accuracy to impact on health outcomes and costs
- The diagnostic research question must take the context of the diagnostic work-up into consideration
- Diagnostic study design varies greatly across the phases of diagnostic test evaluation

Learning objectives

- To describe the use of diagnostic tests in practice
- To frame a diagnostic research question
- To know about the phases of diagnostic test evaluation
- To know the different diagnostic study designs
Brain natriuretic peptide (BNP) for dx'ing CHF

Cut-offs between 15 and 100 pg/ml used

Toma et al Cardiovascular Medicine 2007;10:27–33

Use of BNP in practice

<table>
<thead>
<tr>
<th>Stage of clinical management</th>
<th>Setting</th>
<th>Available information</th>
<th>Purpose of test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early diagnostic work-up</td>
<td>Primary care ER</td>
<td>Patient history Physical exam</td>
<td>Triage</td>
</tr>
<tr>
<td>Diagnosis not yet established</td>
<td>ER</td>
<td>Chest X-ray ECG</td>
<td>Add-on Replacement</td>
</tr>
<tr>
<td>Diagnosis established</td>
<td>Specialized care ER</td>
<td>echocardiography</td>
<td>Prognostic information</td>
</tr>
<tr>
<td>Under treatment</td>
<td>Primary or specialized care</td>
<td>Diagnostic work-up + treatments</td>
<td>Monitor disease process</td>
</tr>
</tbody>
</table>

BNP for triage to avoid unneeded work-up

Heart failure among differential diagnoses after taking patient history and physical exam

Positive test result: 100% Heart failure work-up ± treatment

Negative test result: 100% Consider other diagnoses

BNP as add-on test

Heart failure suggested after patient history, physical exam, ECG and Chest x-ray

Positive test result: 100% Heart failure work-up ± treatment

Negative test result: 100% Consider other diagnoses

Health outcomes
BNP used as replacement for echocardiography
Heart failure suggested after patient history, physical exam, ECG and Chest x-ray

- Positive test result: 100% Treatment
- Negative test result: 100% Consider other diagnoses
- Health outcomes:
  - 100%
  - 50%
  - 0%

BNP for monitoring
Heart failure diagnosis and treatment established

- Not in therapeutic range
  - Adapt treatment
  - In therapeutic range
    - Treatment unchanged

Diagnostic research questions
Information before testing: What is the prior probability?

- Positive test result: 100%
- Negative test result: 100%
- Heart failure work-up: a treatment
- Is decision making influenced?: 100%
- Is the health outcome influenced?: 100%
- At acceptable costs?: 100%

Framing a diagnostic research question

PICO
- Population (setting, patient characteristics, stage of diagnostic work-up)
- Index test (test of interest)
- Control (Reference) test (determines disease status)
- Outcome (test accuracy measures, management decisions, health outcomes)
Test phases for diagnostic tests

Phase I: Investigates whether test results are different for patients ± disease.

Phase II: Investigates whether patients with disease are more likely to have positive test results compared to patients without disease.

Phase III: Investigates how well the test distinguishes between patients ± disease in patients suspected of having the disease.

Phase IV: Investigates how informative a test is considering additional information available at the moment of testing.

Phase V: Investigates whether using the test leads to better health outcomes.

Phase VI: Investigates whether using the test leads to better health outcomes at acceptable costs.

Phases in the development of a diagnostic test

(From Nierenberg and Feinstein, JAMA, 1981;256:1699–1702)

<table>
<thead>
<tr>
<th>Phase</th>
<th>Case</th>
<th>Control</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Narrow, typical disease spectrum</td>
<td>Healthy</td>
<td>Performance of procedures, clinical pharmacology</td>
</tr>
<tr>
<td>II</td>
<td>Narrow, typical spectrum of disease</td>
<td>Healthy</td>
<td>Course distinctions</td>
</tr>
<tr>
<td>III</td>
<td>Expanded spectrum of disease</td>
<td>Healthy</td>
<td>Subtle distinctions</td>
</tr>
<tr>
<td>IV</td>
<td>+ appropriate co-morbidity</td>
<td>With diseases of co-morbidity needing discrimination</td>
<td>Clinically relevant test</td>
</tr>
<tr>
<td>V</td>
<td>Full spectrum</td>
<td>Full spectrum</td>
<td>Clinical trial, with clinical outcomes</td>
</tr>
</tbody>
</table>

Phases of diagnostic test evaluation

Phase I: Healthy subjects

- Patient with heart failure
  - Test positive: 90, 20, PPV: 82%, DOR: 36
  - Test negative: 10, 80, NPV: 90%, -LR: 0.13

Phase II: Patient with heart failure

- Test positive
  - Sens: 90%, Spec: 85%

- Test negative
  - Sens: 85%, Spec: 72%

Probability of heart failure?

Phase III: Patient with heart failure

Test positive: 85, 250, PPV: 25%

Test negative: 15, 650, NPV: 98%, +LR: 3.0, -LR: 0.21

Phase IV: Patients suspected of having disease

- Age, gender, smoking and coronary heart disease status known
- ± Probability of heart failure?
**Phases of diagnostic test evaluation**

**Phase V**
- Randomized trial
  - Patients suspected of having heart failure
  - Treat and follow-up
  - Health outcome
- Before-after study
  - Patients suspected of having heart failure until 2000
  - Treat and follow-up
  - Health outcome
- or
  - Patients suspected of having heart failure from 2001
  - Treat and follow-up
  - Health outcome

**Study designs for the evaluation of diagnostic tests**

<table>
<thead>
<tr>
<th>Phase</th>
<th>Study Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase I</td>
<td>Cross-sectional case-control study (pro- or retrospective)</td>
</tr>
<tr>
<td>Phase II</td>
<td>Cross-sectional case-control study (pro- or retrospective)</td>
</tr>
<tr>
<td>Phase III</td>
<td>Cross-sectional study of patients suspected of having disease (pro- or retrospective)</td>
</tr>
<tr>
<td>Phase IV</td>
<td>Cross-sectional study of patients suspected of having disease</td>
</tr>
<tr>
<td>Phase V</td>
<td>Randomized trial or before-after study</td>
</tr>
<tr>
<td>Phase IV</td>
<td>Cost effectiveness study</td>
</tr>
</tbody>
</table>

**Diagnostic test framework**

**GOOD**
- Dx test reader
- Ref. Test
- Reader
- Population
- Test
- Reference standard

**BAD**

**Biases in dx testing**

<table>
<thead>
<tr>
<th>Spectrum bias</th>
<th>Information bias</th>
<th>Measurement bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verification</td>
<td>Test-review</td>
<td>Imperfect gold-standard</td>
</tr>
<tr>
<td>Workup</td>
<td>Diagnostic review</td>
<td>Imperfect test reading</td>
</tr>
<tr>
<td>Selection</td>
<td>Incorporation</td>
<td>Context (prevalence)</td>
</tr>
<tr>
<td></td>
<td>Reading order</td>
<td>Outlier/no result</td>
</tr>
</tbody>
</table>
Is Sensitivity a Property of the Test Alone?

No!!

Is Specificity a Property of the Test Alone?

No!!

Sensitivity of a test changes as the composition of the case population changes, with different proportions of mild, moderate and severe cases.
Spectrum Bias - Specificity

Specificity of a test usually goes down when the non-diseased population includes more people with symptoms that mimic the disease.

Diagnostic test biases

- Spectrum bias
  - Spectrum of disease severity/subtype in cases varies from target population
  - Spectrum of comorbidities in controls varies from target population
- Information bias
  - Interpretation of dx test is affected by gold standard (or patient correlates), or vice versa.
- [Imperfect] Measurement bias
  - Bias introduced by imperfect dx or referent test reading.

Mathematics of Diagnostic testing

Topics

- Basic goals of diagnostic testing
  - Threshold model
  - Link to decision analysis
- Mathematics of diagnostic testing (review)
  - Bayes theorem
  - Sensitivity, Specificity, Odds, Likelihood ratio, PV
  - Tests with multiple cutpoints
    - ROC curves
    - Choosing optimal cutpoints
    - Distinction between binary vs. continuous LRs.
**Goal of a diagnostic test?**

- Not!!! .... just to distinguish between diseased and non-diseased patients.
- To improve the overall health outcomes (or reduce cost or suffering) in a group in whom it is applied.
- To do more good than harm.

**Threshold model of testing**

<table>
<thead>
<tr>
<th>Go Home</th>
<th>More testing</th>
<th>Reference standard</th>
<th>Treat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Probability of disease</td>
<td>100%</td>
</tr>
</tbody>
</table>

**Basic quantitative concepts**

- **Sensitivity**: Proportion of persons with disorder who will test positive
- **Specificity**: Proportion of persons without disorder who will test negative
- **Likelihood Ratio**: Change in disease odds from before to after test.
- **Predictive Value**: Probability of disease (positive test) or non-disease (negative test) after test.
- **Posterior probability**: Probability of disease after test.

**Sensitivity and Specificity**

![Graph showing sensitivity and specificity](image)
The 2x2 Nightmare

<table>
<thead>
<tr>
<th></th>
<th>Test Positive</th>
<th>Test Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease present</td>
<td>True Positive</td>
<td>False Negative</td>
</tr>
<tr>
<td>Disease absent</td>
<td>False Positive</td>
<td>True Negative</td>
</tr>
</tbody>
</table>

Sensitivity = True Pos / Tot. Disease
Specificity = True Neg / Tot. Well

Pos. Pred. Value = PV(+) = True Pos / Tot. Pos
Neg. Pred. Value = PV(-) = True Neg / Tot. Neg

Diagnostic calculations

<table>
<thead>
<tr>
<th></th>
<th>Test Positive</th>
<th>Test Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease present</td>
<td>95</td>
<td>5</td>
</tr>
<tr>
<td>Disease absent</td>
<td>80</td>
<td>320</td>
</tr>
</tbody>
</table>

Sensitivity = 95/100 = 95%
Specificity = 320/400 = 80%

PV(+) = 95 / 175 = 54%
PV(-) = 320/325 = 98.5%

Bayes Theorem

Pr (H₀ | Data) = Pr (H₀) × Pr (Data | H₀)
Pr (H₁ | Data) = Pr (H₁) × Pr (Data | H₁)

Post-test Odds = Pre-test Odds × Likelihood Ratio
OR
Likelihood ratio = Post-test odds / Pre-test odds

Semantic confusion

- Positive predictive value
  = Pr(Dis | T⁺) = The posterior (post-test) probability of disease after a positive test. It is the probability that you are right if you classify someone as diseased after a positive test.

- Negative predictive value
  = Pr(No Dis | T⁻) = 1 - Pr(Dis | T⁻)
  is the complement of the posterior probability of disease after a negative test. It is the probability that you are right if you classify someone as non-diseased after a negative test.

  = 1 - posterior probability after a negative test
Mathematics of Diagnostic Testing
Steven Goodman, MD, PhD

Conceptual/Computational Model for Analyzing Diagnostic Test Information

From probability to odds, and back

\[ \text{Odds} = \frac{\text{Prob}}{1 - \text{Prob}} \]

\[ \text{Prob} = \frac{\text{Odds}}{1 + \text{Odds}} \]

Bayes theorem

Odds Practice

\[ \text{Odds} = \frac{\text{Prob}}{1 - \text{Prob}} \]

p=1
\[ \text{Odds} = \frac{.1}{1 - .1} = .1/ .9 = 0.11 \]
p=.5
\[ \text{Odds} = \frac{.5}{1 - .5} = 1 \]
p=.9
\[ \text{Odds} = \frac{.9}{1} = 9 \]
p=.95
\[ \text{Odds} = \frac{.95}{.05} = 19 \]
Range of odds is 0→∞
Higher Odds

Odds = 0.1
\[ p = 0.1/(1+0.1) = 0.09 \]
Odds = 0.5
\[ p = 0.5/(1.5) = 0.33 \]
Odds = 5
\[ p = 5/6 = 0.83 \]
Odds = 24
\[ p = 24/25 = 0.96 \]

\[ \text{Prob} = \frac{\text{Odds}}{1 + \text{Odds}} \]

\[ \text{Range of } p \text{ is } 0 \rightarrow 1 \]

Likelihood Ratio Defn.

- Likelihood Ratio (LR) is the chance of observing a given result when the patient has disease, divided by the probability of observing that result if they are well.

\[ \text{LR} = \frac{p(\text{Test Result} | \text{Pt. has disease})}{p(\text{Test Result} | \text{Pt. does not have disease})} \]

Likelihood Ratio Eqns.

\[ \text{LR(Positive)} = \frac{p(\text{Positive} | \text{Pt. has disease})}{p(\text{Positive} | \text{Pt. does not have disease})} \]

\[ \text{LR(Negative)} = \frac{p(\text{Negative} | \text{Pt. has disease})}{p(\text{Negative} | \text{Pt. does not have disease})} \]

Diagnostic Tests acronyms

\[ \text{LR(Positive)} = \frac{\text{Sens}}{1-\text{Spec}} \]
\[ \text{LR(Negative)} = (1-\text{Sens})/\text{Spec} \]

- SnNout
  * A highly Sensitive test, that when negative rules out the disorder. (i.e. LR(Negative) small)

- SpPin
  * A highly specific test, that when positive rules in the disorder. (i.e. LR(Positive) large)
**Calculation of LRs**

\[
\text{LR(+) = Sens/(1-Spec)}
\]
\[
\text{LR(-) = (1-Sens)/Spec}
\]

- Sens 95% --> LR (+) = 9.5
- Spec 90% --> LR (-) = 1/18
- Sens 65% --> LR (+) = 1
- Spec 35% --> LR (-) = 1
- Sens 80% --> LR (+) = 16
- Spec 95% --> LR (-) = 1/4.8
- Sens 90% --> LR (+) = 180
- Spec 99.5% --> LR (-) = 1/10

**Value of LRs**

- The degree to which dz probability is affected by a test result.
- Tell us exactly how to combine sensitivity and spec to inform us about dx.
- Are a measure of the “strength of the evidence” of a test result for the hypothesis that the patient has dz vs. patient does not.
- Does not require a dichotomous test outcome; can be calculated for any test result

**What is a “good” LR?**

- 3-5 “Fair”
- 8-12 “Moderate”
- 12-20 “Strong”
- > 20 V. Strong

**Diagnostic calculations**

**Prevalence = 33%**, **Sensitivity = 80%**, **Specificity = 90%**

What is the probability of disease after a positive test?

- Prevalence odds = 33/66 = 0.5
- LR(+) = 80/(100-90) = 8
- Odds of disease (+) = LR(+)*x Prev. Odds
  = 8 x 0.5 = 4
- Probability of disease (+) = 4/(1+4) = 80%
Mathematics of Diagnostic Testing
Steven Goodman, MD, PhD

Diagnostic calculations

- Prevalence = 33%, Sensitivity = 80%, Specificity = 90%
- What is the probability of disease after a negative test?
  - LR(-) = 20/90 = 0.22
  - Prevalence odds = 33/67 = 0.5
  - Odds of disease (-) = LR(-) x Prev. Odds = 0.22 x 0.5 = 0.11
  - Probability of disease (-) = 0.11/1.11 = 10%

Computational Model for Analyzing Diagnostic Test Information

- Infinite Odds
- Has Disease
- LR(+) = 8
- Odds of disease after positive test = 8 x 0.5 = 4
- Odds of disease before testing = 0.5

Conceptual/Computational Model for Analyzing Diagnostic Test Information

- Infinite Odds
- Has Disease
- LR(+) = 8
- Odds of disease after positive test = 8 x 0.5 = 4
- Odds of disease before testing = 0.5

Comparative Evaluation of Immunochemical Fecal Occult Blood Tests for Colorectal Adenoma Detection

Background: Different immunochemical fecal occult blood tests (FOPR) have been proposed for responsive colorectal cancer screening. Logistic, logit, or probit regression studies that allow calculation of these tests' abilities to discriminate between patients with and without colorectal cancer.

Objective: To determine and compare performance characteristics of 4 different immunochemical FOPR for screening colorectal adenomas among adults who underwent screening colorectal examinations.

Design: Prospective screening study from January 2005 to December 2007.

Setting: 35 gastroenterology positions in Germany that did screening colonoscopies.

Participants: 1,159 participants at average risk for colorectal neoplasia who were undergoing screening colonoscopy for men aged 65 years, 55% men.

Measurements: 4 different immunochemical FOPR were done with stool samples collected before bowel preparatory for colonoscopy. Performance characteristics (sensitivity, specificity, predictive values, and likelihood ratio of tests) were measured for each individual. These 4 different FOPR were blinded to colorectal results, and colorectal cancers were divided into 3 groups.

Results: Overall, 390 participants (33%) had adenomas and 700 participants (67%) had no adenomas. The performance characteristics varied widely among tests, the 2 best-performing tests were ImmunoCard (3) and ImmunoCard (Gastro). Sensitivity was 69%, specificity was 92%, and prevalence was 33%. LR(+)=8, sensitivity was 91%, specificity was 92%, and prevalence was 33%. The 3rd best-performing test was also blind to colorectal results, colorectal cancers were divided into 3 groups.

Conclusions: Qualitative immunochemical FOPR could be an option for future colorectal cancer screening because they showed better performance characteristics than previously used programs.

Funding: The German Research Foundation (Deutsche Forschungs- gemeinschaft) and the Federal Ministry of Education and Science support the large differences in diagnostic performance among tests, while evaluations of the different test versions is important.

Annals of Internal Medicine
Volume 150, Number 15, October 2006
www.annals.org
\[ LR(+) = \frac{Sens}{1 - Spec} = \frac{35.8}{18.1} = 1.98 \]
\[ LR(-) = \frac{1 - Sens}{Spec} = \frac{64.2}{81.9} = 0.78 \]
Predictive value calculation

- **Pre-test probability** = 405/(405+914) = 30.7%
- **Prior odds** = 30.7/69.3 = 0.44
- **Post-test odds** = 0.44 x LR(+) = 0.44 * 1.98 = 0.87
- **Post-test probability** = 0.87/1.87 = 47%

- **Negative predictive value calculation**
  - **Pre-test probability** = 405/(405+914) = 30.7%
  - **Prior odds** = 30.7/69.3 = 0.44
  - **Post-test odds** = 0.44 x LR(-) = 0.44 * 0.78 = 0.34
  - **Post-test probability** = 0.34/1.34 = 26%
  - Predictive value (-) = 100-26 = 74%
Decisions about testing

Should we be testing?

**Diagnostic setting**

Sensitivity = 95%, Specificity = 95%, Prevalence = 50%

- Pre-test odds = 50/50 = 1
- LR(+) = 95/5 = 19
- Disease Odds(+) = 1x19 = 19
- PV(+)= 19/20 = 95%

- For every 19 true cases worked-up, one patient will have an unnecessary work-up.

**For this test to be worth using in this setting:**

19 x benefit (true positive) > harm (false positive)

---

**Should we test? (Cont.)**

**Screening setting: Positive test**

Sensitivity = 95%, Specificity = 95%, Prevalence = 1/500

- Pre-test odds = 1/500
- LR(+) = 95/5 = 19
- Disease Odds(+) = (1/500)x19 = 0.04
- PV(+) = 0.04/1.04 = 4%

- For every single true case worked-up, 24 patients will have an unnecessary work-up.

**For this test to be worth using in this setting:**

Benefit (true positive) > 24 x Harm (false positive)

---

**Effect of lowering prevalence**

![Graph showing normal and diseased distribution with effect of lowering prevalence indicated]
Decisions about screening

- There must be a clear benefit that results from screening, and this must outweigh the cost of testing (infrastructure, and harm to false positives and false negatives).
- Alternatives to screening/testing include:
  - Doing nothing - waiting for disease to become manifest. (e.g. PSA controversy)
  - Doing "next step" on everyone. (e.g. Strep throat)
  - Screen only subsets at higher risk.

Key Dx Test Questions

- "Phase" of study? Retrospective vs. prospective?
- Valid reference standard, done in all patients, or in a defined random sample of cases and controls.
- Description of disease spectrum, comorbidities and other clinical characteristics in all patients.
- Clear test procedures, inc. training, interpretations and reproducibility. Done blind to reference standard?
- Comparison with alternatives. Incremental value.
- Appropriate indices, w/variability.
- How is it proposed to be used, and has this use been shown to impact clinical outcomes?

Puzzler #1

- PET scanning is being assessed as a Dx test for schizophrenia.
- 20 schizophrenics have already had the test as part of a pilot feasibility study.
- 20 normal controls are chosen among general medicine clinic patients for comparison.
- Results: PET scan positive in 17/20 schizophrenics, 9/20 controls.
- Conclusion: Useful test.

Puzzler #2

- New radioisotope is being evaluated for use in diagnosing MI’s.
- Test is applied to large ER population presenting with angina.
- Gold standard is combination of ECG changes and cardiac enzyme elevation.
- Sensitivity = 96%, Specificity = 93%
- Conclusion: Useful screening test.
- Recommendation: Implement in all ERs.
Liquid Crystal Thermography as a Screening Test

**Background:** Poor correlation between clinical signs and objectively proven DVT. Venograms have significant expense and morbidity, making a universal screening test for DVT’s desirable.

**Patients:** Patients with sx or signs suggesting unilateral, lower limb, DVT were studied with a standardized clinical exam and LCT before X-ray venography.

**Diagnostic test:** Criteria have been described... “if there is a homogenous area of increased temperature in the symptomatic limb...This was done w/o knowledge of either the side of the suspected thrombus or the result of the venogram.”

<table>
<thead>
<tr>
<th>Symptom/Sign/Test</th>
<th>Venogram Result</th>
<th>Positive (n=51)</th>
<th>Negative (n=55)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain/numbness</td>
<td></td>
<td>33 (64.7%)</td>
<td>18 (32.7%)</td>
</tr>
<tr>
<td>Ankle edema</td>
<td></td>
<td>25 (75%)</td>
<td>29 (44%)</td>
</tr>
<tr>
<td>Warm to touch</td>
<td></td>
<td>28 (80%)</td>
<td>17 (35.4%)</td>
</tr>
<tr>
<td>+Thomas’s sign</td>
<td></td>
<td>21 (66.6%)</td>
<td>30 (44.5%)</td>
</tr>
<tr>
<td>Thermogram +</td>
<td></td>
<td>24</td>
<td>17 (31.8%)</td>
</tr>
</tbody>
</table>

**Sensitivity** (34/35) = 97.1%

**Specificity** (28/45) = 62.2%

**LR(+) =** 2.5

**LR(-) =** 1/21

**PV(+) =** 34/51 = 66.7%

**PV(-) =** 28/29 = 96.5%

\[ n.b., 95\% CI: 82\% to 99.9\% \]

**Conclusions:** We would propose that any patients in whom DVT is clinically suspected should undergo LCT. Anticoagulants may be withheld from those with a negative thermogram. Since a positive thermogram still has a 1/3 chance of being a false-positive, patients with a positive LCT could undergo [technetium] venoscanning, which would give a correct diagnosis in 2/3rds of them. The remainder would then undergo the definitive test, X-ray venography. This diagnostic approach...would reduce morbidity, time and cost; the chance of missing a DVT because of a false-negative thermogram on initial screening would be 1.25%.”

PET in Breast Cancer Staging


**Goal:** Find non-invasive means to dx spread of breast ca to local nodes

**Patients:** 50 women w/breast cancer in Aberdeen Royal Infirmary Breast Clinic.

**Test:** PET scan, read by two observers, blinded to clinical information and surgical staging.

**Reference test:** Surgical exploration of axilla with nodal pathology or FNA.
PET in Breast Cancer Staging

- Results of pathology:
  - Total nodes examined: 425
  - Total # of patients with involved nodes: 21

- Results of PET
  - Sensitivity: 90% (19/21) [95% CI 70% to 99%]
  - Specificity: 97% (28/29) [95% CI 82% to 99.9%]
  - PPV: 95%
  - NPV: 96%

Results of pathology:
- Total nodes examined: 425
- Total # of patients with involved nodes: 21

Results of PET:
- Sensitivity: 90% (19/21) [95% CI 70% to 99%]
- Specificity: 97% (28/29) [95% CI 82% to 99.9%]
- PPV: 95%
- NPV: 96%

PET in Breast Cancer Staging

<table>
<thead>
<tr>
<th>TNM Stage</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>100 (1/1)</td>
<td>100 (7/7)</td>
</tr>
<tr>
<td></td>
<td>(CI 2.5%-100%)</td>
<td>[82%-99%]</td>
</tr>
<tr>
<td>T2</td>
<td>78 (7/9)</td>
<td>92 (11/12)</td>
</tr>
<tr>
<td></td>
<td>[40%-97%]</td>
<td>[66%-99%+]</td>
</tr>
<tr>
<td>T3</td>
<td>100 (5/5)</td>
<td>100 (4/4)</td>
</tr>
<tr>
<td></td>
<td>[48%-100%]</td>
<td>[40% - 100%]</td>
</tr>
<tr>
<td>T4</td>
<td>100 (6/6)</td>
<td>100 (4/4)</td>
</tr>
<tr>
<td>N1 and N2</td>
<td>91 (10/11)</td>
<td>100 (3/3)</td>
</tr>
</tbody>
</table>

Claims...
- “The results from this study show that PET can accurately (94%) and reliably (PV>94%) stage the axilla…”
- “…a preponderance of patients with large or locally advanced cancer.”
- “PET...had a sensitivity and specificity of 100% in seven patients with T1N0 disease.
- “On the evidence obtained...PET could help determine treatment options in postmenopausal women....PET may have the advantage of obviating the need for additional staging investigations.”

Key Dx Test Questions
- Retrospective vs. prospective study.
- Valid reference standard, done in all patients, i.e. not affected by dx. test result.
- Description of disease spectrum, comorbidities and other clinical characteristics in all patients.
- Clear test procedures, inc. interpretations and reproducibility. Done blind to reference standard?
- Comparison with alternatives.
- Appropriate indices, w/variability.
- How is it proposed to be used, and has this use been shown to impact clinical outcomes?