Mapping the landscape and quality of TB diagnostic research

Madhukar Pai, MD, PhD
Laurence Brunet, MSc
Jessica Minion, MD
Karen Steingart, MD, MPH
Andrew Ramsay, MSc
Christian Lienhardt, MD, PhD

Contact: madhukar.pai@mcgill.ca

Rationale

- Research in TB diagnostics is an active field, but there has been no systematic mapping of existing TB diagnostic research
- While concerns have been expressed about poor quality of TB diagnostic studies, this has not been formally assessed
Goals of this project by STP RM & NDWG

- Map the landscape of current TB diagnostic research
  - What % of TB research is focused on diagnosis?
  - Where is the research output from?
  - What tests are being evaluated?
  - What outcomes are commonly reported?

- Assess the quality of TB diagnostic accuracy studies
  - Methodological quality of TB diagnostic accuracy studies
  - Quality of reporting

Methods

- Map the landscape of current TB diagnostic research
  - Bibliometric analysis of citations
  
  - PubMed and EMBASE were searched by a librarian for all original TB citations in a two year period – 2007–2008
    - For EMBASE, the search strategy was: exp *Mycobacterium Tuberculosis/ or exp *Tuberculosis or exp Tuberculosis/di [Diagnosis] or tuberculosis.m_titl. limit to yr=“2007 – 2008” not (book or book series or editorial or letter or “review”)
Methods

- Map the landscape of current TB diagnostic research
  - All the citations (titles and abstracts) were read and coded by a trained researcher after pilot testing and standardization
  - A second reviewer coded a subset of the citations
  - UK Clinical Research Collaboration’s Health Research Classification System (HRCS) was used to retrieve details on the type of research of each study.
  - Additional information was collected for the diagnosis studies on: study design and type of outcome reported, purpose of the test, technology platform, study participants, study population, reporting of HIV status, use of commercial vs. in-house test, country where study was done, etc.
Methods

- Assess the quality of TB diagnostic accuracy studies
  - We used QUADAS and STARD checklists to assess the methodological and reporting quality of TB diagnostic studies published in a two year period
  - We also used several diagnostic meta-analyses to assess quality of the included studies in these systematic reviews
Results: bibliometric/citation analysis

Total unique citations on TB from PubMed & Embase for 2007-2008: 6459

- Abstracts available: 5438 (84.2%)
- Original Study: 4266 (78.4%)
- Abstracts excluded: Not original study 215 (4.0%)
- Citations excluded: No abstract available 1021 (15.8%)

- Coded as per UK Health Research Classification System (HRCS)

1. Underpinning Research
   - Biological 584 (83.7%)
   - Psychological, socioeconomic 59 (8.5%)
   - Chemical, physical 9 (1.3%)
   - Methodology 52 (7.8%)
   - Resources 7 (1.1%)

2. Etiology
   - Biological 1607 (37.7%)
   - Physical 9 (1.5%)
   - Methodology 52 (11.3%)

3. Prevention of Disease and Conditions, and Promotion of Well Being
   - Behaviour 209 (4.9%)
   - Environmental 14 (6.7%)
   - Nutrition, chemoprevention 0 (0%)

4. Detection, Screening and Diagnosis
   - Behaviour 132 (20.9%)
   - Environmental 12 (1.7%)
   - Psychological, social, economic 92 (5.7%)

5. Development of Treatments and Therapeutic Interventions
   - Pharmaceutical 316 (96.4%)
   - Celluar/gene 3 (0.9%)

6. Evaluation of Treatments and Therapeutic Interventions
   - Pharmaceutical 189 (68.2%)
   - Cellular/gene 1 (0.4%)

7. Management of Diseases and Conditions
   - Individual care needs 155 (44.4%)
   - End of life care 1 (0.3%)
   - Resources 4 (1.1%)

8. Health and Social Care Services Research
   - Organisation delivery of services 199 (58.4%)
   - Health services economics 19 (5.6%)
   - Policy, ethics research governance 41 (11.9%)
   - Methodology 915 (26.6%)
   - Resources 0 (0%)

9. Impact
   - 6 (0.9%)

10. Population
    - Screening 12 (1.7%)

Citations excluded: Not original study 315 (4.9%)
Citations excluded: No abstract available 1021 (15.8%)
Distribution of major types of TB research activities [N=6459]

Distribution of study types within diagnostic research [N=699]
Distribution of phases within evaluation studies of diagnostics [N=584]

Countries accounting for the majority of diagnostic studies

<table>
<thead>
<tr>
<th>Country</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>86</td>
<td>12.3</td>
</tr>
<tr>
<td>China</td>
<td>50</td>
<td>7.1</td>
</tr>
<tr>
<td>USA</td>
<td>47</td>
<td>6.7</td>
</tr>
<tr>
<td>Japan</td>
<td>44</td>
<td>6.3</td>
</tr>
<tr>
<td>Brazil</td>
<td>36</td>
<td>5.1</td>
</tr>
<tr>
<td>Russia</td>
<td>36</td>
<td>5.1</td>
</tr>
<tr>
<td>South Africa</td>
<td>30</td>
<td>4.3</td>
</tr>
<tr>
<td>Turkey</td>
<td>29</td>
<td>4.1</td>
</tr>
<tr>
<td>Republic of Korea</td>
<td>23</td>
<td>3.3</td>
</tr>
</tbody>
</table>
Distribution of outcomes reported in abstracts of diagnostic studies [N=699]

Distribution of the purpose of the test within diagnostics studies [N=699]
Distribution of the technology platform of the test within diagnostics studies [N=699]

Results: quality and reporting of diagnostic accuracy studies
Quality and Reporting of Diagnostic Accuracy Studies in TB, HIV and Malaria: Evaluation Using QUADAS and STARD Standards

Patricia Scolari Fontela, Mitika Pant Pai, Ian Schiller, Nandini Dendukuri, Andrew Ramsay, Madhukar Pai

Background: Methodological quality and reporting are known concerns with diagnostic accuracy studies. In 2003, the QUADAS tool and the STARD standards were published for evaluating the quality and improving the reporting of diagnostic studies, respectively. However, it is unclear whether these tools have been applied to diagnostic studies of infectious diseases. We performed a systematic review on the methodological and reporting quality of diagnostic studies in TB, malaria and HIV.

Methods: We identified diagnostic accuracy studies of commercial tests for TB, malaria and HIV through a systematic search of the literature using PubMed and EMBASE (2004–2005). Original studies that reported sensitivity and specificity data were included. Two reviewers independently extracted data on study characteristics and diagnostic accuracy, and used QUADAS and STARD to evaluate the quality of methods and reporting, respectively.

Findings: Majority (100%) of 219 articles met inclusion criteria. All studies had design deficiencies. Study quality indicators that were met in less than 25% of the studies included adequate description of withdrawals (9%) and reference test execution (15%). Absence of index test bias (15%) and reference test review bias (54%) and report of uninterpretable results (25%). In terms of quality of reporting, 9 STARD indicators were reported in less than 25% of the studies: methods for calculation and estimation of reproducibility (25%), absence of disease progression bias (25%), and absence of differential verification bias (25%). Absence of index test review bias (25%) and clinical review bias (25%).

Conclusion: Recently published diagnostic accuracy studies on commercial tests for TB, malaria and HIV have moderate to low quality and are poorly reported. The frequent use of tests such as QuantiFERON and T-Spot may be necessary to improve the methodological and reporting quality of future diagnostic accuracy studies in infectious diseases.

Quality of TB accuracy studies using QUADAS [N=45]

<table>
<thead>
<tr>
<th>Quality Item</th>
<th>45 studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adequate spectrum composition</td>
<td>26 (58%)</td>
</tr>
<tr>
<td>Clear description of selection criteria</td>
<td>21 (47%)</td>
</tr>
<tr>
<td>Adequate reference standard</td>
<td>44 (98%)</td>
</tr>
<tr>
<td>Absence of disease progression bias</td>
<td>42 (93%)</td>
</tr>
<tr>
<td>Absence of partial verification bias</td>
<td>44 (98%)</td>
</tr>
<tr>
<td>Absence of differential verification bias</td>
<td>42 (93%)</td>
</tr>
<tr>
<td>Absence of incorporation bias</td>
<td>45 (100%)</td>
</tr>
<tr>
<td>Absence of index test review bias</td>
<td>6 (13%)</td>
</tr>
<tr>
<td>Absence of reference test review bias</td>
<td>7 (15%)</td>
</tr>
<tr>
<td>Absence of clinical review bias</td>
<td>14 (31%)</td>
</tr>
<tr>
<td>Report of uninterpretable results</td>
<td>9 (20%)</td>
</tr>
<tr>
<td>Description of withdrawals</td>
<td>3 (7%)</td>
</tr>
</tbody>
</table>
17 meta-analysis with over 500 diagnostic studies

- 52% (range 16 – 100%) of the trials used a prospective data collection design.
- 30% (range 0 – 95%) of the trials used a consecutive or random sampling method to recruit subjects.
- 75% (range 43 – 100%) of the trials used a cross-sectional design, and the case–control approach was used in about 25% of the studies.
- Any form of blinding was used in only 35% (range 0 – 78%) of the trials.
- In most studies (87%; range 10 – 100%), the index test results were verified by a reference standard test.

<p>| Table 2. Methodological quality of studies on tuberculosis diagnosis in recent updated meta-analyses |
|-------------------------------------------------|----------------|----------------|------------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Meta-analysis No. of studies</th>
<th>Diagnostic test</th>
<th>Prognostic test</th>
<th>Pre-test sample</th>
<th>Consecutive subjects</th>
<th>Blinded trial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>14</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Criteria歌</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Standard</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Source</td>
<td>4</td>
<td>4</td>
<td>12</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

关于15% of all TB papers were mainly focused on TB diagnosis.
- Of these, about 85% were evaluation studies of tests and markers.
- Of these evaluation studies, about 85% are early phase studies of test accuracy; there are very little data on impact on patient outcomes.
- Most test accuracy studies are of moderate to low quality and are poorly reported.
- Essential methodological and design elements are often either not reported or poorly reported.
- These results have important implications for policy making.

Conclusions

- About 15% of all TB papers were mainly focused on TB diagnosis.
- Of these, about 85% were evaluation studies of tests and markers.
- Of these evaluation studies, about 85% are early phase studies of test accuracy; there are very little data on impact on patient outcomes.
- Most test accuracy studies are of moderate to low quality and are poorly reported.
- Essential methodological and design elements are often either not reported or poorly reported.
- These results have important implications for policy making.
WHO policy process

- According to WHO, in order to consider a global policy change, WHO must have solid evidence, including clinical trials or field evaluations in high TB prevalence settings.
- Policy process includes a comprehensive review of the evidence, as well as expert opinion and judgment.
- All WHO guidelines will be approved by a Guideline Review Committee.
- All guidelines and policies will explicitly incorporate evidence using the GRADE approach.

The GRADE approach

Clear separation of 2 issues:

1) 4 categories of quality of evidence: ++++ (High), ++++ (Moderate), ++ + + (Low), ++ ++ (Very low)?
   - methodological quality of evidence
   - likelihood of bias
   - by outcome and across outcomes

2) Recommendation: 2 grades – weak/conditional or strong (for or against)?
   - Quality of evidence only one factor
   - Balance of benefits and downsides, values and preferences, resource use

*www.GradeWorking-Group.org

---

**RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS**

**GRADE: grading quality of evidence and strength of recommendations for diagnostic tests and strategies**

The GRADE system can be used to grade the quality of evidence and strength of recommendations for diagnostic tests or strategies. This article explains how patient-important outcomes are taken into account in this process.

**SUMMARY POINTS**

As for other interventions, the GRADE approach to grading the quality of evidence and strength of recommendations for diagnostic tests or strategies provides a comprehensive and transparent approach for developing recommendations.

Cross-sectional or cohort studies can provide high-quality evidence of test accuracy.

However, test accuracy is a surrogate for patient-important outcomes, so such studies often provide low-quality evidence for recommendations about diagnostic tests, even when the studies do not have serious limitations.

Inferring from data on accuracy that a diagnostic test or strategy improves patient-important outcomes will require the availability of effective treatment, reduction of test-related adverse effects or anxiety, or improvement of patients' wellbeing from prognostic information.

Judgments are thus needed to assess the directness of test results in relation to consequences of diagnostic recommendations that are important to patients.

BMJ 2008
Clinical impact of a test result on individual patient outcome
- This is what GRADE needs
- Ideally, needed before policy (but currently not happening)
- Collected at the individual level (as in a clinical trial)
- E.g. If Xpert is used instead of smear microscopy, will help initiate TB treatment quicker and ensure cure?

Epidemiological impact of introducing a test on disease control
- Public health or "societal" impact
- Collected after policy and scale-up
- Collected at the ecological/population level
- E.g. If Xpert is scaled-up in a country, will it help reduce TB transmission and cut TB incidence rates?
GRADE expectations are met in other fields that are well ahead of TB...

- Example: Rapid diagnostics tests (RIDTs) for influenza
  - 100+ accuracy studies
  - 20+ impact studies (including several diagnostic RCTs)

Test accuracy studies

[Graph showing meta-analysis with data points and summary line]

Chartrand C et al.
Impact studies

Impact of the Rapid Diagnosis of Influenza on Physician Decision-Making and Patient Management in the Pediatric Emergency Department

Results of a Randomized, Pragmatic, Controlled Trial

Arun E. Busse, DSC, MD; Kathy S. Mannino, MD; Loren J. Levy, MD; Arlin E. K莖son, MD; MPH; and Donald R. Krockenberger, MD

ABSTRACT: Objective. To determine the impact of the rapid diagnosis of influenza on physician decision-making and patient management. Utilizing blinded algorithms, pediatric emergency department physicians were randomized to perform rapid influenza diagnosis with a commercially available rapid influenza diagnostic test (MODS) or not. The physician decision evaluation tool (PDET) was used to assess the quality and impact of physician decision-making. Methods: A physician decision evaluation tool (PDET) was used to assess the quality and impact of physician decision-making. Results: When rapid influenza diagnostic test results were compared between physician groups, no significant differences were observed. Conclusion: Pediatric emergency department physicians randomized to the rapid influenza diagnostic test arm did not change their diagnostic or therapeutic decisions when compared to a group who did not receive the rapid influenza diagnostic test results.

Impact of Rapid Diagnosis on Management of Adults Hospitalized With Influenza

Arun E. Busse, MD; Yoshio Kubota, MD; S. B. Pacheco; and Donald R. Rock, MD

Background: Rapid influenza testing increases providers’ awareness and facilitates quick evaluation, thereby reducing the time to appropriate interventions. No randomized controlled trials comparing rapid influenza diagnostic tests (MODS) to traditional clinical evaluation of patients with influenza-like illness (ILI) in a hospital setting are currently available. This study was undertaken to determine whether rapid influenza diagnostic tests would increase the initiation and length of time to appropriate intervention for patients with ILI.

Methods: A prospective, blinded, randomized comparison of patients with ILI presenting to an urban hospital emergency department (1999-2003). The management of patients with rapid influenza test results was compared with the management of patients with traditional clinical evaluation of patients with ILI.

Results: Rapid influenza diagnostic tests were associated with an increased initiation of antiviral therapy (30.5% vs 16.5% for patients with rapid influenza test results vs traditional clinical evaluation of patients, respectively, P<.0001). The management of patients with rapid influenza test results was associated with an increased length of time to appropriate antiviral therapy initiation (20.5% vs 17.6% for patients with rapid influenza test results vs traditional clinical evaluation of patients, respectively, P<.0001).

Conclusion: Rapid influenza diagnostic tests are associated with an increased initiation and length of time to appropriate intervention for patients with ILI.

In TB, since we have mostly accuracy data:
example from WHO EGM on tests for drug-resistant TB

<table>
<thead>
<tr>
<th>Test, # Studies (participants)</th>
<th>Design</th>
<th>Limitations</th>
<th>Directness</th>
<th>Inconsistency</th>
<th>Imprecise or sparse data</th>
<th>Publication Bias</th>
<th>Evidence Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODS, 9 (1474)</td>
<td>CS &amp; CC</td>
<td>Low</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>NTA, 19 (2304)</td>
<td>CS &amp; CC</td>
<td>Low</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>CRI, 31 (2498)</td>
<td>CS &amp; CC</td>
<td>Low</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
</tr>
<tr>
<td>TLA, 3 (439)</td>
<td>CS &amp; CC</td>
<td>Low</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Phage, 12 (2935)</td>
<td>CS &amp; CC</td>
<td>Moderate/High</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>P&lt;.001</td>
<td>Very low</td>
</tr>
<tr>
<td>LPA, 12 (4937)</td>
<td>CS &amp; CC</td>
<td>Low</td>
<td>No evidence</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>Moderate</td>
</tr>
</tbody>
</table>

Regardless of study quality, precision, consistency ... accuracy studies will never lead to High Quality Evidence.
There are 45+ systematic reviews on TB tests, but almost all focus on sensitivity and specificity (accuracy)

Conclusions

- Test accuracy studies need to be done better and reported better
- Need to go beyond test accuracy and generate evidence on:
  - Impact of test on patient important outcomes
  - Impact of test on diagnostic thinking and decision making
  - Incremental or added value beyond what is already in place
  - Time to diagnosis and treatment
  - Cost–effectiveness
Clinical impact is therefore a key part of demonstration studies and evidence for scale-up

Figure 4: Schematic showing the pathway to tuberculosis diagnostics, from concept to delivery
Source: Stop TB Partnership’s New Diagnostics Working Group. Pathways to better diagnostics for tuberculosis: a blueprint for the development of TB diagnostics (2009), and reproduced with permission from author and publisher.

Acknowledgements

- Andy Ramsay
- Karen Steingart
- Rick O’Brien
- Karin Weyer
- Holger Schunemann
- Michael Kimerling
- Frank Cobelens
- Susan van den Hof
- Bertie Squire
- Christian Lienhardt