Where We Are

• So far, we have introduced 2 different approaches to interpreting continuously scaled diagnostic tests
  – Selection of the optimal 2x2 table
  – SSLR
• In what situations does the 2x2 approach to interpreting diagnostic tests yield treatment decisions equivalent to those derived from the stratum-specific likelihood ratio approach? In what situations do they differ?
• We address these questions in this last module

Outline

• Treatment Decisions When No Test Is Available
  – The treatment threshold
• Two “models” of decision making
  – "One (test) and done"
  – Continuous updating
• When the treatment decisions are the same and when they differ
Treatment Decisions When No Test Is Available

• Suppose you are in a remote health center and a child aged between 3 and 36 months presents with a rectal temperature >39°C
• The child has no obvious focal infection for which timely antibiotic therapy is indicated (e.g., otitis media), nor has she received antibiotics during the preceding 48 hours
• She has no “toxic” clinical appearance necessitating immediate hospitalization, nor a specific viral infection (e.g., varicella), a known immune-deficiency condition, or chronic illness that would alter the standard approach to febrile illness (e.g., hemoglobinopathy)
• What information goes into your decision either to treat the patient empirically or to watchfully wait?

Probability of Disease

• All else equal, the higher the probability of disease, the more likely you are to treat empirically
• The lower the probability of disease, the less likely you are to treat empirically

Cost of Mistakes

• All else equal, the higher the cost of mistakenly withholding treatment compared with mistakenly treating the patient, the more likely you are to treat empirically
• The lower the cost of mistakenly withholding treatment compared with mistakenly treating the patient, the less likely you are to treat empirically
The Treatment Threshold

- Information about the probability of disease and cost of mistakes can be combined to identify a probability of disease (p*) where the expected costs of mistakenly treating and mistakenly withholding treatment are equal.
  - If our probability of disease is above p*, the cost of mistakes from withholding treatment will exceed those of mistakenly treating, and we should treat.
  - If it is below p*, the cost of mistakes from treatment will exceed those of mistakenly withholding treatment, and we should withhold treatment.

Definitions

- $OTP = \text{Value of outcome given true positive}$
- $OFN = \text{Value of outcome given false negative}$
- $OTN = \text{Value of outcome given true negative}$
- $OFP = \text{Value of outcome given false positive}$
- $OTP - OFN = C_{FN} = \text{Cost of false negative}$
- $OTN - OFP = C_{FP} = \text{Cost of false positive}$

- Expected outcome of treatment ($E_{Treat}$)
  $$E_{Treat} = pOTP + (1-p)OFP$$

- Expected outcome of no treatment ($E_{NoTreat}$)
  $$E_{NoTreat} = pOFN + (1-p)OTN$$

- Treatment threshold $p^* = (E_{Treat} = E_{NoTreat})$

Deriving the Treatment Threshold

Solve for $p^*$ such that $(E_{Treat} = E_{NoTreat})$

1. $pOTP + (1-p)OFP = pOFN + (1-p)OTN$
2. $pOTP + OFP - pOFP = pOFN + OTN - pOTN$
3. $(pOTP - pOTN) + (pOFP - pOFN) = (OTN - OTP)$
4. $p(OTP - OFN) + (OTN - OFP) = (OTN - OTP)$
5. $p(OFN + OPF) = C_{FP}$
6. $p^* = C_{FP} / (CFN + C_{FP})$
Treatment Decisions and the Threshold
\[
\frac{C_{FP}}{(C_{FN} + C_{FP})} = \text{Treatment Threshold}
\]
- At the end of any testing sequence (e.g., no tests or 1, 2, 3+ tests):
  - If the (posterior) probability is less than this threshold, watchfully wait because the expected costs of mistaken treatment exceed those of mistaken withholding of treatment
  - If the probability is greater than the threshold, treat empirically because the expected costs of mistaken withholding of treatment exceed those of mistaken treatment of the patient

Definition of the Costs of Mistakes
- The difference in the net value of treating someone correctly and the net value of treating them incorrectly
- Can be estimated by use of a cost-benefit framework (monetizing both costs and outcomes)
- Also can be estimated by use of a cost-effectiveness (NMB) framework:
  - \( C_{FN} = (W e_{FN}) + c_{FN} \)
  - \( C_{FP} = (W e_{FP}) + c_{FP} \)
  where \( W \) = willingness to pay

Cost-Effectiveness Equivalent of the Treatment Threshold
- As used here, \( C_{FN} \) and \( C_{FP} \) are being considered in a cost-benefit framework (i.e., each includes costs and the monetized value of benefits)
- Under a cost-effectiveness framework (in which \( C_{FN} \) equals a combination of \( c_{FN} \) and \( e_{FN} \) and \( C_{FP} \) equals a combination of \( c_{FP} \) and \( e_{FP} \)):
  \[
p^* = \frac{W e_{FP} + c_{FP}}{W(e_{FP} + e_{FN}) + (c_{FP} + c_{FN})}
\]
Ratio of Costs Sufficient

• As indicated previously, knowing the ratio of cost is sufficient
  – If we are able to say that the cost of false positives is 1/3 the cost of false negatives, we know that:
    \[ C_{FP} = \frac{1}{3} C_{FN} \]
  \[ \Rightarrow p^* = \frac{1}{3} C_{FN} / \left( \left( \frac{1}{3} C_{FN} \right) + C_{FN} \right) = 0.25 \]

Treatment Threshold and Costs

• If we instead have an idea of our treatment threshold, we can infer our relative valuation of \( C_{FP} \) and \( C_{FN} \)
  – If our \( p^* = 0.25 \), then
    \[ 0.25 = \frac{C_{FP}}{C_{FN} + C_{FP}} \]
    \[ 0.25 C_{FN} + 0.25 C_{FP} = C_{FP} \]
    \[ 0.25 C_{FN} = 0.75 C_{FP} \]
    \[ 1/3 C_{FN} = C_{FP} \]

• Implication: We are implicitly making assumptions about \( C_{FP} \) and \( C_{FN} \) whenever we make a treatment decision

Treatment Threshold and The Optimal 2x2 Table

• The treatment threshold is linked to the definition of the optimal 2x2 table (and thus to the definition of positive and negative tests)
  – In the optimal 2x2 table, negative findings include those test results that yield post-test probabilities that leave us below the treatment threshold
  – Positive findings include those test results that yield post-test probabilities that leave us above the threshold
Two “Models” of Decision Making

• "One (Test) and Done"
• Continuous Updating -- "World Enough and Time" -- decision making

* Andrew Marvell, “To his coy mistress”

One (Test) and Done Decision Making

• Characterized by the need to make a treatment decision quickly without a large number of opportunities to collect additional data
• The primary concern for this type of decision making is not the calculation of the exact post-test probability of the outcome, but instead the determination of whether the post-test probability is high enough to initiate treatment (i.e., above the treatment threshold)

One (Test) and Done, 2x2 Tables, and SSLR

• In “One (Test) and Done” decision making, use of the optimal 2x2 table and SSLR yield identical treatment decisions, because in the optimal 2x2 table:
  – All SSLR that yield post-test probabilities above the treatment threshold will be classified as positive tests;
  – All SSLR that yield post-test probabilities below the treatment threshold will be classified as negative tests
Demonstration

- WBC SSLR for bacteremia

<table>
<thead>
<tr>
<th>Cut-off</th>
<th>SSLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥25</td>
<td>7.624</td>
</tr>
<tr>
<td>≥20, &lt;25</td>
<td>3.073</td>
</tr>
<tr>
<td>≥15, &lt;20</td>
<td>1.793</td>
</tr>
<tr>
<td>≥10, &lt;15</td>
<td>0.7920</td>
</tr>
<tr>
<td>≥0, &lt;10</td>
<td>0.1791</td>
</tr>
</tbody>
</table>

- Assume that \( C_{FP} = C_{FN} \)
  \[ p^* = 0.5 \]

Pre-Test Probabilities Below the Threshold

- Resulting post-test probabilities:

<table>
<thead>
<tr>
<th>Probability of Disease</th>
<th>0.00</th>
<th>0.25</th>
<th>0.50</th>
<th>0.75</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ C_{FP} = C_{FN}; \frac{C_{FP}}{C_{FP} + C_{FN}} = 0.5 = \text{Threshold} \]

Likelihood ratios defined for WBC for bacteremia

Strata yielding post-test probabilities > the threshold

<table>
<thead>
<tr>
<th>Pre-test</th>
<th>Strata</th>
<th>OOS *</th>
<th>Tangency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>≥ 20</td>
<td>2.333</td>
<td>～ 20</td>
</tr>
<tr>
<td>0.358</td>
<td>≥ 15 or ≥ 20</td>
<td>1.793</td>
<td>～ 15 or ≥ 20</td>
</tr>
<tr>
<td>0.4</td>
<td>≥ 15</td>
<td>1.500</td>
<td>～ 15</td>
</tr>
</tbody>
</table>

* Because we assumed that \( C_{FP} = C_{FN} \), the costs cancel out of the OOS
Pre-Test Probabilities Above the Threshold

• Resulting post-test probabilities:

\[ \begin{array}{c|ccccc}
\text{Probability of Disease} & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \\
\hline
\text{Negative Test} & \text{Prior = 0.60} & \text{Prior = 0.51} & \text{Prior = 0.558} & \\
\text{Positive Test} & \text{20 - 25} & \geq 25 & \\
\end{array} \]

\[ C_{fp} = C_{fn}; \frac{C_{fp}}{C_{fp} + C_{fn}} = 0.5 = \text{Threshold} \]

Likelihood ratios defined for WBC for bacteremia

Optimal 2x2 Tables

<table>
<thead>
<tr>
<th>Strata yielding post-test probabilities ≥ the threshold</th>
<th>Optimal 2x2 Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Strata</td>
</tr>
<tr>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>0.51</td>
<td>≥ 15</td>
</tr>
<tr>
<td>0.558</td>
<td>≥ 10 or ≥ 15</td>
</tr>
<tr>
<td>0.60</td>
<td>≥ 10</td>
</tr>
</tbody>
</table>

* Because we assumed that CFP = CFN, the costs cancel out of the OOS

Classification of WBC Counts Between 20 and 25

\[ \begin{array}{c|ccccc}
\text{Probability of Disease} & 0.00 & 0.25 & 0.50 & 0.75 & 1.00 \\
\hline
\text{Negative Test} & \text{Prior = 0.175} & \text{Prior = 0.2455} & \text{Prior = 0.325} & \\
\text{Positive Test} & \text{Posterior = 0.395} & \text{Posterior = 0.50} & \text{Posterior = 0.597} & \\
\text{OOS} & 4.714 & 3.073 & 2.077 & \\
\text{Neg Test} & \text{Post of Neg Test} & \text{Positive Test} & \\
\end{array} \]

\[ C_{fp} = C_{fn}; \frac{C_{fp}}{C_{fp} + C_{fn}} = 0.5 = \text{Threshold} \]

Likelihood ratios defined for WBC for bacteremia
Classification of the 20-25 Stratum

<table>
<thead>
<tr>
<th>Probabilities</th>
<th>Classification of Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>Post-test</td>
</tr>
<tr>
<td>0.174</td>
<td>0.395</td>
</tr>
<tr>
<td>0.2455</td>
<td>0.50</td>
</tr>
<tr>
<td>0.325</td>
<td>0.597</td>
</tr>
</tbody>
</table>

* Because we assumed that CFP = CFN, the costs cancel out of the OOS

"Proof"

- Is it possible for 1) a stratum to be classified as a negative test if 2) the post-test probability resulting from its stratum specific likelihood ratio to be above the treatment threshold?

1) We classify strata as negative test results by comparing the SSLR to the OOS; they are classified as negative results if:

\[ \text{SSLR}_i \times \text{OOS} = \frac{(1-p)CFP}{pCFN} \]

"Proof" (2)

2) The post-test probability resulting from an SSLR is above the treatment threshold if:

\[ \text{Post-Test Prob} = \frac{\text{SSLR}_i p}{(\text{SSLR}_i p) + (1-p)} > \frac{CFP}{CFP+CFN} \]

a) Multiply through by the denominators:

\[ (\text{SSLR}_i p CFP) + (\text{SSLR}_i p CFN) > (\text{SSLR}_i p CFP) + ((1-p) CFP) \]

b) Cancel SSLR_i p CFP:

\[ (\text{SSLR}_i p CFN) > ((1-p) CFP) \]

c) Divide through by p CFN:

\[ \text{SSLR}_i > \frac{(1-p)CFP}{pCFN} \]
“Proof” (3)

- Classification as a negative test implies:
  \[ \text{SSLR}_i < \frac{(1-p)C_{FP}}{pC_{FN}} \]

- Having a post-test probability above the threshold implies:
  \[ \text{SSLR}_i > \frac{(1-p)C_{FP}}{pC_{FN}} \]

- We can develop an analogous set of equations that show a similar contradiction for positive strata and post-test probabilities below the treatment threshold

Continuous Updating Decision Making

- Patient has a problem that eventually needs to be treated
- Have time to perform a large number of tests
- Continuously update our probabilities after each round of testing
- At the end of the “Continuous Updating” testing process, we still use the same decision criterion as One and Done decision making: \( p \leq \frac{C_{FP}}{C_{FN} + C_{FP}} \)

Continuous Updating Decision Making (2)

- Continuous Updating decision making is not simply concerned about \( >p^* \) / \( <p^* \)
- Rather, the magnitude of the difference between our post-test probability and \( p^* \) is important
- Because distance from \( p^* \) matters, we don’t want to combine all “positive” test results and all “negative” test results, because doing so dilutes the value of the test’s information
Post-Test Probabilities Differ

- Use of the optimal 2x2 table generally will yield different post-test probabilities than use of SSLR
- The post-test probabilities from the SSLR will generally be more discriminating than those from the optimal 2x2 table

Illustration of Differences

- Illustrate these issues by use of data for a second test for bacteremia: interleukin-6 (IL-6)

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>D+</th>
<th>D-</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥10^3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>≥10^2, &lt;10^3</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>5</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>46</td>
</tr>
</tbody>
</table>


LR+ and LR- for the 4 2x2 Tables

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>D+</th>
<th>D-</th>
<th>LR+</th>
<th>LR-</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥10^3</td>
<td>0</td>
<td>0</td>
<td>--</td>
<td>1.000</td>
</tr>
<tr>
<td>≥10^2</td>
<td>4</td>
<td>2</td>
<td>4.182</td>
<td>0.855</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>17</td>
<td>11</td>
<td>3.231</td>
<td>0.299</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>22</td>
<td>46</td>
<td>1.000</td>
<td>--</td>
</tr>
</tbody>
</table>
### SSLR for the 3 IL-6 Strata

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>D+</th>
<th>D-</th>
<th>SSLR</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;10^3</td>
<td>4</td>
<td>2</td>
<td>4.182</td>
</tr>
<tr>
<td>10^2 - 10^3</td>
<td>13</td>
<td>9</td>
<td>3.020</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>5</td>
<td>35</td>
<td>0.299</td>
</tr>
</tbody>
</table>

### Differing Post-Test Probabilities, Example 1

- Suppose the optimal 2x2 table combines 10^2-10^3 and >10^3 strata into a positive test (e.g., an OOS of 2.5)
- Suppose the test result is >10^3:
  - SSLR, >10^3: 4.182
  - LR+, >10^2: 3.231
- Suppose the test result is 10^2 - 10^3:
  - SSLR, 10^2 - 10^3: 3.020
  - LR+, >10^2: 3.231
- For all 3 LR and SSLR, the post-test probability increases
- For >10^3, the 2x2 approach yields too little an increase in posterior probability, while for >10^2, it yields too great an increase

### Differing Post-Test Probabilities, Example 2

- Suppose the optimal 2x2 table combines 10^2-10^3 and <10^2 strata into a negative test (e.g., an OOS of 3.5)
- Suppose the test result is <10^3:
  - SSLR, <10^3: 0.299
  - LR-, <10^3: 0.855
- Suppose the test result is 10^2 - 10^3:
  - SSLR, 10^2 - 10^3: 3.020
  - LR-, <10^3: 0.855
- For <10^3, the posterior probabilities shift in the correct direction, although not by enough
- For test results 10^2 - 10^3, the 2x2 approach shifts the posterior in the wrong direction (use of an LR of 0.855 rather than 3.020)
Differing Post-Test Probabilities But One Test

- If we base our treatment decision on this one test alone, the fact that the post-test probabilities differ, or that they move in different directions has no effect on the treatment decision
  - In the case where all the post-test probabilities are greater than the pre-test probability, all of the results leave us above the treatment threshold
  - In the case where some of the post-test probabilities are greater than the pre-test probability and some are less than this probability, all the results leave us below the treatment threshold

Chaining LRs

- These variations in the difference between the pre-test and post-test probabilities can lead to different treatment decisions only when we use multiple individual tests
- Suppose we use 2 tests for bacteremia, IL-6 and absolute neutrophil count (ANC) and stratify them as follows:

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>D+</th>
<th>D-</th>
<th>ANC Level</th>
<th>D+</th>
<th>D-</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥10^3</td>
<td>4</td>
<td>2</td>
<td>≥15K</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>≥10^3, &lt;10^3</td>
<td>13</td>
<td>9</td>
<td>&gt;10K, &lt;15K</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>5</td>
<td>35</td>
<td>≤10K</td>
<td>10</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td>22</td>
<td>46</td>
<td>Total</td>
<td>22</td>
<td>46</td>
</tr>
</tbody>
</table>

SSLR_i

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>SSLR_i</th>
<th>ANC Level</th>
<th>SSLR_i</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥10^3</td>
<td>4.182</td>
<td>≥15K</td>
<td>7.318</td>
</tr>
<tr>
<td>10^2 - 10^3</td>
<td>3.020</td>
<td>&gt;10K, &lt;15K</td>
<td>1.742</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>0.299</td>
<td>≤10K</td>
<td>0.550</td>
</tr>
</tbody>
</table>
Chained Post-Test Probabilities

- At least 2 methods exist for calculating the post-test probability
- Start with your prior; use the SSLR for one of the tests to calculate a posterior; use this posterior as your new prior, and use the SSLR from the second test to calculate a posterior
- Equivalently, multiply the likelihood ratios from the 2 test together and use this “combined” likelihood ratio to calculate a posterior
- Both methods assume that the results of the 2 tests are independent of one another

Independence

- Mistakes from one test occur equiproportionally among the other test’s strata
- If 59% of those in whom disease is present have an IL-6 between $10^2$ and $10^3$, then:
  - Among those in whom disease is present and ANC is $>15K$, 59% have an IL-6 between $10^2$ and $10^3$
  - Among those in whom disease is present and ANC is $>10K$ and $<15K$, 59% have an IL-6 between $10^2$ and $10^3$
  - Among those in whom disease is present and ANC is $<10K$, 59% have an IL-6 between $10^2$ and $10^3$

Expected and Observed Combined SSLR

<table>
<thead>
<tr>
<th>Expected (SSLR_{IL6} * SSLR_{ANC})</th>
<th>ANC Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>IL-6</td>
<td>ANC Levels</td>
</tr>
<tr>
<td>$&gt;10^3$</td>
<td>$&gt;15K$</td>
</tr>
<tr>
<td>$2.091$</td>
<td>0</td>
</tr>
<tr>
<td>$10^2 - 10^3$</td>
<td>10.45</td>
</tr>
<tr>
<td>$&lt;10^2$</td>
<td>$\infty$</td>
</tr>
</tbody>
</table>
Violation of the Independence Assumption

- This assumption is violated for ANC<10,000 (p=0.000)

<table>
<thead>
<tr>
<th>IL-6 Level</th>
<th>ANC&lt;10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>D+</td>
<td>D-</td>
</tr>
<tr>
<td>&gt;10^3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>75%</td>
</tr>
<tr>
<td>10^2 - 10^3</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>56%</td>
</tr>
<tr>
<td>&lt;10^2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6%</td>
</tr>
</tbody>
</table>

- p = 0.58 and 0.82 for ANC ≥ 15,000 and ANC between 10,000 and 15,000, respectively

Use of More Than One Test

- If we base our treatment decision on more than one test and use either multiple 2x2 tables or multiple SSLR, the estimated post-test probabilities will differ
  - If the independence assumption holds, the SSLR approach will yield more discriminating posterior probabilities
  - If the independence assumption does not hold, it is better to treat the 2 tests as 1
    - Empirically derive sensitivities and specificities or LR for combinations of test results (see "observed" Table, slide 42)
    - By doing so, we return to One (Test) and Done decision making, and the 2x2 and SSLR approaches will yield the same treatment decisions

Which Decision Mode, When?

- When are medical decisions more like One and Done and when are they more like Continuous Updating?
  - One and Done
    - Life and death
    - Only one chance to affect outcome (homeless, AIDS patients, etc.)
    - Concern about patient lack of follow-up
    - High degree of correlation among tests
Which Decision Mode, When? (2)

- Continuous Updating
  - Not life and death
  - Good continuity of care
  - Patient likely to follow-through with moderately complex diagnostic/treatment plan
  - Test results are uncorrelated

Role of Updating of Probabilities in Routine Clinical Practice

- Evidence-based diagnostic test types routinely suggest that healthcare professionals should become probabilists and use the math we've described in this lecture for complex medical decision making
- Ironically, the evidence base for this recommendation is thin or nonexistent (few if any trials)
- After 1) teaching this material for many years, 2) seeing how innumerate lots of good clinicians are, and 3) teaching clinicians how to estimate p*, I tend to see medical education as teaching pattern recognition rather than probability updating

"Advanced" Technical Issues

- The simple version of the math that we teach you is appropriate for decision making when there are two options (e.g., disease or no disease)
  - It becomes much more complicated when there are more than 2 options
- For complicated decision making, the simple math tends to assume that the test characteristics are independent of one another
  - Not much evidence exists on independence or lack thereof of different test’s results
Important For Researchers Developing or Evaluating Tests

• The fact that clinicians may be pattern recognizers and not probabilists does not mean that when we are developing and evaluating diagnostic tests that we can ignore the principals we’ve discussed in these lectures
• These are the tools we use to identify the patterns that clinicians should look for and are essential to determining when tests should be used and when they should not be