How to evaluate molecular testing: How to use it in your practice?

Susan J. Mandel, MD MPH
Professor of Medicine and Radiology
Perelman School of Medicine
University of Pennsylvania
Disclosures

• No financial disclosures related to this presentation
To reflect back almost 50 years . . .

THE NODULAR THYROID GLAND AND CANCER*
A Practical Approach to the Problem

New England Journal of Medicine 1964

How to select that patient with a nodular thyroid gland in whom there is the greatest likelihood of finding a cancer is still the subject of considerable disagreement.

2013
Ultrasound
FNA
Cytology
Molecular analysis
Goal of diagnostic evaluation of thyroid nodules:

• Once we have decided that a nodule has the potential to be clinically relevant
  – FIND CANCER
How good is our DX of cancer?

- 2 expert thyroid surgical pathologists

<table>
<thead>
<tr>
<th>765 thyroid tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Appendix Table 4. Interrater Concordance of Thyroid Histopathologic Diagnosis Between the 2 Central HP Members</strong>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Benign</th>
<th>Cancer</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HP Member 1</strong></td>
<td>455</td>
<td>51</td>
<td>506</td>
</tr>
<tr>
<td><strong>HP Member 2</strong></td>
<td>23</td>
<td>236</td>
<td>259</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>478</td>
<td>287</td>
<td>765</td>
</tr>
</tbody>
</table>

- 90.3% agreement for initial diagnosis
- After conferral 98.5% agreement

Cibas Ann Intern Med 2013:159:325
When should molecular testing be considered in patients with thyroid nodules?

Never
if there is another gold standard dx test

Always
if molecular testing always provides the correct answer

Sometimes
The ACCE Model System for evaluating genetic tests

http://www.cdc.gov/genomics/gtesting/ACCE/
Analytical Validity: Accuracy and reliability to measure genotype of interest

- How often is it positive when a mutation is present?
- What range of patient specimens have been tested?
- Quality control program
- Robustness- concordance of results in multiple labs

http://www.cdc.gov/genomics/gtesting/EGAPP/recommend/method.htm
BRAF V600E mutation testing: the impact of detection sensitivity

- Percentage of detectable mutant DNA in the admixture of normal DNA

<table>
<thead>
<tr>
<th>Method</th>
<th>Detection sensitivity</th>
<th>False positive rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanger sequencing</td>
<td>20%</td>
<td>0%</td>
</tr>
<tr>
<td>Dual priming oligonucleotide-PCR</td>
<td>2%</td>
<td>1.4%</td>
</tr>
<tr>
<td>MEMO sequencing</td>
<td>0.1%</td>
<td>0.08%</td>
</tr>
</tbody>
</table>

Lee 2012 J Clin Endocrinol Metab 97:2299
And this IMPACTS clinical practice.

- 87 yo woman with 1.7cm nodule with FLUS cytology
- BRAF V600E mutation present by QUEST assay run in duplicate with detection sensitivity of 0.1%
- Total thyrx and pathology benign
- Local hospital performed Sanger sequencing (detection sensitivity 10-15%)→ no BRAF V600E mutation detected
The ACCE Model System for evaluating genetic tests

http://www.cdc.gov/genomics/gtesting/ACCE/
Clinical Validity:
Accuracy with which a test predicts the disorder of interest

- the sensitivity, specificity, and predictive values of a test in relation to a particular disorder
- evaluation of testing in a similar population to which test is targeted
- longitudinal cohort studies

http://www.cdc.gov/genomics/gtesting/EGAPP/recommend/method.htm
<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Pos</strong></td>
<td><strong>TP</strong></td>
<td><strong>FP</strong></td>
</tr>
<tr>
<td><strong>True Positive</strong></td>
<td><strong>True</strong></td>
<td><strong>False</strong></td>
</tr>
<tr>
<td><strong>Test Neg</strong></td>
<td><strong>FN</strong></td>
<td><strong>TN</strong></td>
</tr>
<tr>
<td><strong>False Negative</strong></td>
<td><strong>False</strong></td>
<td><strong>True</strong></td>
</tr>
</tbody>
</table>

**Total cancer** = \( TP + FN \)

**Total benign** = \( FP + TN \)
**Sensitivity** -- ability of test to identify disease, so the fraction of those with disease who test positive

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

**Specificity** -- ability of test to identify those without disease, so the fraction of those without disease who will test negative

$$\text{Specificity} = \frac{TN}{TN + FP}$$

<table>
<thead>
<tr>
<th>Test Pos</th>
<th>Cancer</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>TP</td>
<td>True Positive</td>
<td>FP</td>
</tr>
<tr>
<td>FN</td>
<td>False Negative</td>
<td>TN</td>
</tr>
</tbody>
</table>

**Total cancer** = TP + FN

**Total benign** = FP + TN
Sensitivity and specificity are characteristics of the test. The population does not affect the results.

- The relevant questions for the clinician and patient--
- What is the chance that a person with a positive test truly has the disease?
- What is the chance that a person with a negative test result is disease free?

Positive and negative predictive values are influenced by the prevalence of disease in the population being tested.
Test: 90% sensitivity and 90% specificity

<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Pos</strong></td>
<td>TP 18</td>
<td>FP 8</td>
</tr>
<tr>
<td><strong>Test Neg</strong></td>
<td>FN 2</td>
<td>TN 72</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>80</td>
</tr>
</tbody>
</table>

**Example 1:**
20% of population has cancer
PPV TP/all positive results = 18/26 = 69%
NPV TN/all negative results = 72/74 = 97%

<table>
<thead>
<tr>
<th></th>
<th>Cancer</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Pos</strong></td>
<td>TP 63</td>
<td>FP 3</td>
</tr>
<tr>
<td><strong>Test Neg</strong></td>
<td>FN 7</td>
<td>TN 27</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>30</td>
</tr>
</tbody>
</table>

**Example 2:**
70% of population has cancer
PPV TP/all positive results = 63/66 = 95%
NPV TN/all negative results = 27/34 = 79%
NPV Decreases as Cancer Prevalence Rises

TEST A Sens 90% Spec 50%

TEST B Sens 60% Spec 95%
TEST: 95% sensitivity
95% specificity
Evaluating molecular tests to apply to your patients:

- How robust is the definition of the studied population?
- Is the cancer risk similar?
- What is the distribution of cancer histologies associated with that cytology classification?
How robust is the definition of the population?

Indeterminate subclassifications not reliably reproducible

---

**Bethesda 6 category classification**

<table>
<thead>
<tr>
<th>FNA diagnosis</th>
<th>Kappa (95% CI)</th>
<th>Agreement*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsatisfactory/nondiagnostic</td>
<td>0.8 (0.66–0.97)</td>
<td>strong</td>
</tr>
<tr>
<td>Benign</td>
<td>0.6 (0.43–0.78)</td>
<td>fair</td>
</tr>
<tr>
<td>FLUS</td>
<td>0.3 (0.18–0.45)</td>
<td>poor</td>
</tr>
<tr>
<td>Follicular neoplasm</td>
<td>0.5 (0.34–0.65)</td>
<td>fair</td>
</tr>
<tr>
<td>Suspicious for malignancy</td>
<td>0.2 (0.06–0.37)</td>
<td>poor</td>
</tr>
<tr>
<td>Malignant</td>
<td>0.6 (0.46–0.79)</td>
<td>fair</td>
</tr>
</tbody>
</table>

INTRAobserver concordance: 68%
INTERobserver concordance: 25%
### Bethesda Classification for Thyroid Cytology

<table>
<thead>
<tr>
<th>Diagnostic Category</th>
<th>BETHESDA Malignancy risk</th>
<th>REPORTED Malignancy risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>~1-5%</td>
<td>0-32%</td>
</tr>
<tr>
<td>II</td>
<td>2-4%</td>
<td>2-7%</td>
</tr>
<tr>
<td>III</td>
<td>5-15%</td>
<td>6-48%</td>
</tr>
<tr>
<td>IV</td>
<td>15-30%</td>
<td>14-34%</td>
</tr>
<tr>
<td>V</td>
<td>60-75%</td>
<td>53-87%</td>
</tr>
<tr>
<td>VI</td>
<td>97-100%</td>
<td>96-100%</td>
</tr>
</tbody>
</table>

Baloch ZW, Diag Cytopath 36:425, 2008; Wang Thyroid 2011 21:243
What is the distribution of cancer histologies associated with that cytology classification?

“Cancer” is not one entity

- Tall cell: BRAF 70-80%
- PTC classic: BRAF 40%
- PTC FV: BRAF 10%
- Follicular cancer: RAS 30%
- RET/PTC1: 30%
- NRAS 10%
Distribution of Malignant Histology

AUS/FLUS ➔

Follicular neoplasm ➔

The ACCE Model System for evaluating genetic tests

http://www.cdc.gov/genomics/gtesting/ACCE/
Clinical Utility:
Likelihood that using the test to guide management will significantly improve health-related outcomes

- Clinical effectiveness analyses
  - Assumptions of model
  - Robustness of sensitivity analyses

- Real world application-
  - Differences in estimated magnitude of treatment effect between RCTs and observational studies

http://www.cdc.gov/genomics/gtesting/EGAPP/recommend/method.htm
And we didn’t talk about ultrasound

Two nodules with FLUS cytology

Based upon US, which nodule has the higher “pre test” cancer risk?
How to evaluate molecular testing:
Can I use it in my practice?

- Know the test
- Know the studied population
- Know your patients