Perils and Pitfalls in Clinical Trials of Diagnostic Tests for Tuberculosis

Richard O’Brien, MD
Foundation for Innovative New Diagnostics
Geneva
Outline of Presentation

- Statement of the problem
- Common errors in diagnostic clinical trials
- Suggested approaches
  - Trials of new case detection tools
  - Trials of new drug susceptibility testing methods
  - Trials of tests for latent TB infection
TB Diagnostics: Trial Design and Regulatory Perspective

• No internationally accepted standards
Application of Methodological Standards in Evaluation of Diagnostic Tests*

- Specify spectrum of evaluated patients (27%)
- Report subgroup results (8%)
- Avoid workup (verification) bias (46%)
- Avoid review bias (38%)
- Provide numeric precision of test indices (11%)
- Report indeterminate results (22%)
- Specify test reproducibility (23%)

TB Diagnostics: Trial Design and Regulatory Perspective

• No internationally accepted standards
• Diagnostics not regulated in most countries
TB Diagnostics: Trial Design and Regulatory Perspective

• No internationally accepted standards
• Diagnostics not regulated in most countries
• Standards needed
  – To promote improved clinical trials
  – To assist regulatory authority
  – To inform technical guidelines (e.g., WHO)
  – To guide potential purchasers
Standards for Reporting of Diagnostic Accuracy (STARD) Initiative (1)*

• Introduction
  – Identify study as one of diagnostic accuracy
  – State research questions/study aims

• Methodology (1)
  – Participants
    • Study population
    • Participant recruitment
    • Participant sampling
    • Data Collection

Standards for Reporting of Diagnostic Accuracy (STARD) Initiative (2)

• Methodology (2)
  – Test methods
    • Choice of reference standard and rationale
    • Technical aspects of material and methods
    • Define rationale for units, cut/offs, categories
    • Number, training, expertise of staff
    • Blinded reading of test and reference standard
  – Statistical methods
    • Methods for calculating test accuracy
    • Methods for calculation test reproducibility
Standards for Reporting of Diagnostic Accuracy (STARD) Initiative (3)

• Results
  – Participants
    • When was study done
    • Clinical and demographic characteristics
    • Number eligible that were not enrolled
  – Test results
    • Time from test to reference standard (and Rx)
    • Distribution of severity of disease (and non-disease)
    • Cross tabulation of results of test and reference std.
    • Any adverse events
Standards for Reporting of Diagnostic Accuracy (STARD) Initiative (4)

- Results (2)
  - Estimates
    - Diagnostic accuracy and statistical uncertainty
    - How indeterminate results handled
    - Estimates of variability of accuracy in subgroups
    - Estimates of test reproducibility

- Discussion
Errors in TB Diagnostic Trials (1)

• Failure to assess test in correct population
  – Overestimating test sensitivity
    • Non-hospitalized vs hospitalized patients
    • AFB+ vs paucibacillary TB
    • HIV- vs HIV+ patients
### Analysis of SeroDxTic Tests Among 465 Hospitalized Patients in Botswana, 2002

<table>
<thead>
<tr>
<th>Test</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICS (PATH)</td>
<td>27</td>
<td>75</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>Osborn Sci</td>
<td>37</td>
<td>63</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>MycoDot</td>
<td>3</td>
<td>99</td>
<td>56</td>
<td>65</td>
</tr>
<tr>
<td>Amer Biologics</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>Anda Bio (Integ)</td>
<td>63</td>
<td>39</td>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>Omega DxTics</td>
<td>43</td>
<td>52</td>
<td>33</td>
<td>63</td>
</tr>
</tbody>
</table>

Errors in TB Diagnostic Trials (2)

- Failure to assess test in correct population
  - Overestimating test sensitivity
    - Non-hospitalized vs hospitalized patients
    - AFB+ vs paucibacillary TB
    - HIV- vs HIV+ patients
  - Overestimating test specificity
    - Inappropriate control group (healthy controls)
MPB-64 patch test
TB patients vs healthy controls

<table>
<thead>
<tr>
<th></th>
<th>Pos</th>
<th>Neg</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TB Patients</strong></td>
<td>52</td>
<td>1</td>
<td>53</td>
</tr>
<tr>
<td><strong>Healthy controls</strong></td>
<td>0</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>52</td>
<td>44</td>
<td>96</td>
</tr>
</tbody>
</table>

Analysis of SeroDxTic Tests Among 465 Hospitalized Patients in Botswana, 2002

<table>
<thead>
<tr>
<th>Test</th>
<th>Sens</th>
<th>Spec</th>
<th>PPV</th>
<th>NPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICS (PATH)</td>
<td>27</td>
<td>75</td>
<td>39</td>
<td>63</td>
</tr>
<tr>
<td>Osborn Sci</td>
<td>37</td>
<td>63</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>MycoDot</td>
<td>3</td>
<td>99</td>
<td>56</td>
<td>65</td>
</tr>
<tr>
<td>Amer Biologics</td>
<td>0</td>
<td>99</td>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>Anda Bio (Integ)</td>
<td>63</td>
<td>39</td>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>Omega DxDxTics</td>
<td>43</td>
<td>52</td>
<td>33</td>
<td>63</td>
</tr>
</tbody>
</table>

Errors in TB Diagnostic Trials (3)

• Inadequate blinding
  – Patients and controls identified
  – Results of gold-standard test known
• Improper resolution of discrepant results
• Failure to consider indeterminate results
• Too small a sample size
FASTPlaque RIF Response Test
Direct Test on AFB+ Sputa

• Sensitivity = 100% (11/11)
• Specificity = 100% (134/134)
• Results reported in 2 days
• Conventional method took a mean ± SD of 33.2 ± 7.2 days

**FASTPlaque RIF Response Test**

<table>
<thead>
<tr>
<th>FASTPlaque</th>
<th>Resistant</th>
<th>Susceptible</th>
<th>MGIT Culture-negative</th>
<th>Contam**</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistant</td>
<td>10</td>
<td>1#</td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Susceptible</td>
<td>0</td>
<td>134</td>
<td>1</td>
<td>14</td>
<td>149</td>
</tr>
<tr>
<td>RIF- &lt;100 plaques*</td>
<td>4</td>
<td>16</td>
<td>5</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>Contaminated</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>155</td>
<td>6</td>
<td>20</td>
<td>195</td>
</tr>
</tbody>
</table>

* less than 100 plaques obtained on the RIF- plate  
** contaminated on either MGIT culture or 7H11 susceptibility test  
# this specimen was found to be rifampicin resistant upon repeat testing by the proportion method
Approach to Trials of New Case Detection Tools

• Evaluation in TB suspects (HIV+/-)
• Multiple trial sites
• Gold standard
  – Culture on solid and liquid media
• Rigorous assessment of all patients for proper assignment
  – Include “clinical” cases
  – Follow-up of patients not diagnosed with TB
TB Case Definitions

• **AFB+:** $\geq 2$ AFB+ sputum smears or 1 AFB+ smear and compatible CXR

• **AFB-/culture+:** $\geq 1$ culture with $\geq 1+$ growth

• **AFB-/culture-:** $\geq 3$ negative smears/cultures, compatible signs/symptoms of TB, compatible CXR, +TST, clinical/CXR response to treatment

• **Non-TB case:** negative bacteriology, stable CXR and/or resolution without TB treatment and/or alternate diagnosis
TB suspect, age > 14 years
no previous TB DX/RX
consent for venipuncture

- Did not complete assessment (n=24)
- Venipuncture, MycoDot™ Test, Mantoux Test, AFB X 3
  - AFB-negative (n=146)
    - Erythromycin orally for 10 days
      - Two-week assessment
        - Symptomatic response (n=66)
        - No response (n=80)
          - Repeat AFB X 3
            - AFB-negative (n=80)
              - Chest x-ray
                - Normal (n=42)
                - Abnormal (n=38)
          - AFB-positive (n=0)
            - Asymptomatic (n=102)
            - Symptomatic (n=6)
              - Reassess for TB
                - Smear(-)/Culture(+), Treatment
                  - Smear(-)/Culture(-), Treatment
                    - Normal (n=42)
                    - Abnormal (n=38)

- Treated for TB but
did not meet case
definition (n=32)
  - TB Case
    - Treatment

- Erythromycin orally for 10 days
  - Two-week assessment
    - Symptomatic response (n=66)
    - No response (n=80)
      - Repeat AFB X 3
        - AFB-negative (n=80)
          - Chest x-ray
            - Normal (n=42)
            - Abnormal (n=38)
          - AFB-positive (n=0)
            - Asymptomatic (n=102)
            - Symptomatic (n=6)
              - Reassess for TB
                - Smear(-)/Culture(+), Treatment
                  - Smear(-)/Culture(-), Treatment
                    - Normal (n=42)
                    - Abnormal (n=38)

Approach to Trials of Drug Susceptibility Testing Methods

- Study in population at risk of DR, e.g., failures of Category I treatment
- FDA gold standard: proportional method on 7H10 solid medium
- Resolution of discrepant results, e.g., rpoB gene sequencing
- Economic analyses important
FIND/TDR Study of Rapid DST Methods in Peru

• To enroll AFB+ patients (1000 new, 250 retreatment): expected 150 MDR TB cases
• Test methods: direct FASTPlaque RIF Response, INNO LiPA, and LJ; indirect MTT
• Standard method: indirect LJ proportional
• All isolates with any RIF-R plus 10% of others: rpoB sequencing, DST on 7H10
Approach to Trials of Tests for Latent TB Infection (1)

• Calculation of test sensitivity and specificity problematic
  – New/previoulsy treated TB patients
  – TB contacts with defined exposure risk
  – Persons at low risk for LTBI

• Gold standard (PPD TST) inadequate
  – Sensitivity measured in TB patients, poor specificity
Study of QuantiFERON-2 and Tuberculin Skin Test in TB Contacts*

- High school outbreak in Denmark
- 125 contacts evaluated
  - BCG-unvaccinated (85)
    - Excellent agreement between QFT and TST in both high (93%) and low (95%) groups
  - BCG-vaccinated (40)
    - QFT positivity in high (53%) and low (6%) groups similar to that seen among non-vaccinated groups

Evaluation of ELISPOT TB Assay In TB Contacts in School Outbreak*

• 535 students in a school outbreak in UK
• TST (Heaf test) and ELISPOT TB assay
• Four exposure groups defined
• Results of ELISPOT more significantly correlated with exposure than TST
  – OR 2.78 (2.22-3.48) vs 2.33 (1.88-2.88), p = 0.03

Approach to Trials of Tests for Latent TB Infection (2)

• Calculation of test sensitivity and specificity problematic
  – New/previously treated TB patients
  – TB contacts with defined exposure risk
  – Persons at low risk for LTBI

• Gold standard (PPD TST) inadequate
  – Sensitivity measured in TB patients, poor specificity

• Large cohort studies needed to demonstrate relationship between test result and future TB disease
Immune Response to ESAT-6 in TB Contacts*

Dark squares = ESAT-6, open circles = PPD