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

McGill | Stop TB Partnership | Francis J. Curry National Tuberculosis Center | TDR For research on diseases of poverty | World Health Organization

1

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

2

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

3

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 PubMed, the search strategy was: ("Mycobacterium tuberculosis"[Majr] OR "Tuberculosis"[Majr] OR "Tuberculosis/diagnosis"[Mesh] OR tuberculosis Field: Title) Limits: Publication Date from 2007/01/01 to 2008/12/31 NOT Field: Title, Editorial, Letter, Meta-Analysis, Practice Guideline, Review, Addresses, Bibliography, Biography, Comment, Dictionary, Directory, Interview, Newspaper Article.
    - 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")

4

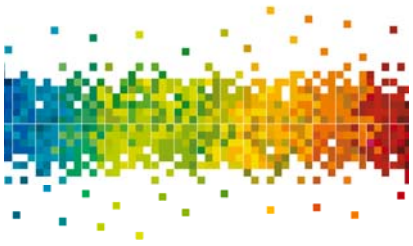
Methods

- 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.

5

UK Clinical Research Collaboration

Health Research Classification System



Overview of the Research Activity Codes

1	**Underpinning Research**
1.1	Normal biological development and functioning
1.2	Psychological and socioeconomic processes
1.3	Chemical and physical sciences
1.4	Methodologies and measurements
1.5	Resources and infrastructure (underpinning)
2	**Aetiology**
2.1	Biological and endogenous factors
2.2	Factors relating to physical environment
2.3	Psychological, social and economic factors
2.4	Surveillance and distribution
2.5	Research design and methodologies (aetiology)
2.6	Resources and infrastructure (aetiology)
3	**Prevention of Disease and Conditions, and Promotion of Well-Being**
3.1	Primary prevention interventions to modify behaviours or promote well-being
3.2	Interventions to alter physical and biological environmental risks
3.3	Nutrition and chemoprevention
3.4	Vaccines
3.5	Resources and infrastructure (prevention)
4	**Detection, Screening and Diagnosis**
4.1	Discovery and preclinical testing of markers and technologies
4.2	Evaluation of markers and technologies
4.3	Influences and impact
4.4	Population screening
4.5	Resources and infrastructure (detection)
5	**Development of Treatments and Therapeutic Interventions**
5.1	Pharmaceuticals
5.2	Cellular and gene therapies
5.3	Medical devices
5.4	Surgery
5.5	Radiotherapy
5.6	Psychological and behavioural
5.7	Physical
5.8	Complementary
5.9	Resources and infrastructure (development of treatments)
6	**Evaluation of Treatments and Therapeutic Interventions**
6.1	Pharmaceuticals
6.2	Cellular and gene therapies
6.3	Medical devices
6.4	Surgery
6.5	Radiotherapy
6.6	Psychological and behavioural
6.7	Physical
6.8	Complementary
6.9	Resources and infrastructure (evaluation of treatments)
7	**Management of Diseases and Conditions**
7.1	Individual care needs
7.2	End of life care
7.3	Management and decision making
7.4	Resources and infrastructure (disease management)
8	**Health and Social Care Services Research**
8.1	Organisation and delivery of services
8.2	Health and welfare economics
8.3	Policy, ethics and research governance
8.4	Research design and methodologies
8.5	Resources and infrastructure (health services)

Research Activity Codes

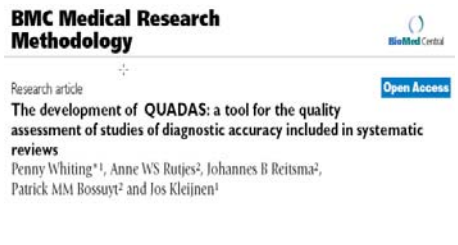
6

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

7

QUADAS tool for quality assessment



BMC Medical Research Methodology

BioMed Central

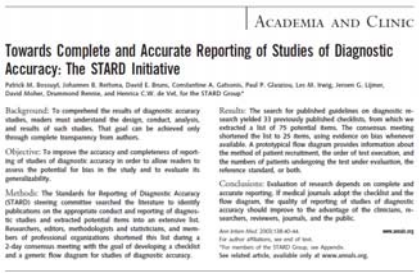
Research article

Open Access

The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews

Penny Whiting*[1], Anne WS Rutjes[2], Johannes B Reitsma[2], Patrick MM Bossuyt[2] and Jos Kleijnen[1]

STARD reporting standards



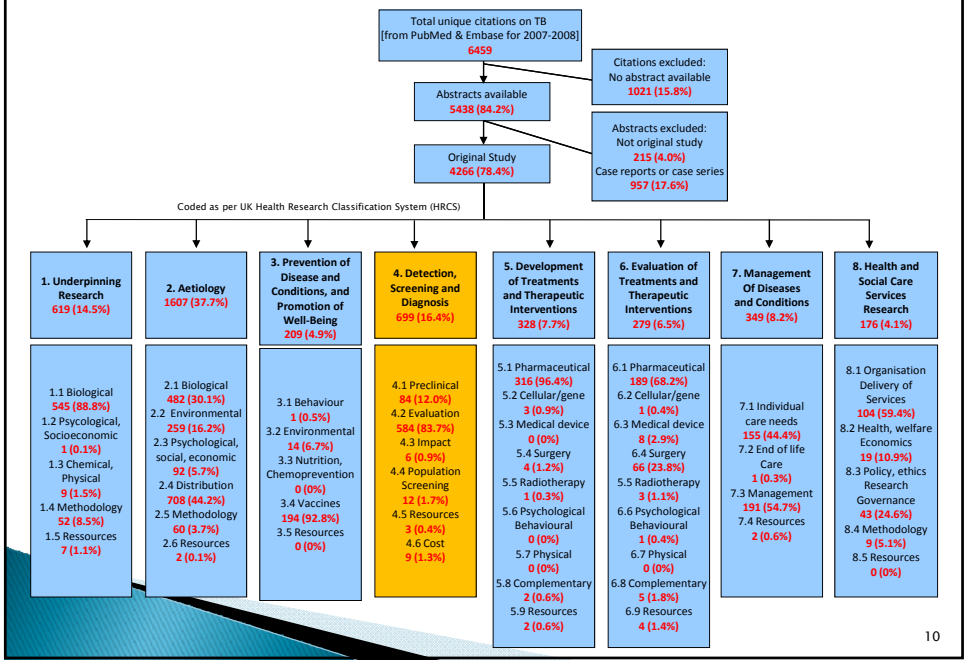
ACADEMIA AND CLINIC

Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative

8

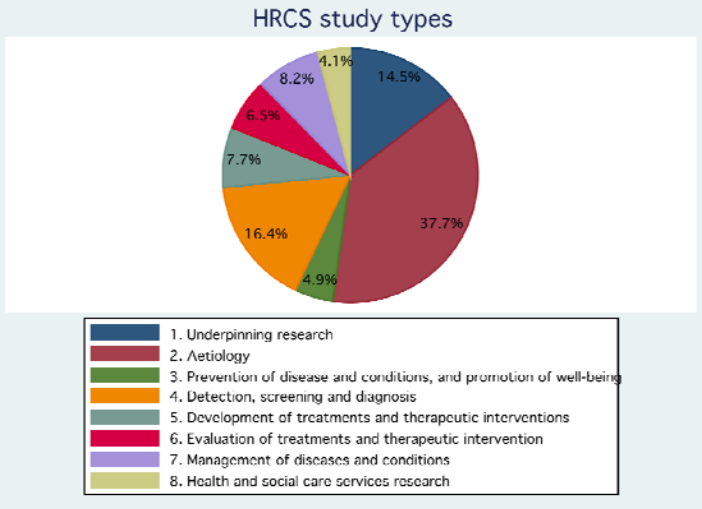
Results: bibliometric/citation analysis

9



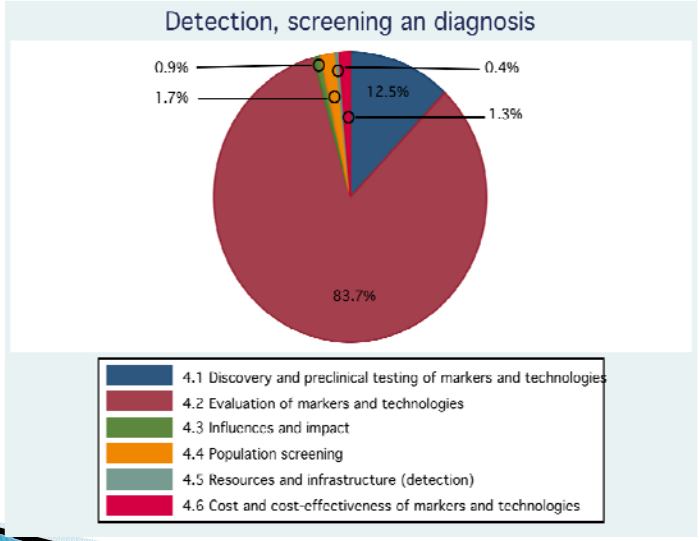
10

Distribution of major types of TB research activities [N=6459]



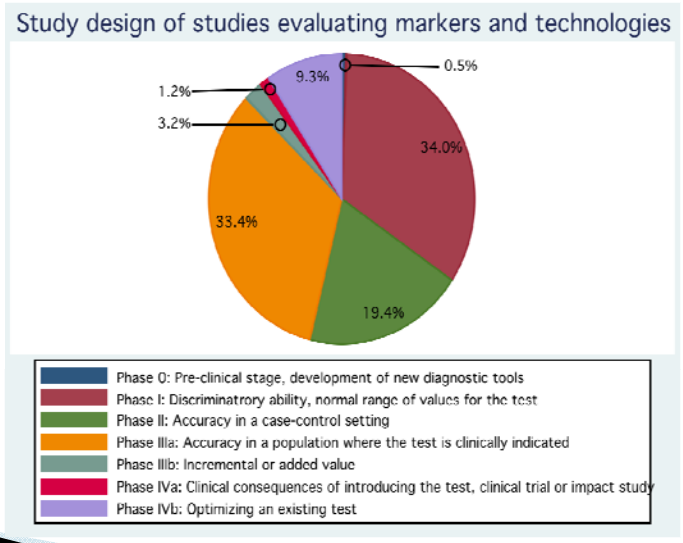
11

Distribution of study types within diagnostic research [N=**699**]



12

Distribution of phases within evaluation studies of diagnostics [N=584]

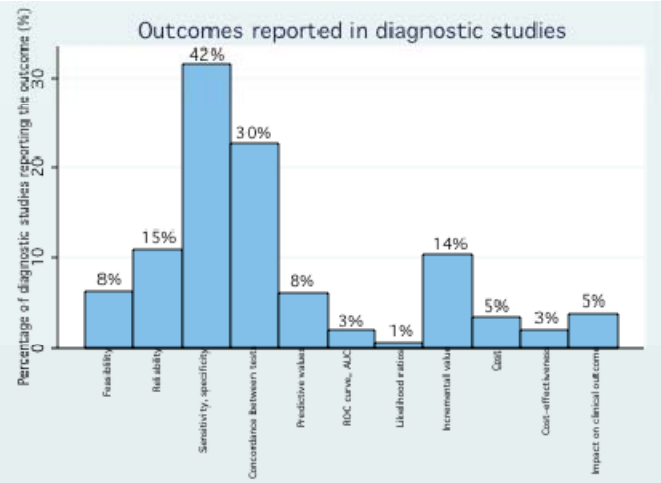


13

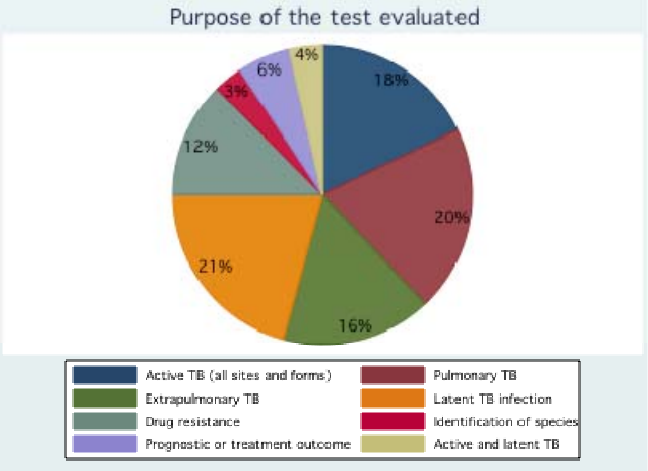
Countries accounting for the majority of diagnostic studies

Country	N	%	Country	N	%
India	86	12.3	Germany	19	2.7
China	50	7.1	Italy	19	2.7
USA	47	6.7	Peru	17	2.4
Japan	44	6.3	UK	15	2.1
Brazil	36	5.1	Taiwan	14	2.0
Russia	36	5.1	Netherlands	13	1.8
South Africa	30	4.3	Spain	12	1.7
Turkey	29	4.1	Iran	10	1.4
Republic of Korea	23	3.3			

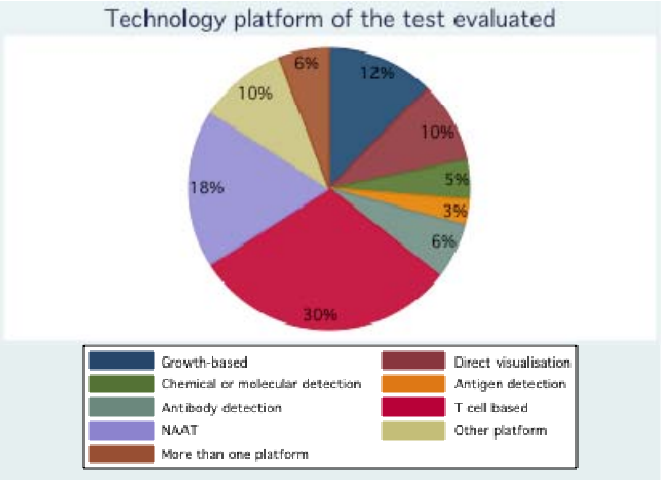
Distribution of outcomes reported in abstracts of diagnostic studies [N=**699]**



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

18

OPEN ACCESS Freely available online

PLoS one

Quality and Reporting of Diagnostic Accuracy Studies in TB, HIV and Malaria: Evaluation Using QUADAS and STARD Standards

Patricia Scolari Fontela[1], Nitika Pant Pai[2], Ian Schiller[2], Nandini Dendukuri[2], Andrew Ramsay[3], Madhukar Pai[1,4]*

1 Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada, 2 Department of Medicine, Division of Clinical Epidemiology, McGill University, Montreal, Canada, 3 Special Programme for Research and Training in Tropical Diseases, World Health Organization, Geneva, Switzerland, 4 Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, Montreal, Canada

Abstract

Background: Poor 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 (2004–2006). 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: Ninety (38%) of 238 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 (6%) and reference test execution (10%), absence of index test review bias (19%) and reference test review bias (24%), and report of uninterpretable results (22%). In terms of quality of reporting, 9 STARD indicators were reported in less than 25% of the studies: methods for calculation and estimates of reproducibility (0%), adverse effects of the diagnostic tests (1%), estimates of diagnostic accuracy between subgroups (10%), distribution of severity of disease/other diagnoses (11%), number of eligible patients who did not participate in the study (14%), blinding of the test readers (16%), and description of the team executing the test and management of indeterminate/outlier results (both 17%). The use of STARD was not explicitly mentioned in any study. Only 22% of 46 journals that published the studies included in this review required authors to use STARD.

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

PLoS One 2009

19

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

Quality item	45 studies n (%)
Adequate spectrum composition	26 (58)
Clear description of selection criteria	21 (47)
Adequate reference standard	44 (98)
Absence of disease progression bias	42 (93)
Absence of partial verification bias	44 (98)
Absence of differential verification bias	42 (93)
Absence of incorporation bias	45 (100)
Absence of index test review bias	6 (13)
Absence of reference test review bias	7 (16)
Absence of clinical review bias	14 (31)
Report of uninterpretable results	9 (20)
Description of withdrawals	3 (7)

Fontela et al. PLoS One 2009

20

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.

Table 2. Methodological quality of studies on tuberculosis diagnostics in recently published meta-analyses.

Meta-analysis	No. of studies	Diagnostic test	Average size of each study	Prospective data collection (%)	Consecutive or random sampling of subjects (%)	Cross-sectional design (%)	Blinded interpretation of test results* (%)	Complete verification of index test results‡ (%)	Ref.
Sarmiento et al. (2003)	16	PCR on respiratory specimens for smear-negative pulmonary TB	NR	50	NR	NR	63	100	[12]
Goto et al. (2003)	40	ADA for TB pleural effusion	137	NR	NR	NR	0	NR	[13]
Pai et al. (2003)	49	NAT for TB meningitis	42	61	49	61	59	94	[14]
Greco et al. (2003)	44	ADA and IFN-γ tests for TB pleural effusion	135	NR	NR	NR	9	NR	[15]
Pai et al. (2004)	40	NAT for TB pleural effusion	60	63	53	70	55	100	[16]
Flores et al. (2005)	84	In-house PCR for pulmonary TB	149	NR	NR	71	34	NR	[17]
Kalantri et al. (2005)	13	Phage amplification tests for pulmonary TB	448	NR	NR	85	23	100	[18]
Pai et al. (2005)	21	Phage-based tests for rifampin resistance	85	NR	38	NR	57	100	[19]
Morgan et al. (2005)	15	Line probe assay for rifampin resistance	91	NR	0	NR	13	100	[20]
Greco et al. (2006)	63	Commercial NAT for pulmonary TB	410	16	32	NR	16	NR	[21]
Steingart et al. (2006)	45	Fluorescence versus conventional sputum smear microscopy for pulmonary TB	493	100	36	NR	49	NR	[22]
Steingart et al. (2006)	83	Direct versus concentrated sputum smear microscopy for pulmonary TB	256	100	21	NR	31	NR	[23]

*At least single-blind. ‡By reference standard.
ADA: Adenosine deaminase; IFN: Interferon; NAT: Nucleic acid amplification test; NR: Not reported; TB: Tuberculosis.

Pai M, O'Brien R. Exp Rev Mol Diagn 2006.

21

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

22

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

Box 1. WHO Policy Process for Tuberculosis

1. Identifying the Need for a Policy Change

The need to formulate new or revised policies may arise from WHO's ongoing monitoring of technical developments or from interested parties submitting requests with supporting documentation for policy or guideline development. WHO receives information about a new technology or approach via many channels, with the most direct lines coming from national TB programs and researchers themselves. To consider a global policy change, WHO must have solid evidence, including clinical trials or field evaluations in high TB prevalence settings.

2. Reviewing the Evidence

WHO may carry out or commission a review of the documentation of the technology's clinical or programmatic performance, including newly published and "grey" research or reviews, "proof of principle" reports, large-scale field trials, and demonstration projects in different resource settings. Standardized evaluation criteria have been and are being developed by the New Diagnostics, New Drugs, and New Vaccines Working Groups of the Stop TB Partnership.

3. Convening an Expert Panel

If the evidence base is compelling, WHO will convene an external panel of experts, excluding all original principal investigators from the studies. The panel will review the evidence and make a recommendation or propose draft policies or guidelines to WHO's Strategic and Technical Advisory Group for Tuberculosis (STAG-TB).

4. Assessing Draft Policies and Guidelines

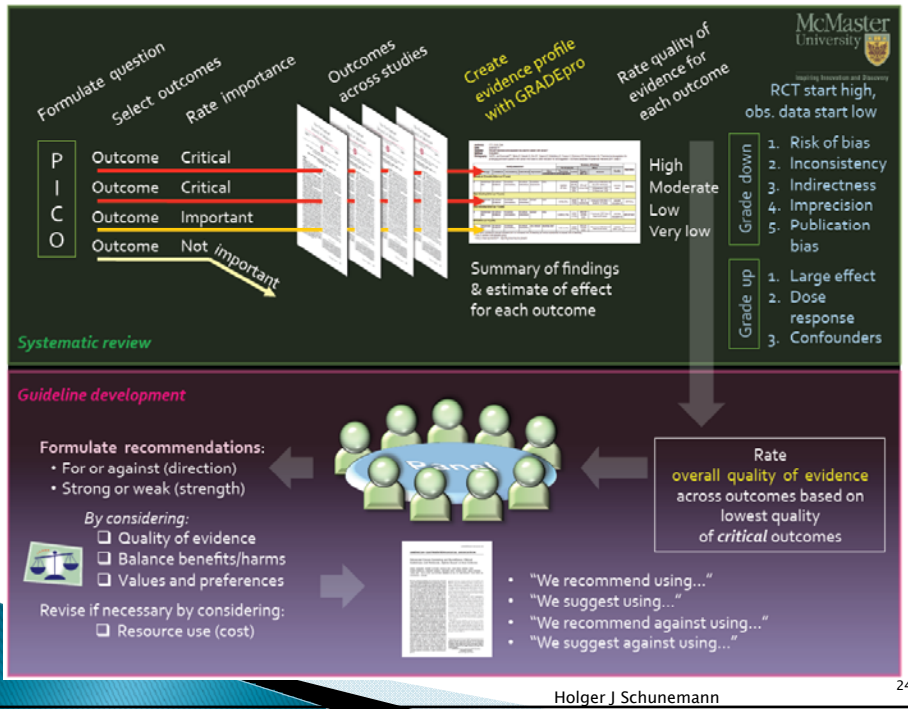
STAG-TB provides objective, ongoing technical and strategic advice to WHO on TB care and control. STAG-TB's objectives are to provide the Director-General, through the Stop TB Department, with an independent evaluation of the strategic, scientific, and technical aspects of WHO's TB activities; review progress and challenges in WHO's TB-related core functions; review and make recommendations on committees and working groups; and make recommendations on WHO's TB activity priorities. STAG-TB reviews the policy drafts and supporting documentation during its annual meeting. STAG-TB may endorse the policy recommendation with or without revisions, request additional information and re-review the evidence in subsequent years, or reject the recommendation.

5. Formulating and Disseminating Policy

New WHO policies and guidelines will be disseminated through different channels to Member States, including through the World Health Assembly, WHO Web site, listservs, and journal publications. WHO also disseminates its recommendations to other agencies and donors engaged in TB control activities.

Source: World Health Organization [7]

http://www.who.int/tb/dots/laboratory/policy/en/index4.html



Holger J Schunemann

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

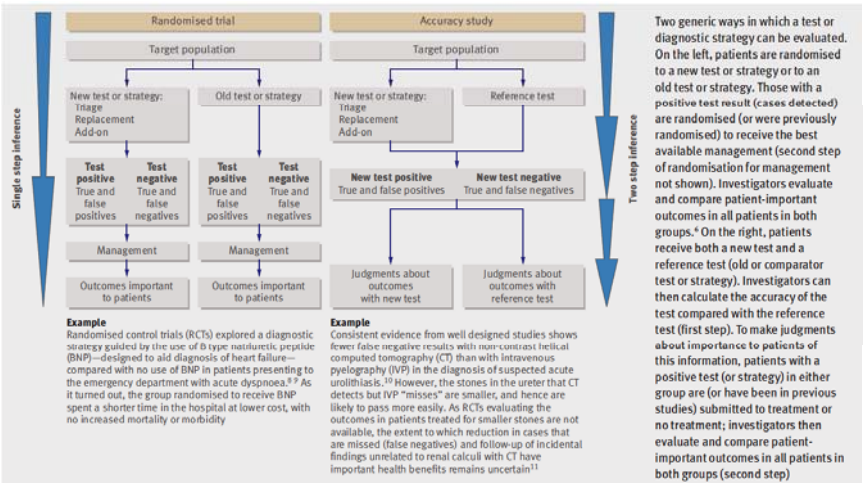
However, test accuracy is a surrogate for patient-important outcomes, so such studies often provide low quality evidence for recommendations about diagnostic tests, even when the studies do not have serious limitations

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 26

GRADE: for high quality evidence, impact on patient-important outcomes needs to be demonstrated



Schunemann H et al. BMJ 2008

Separating clinical from epidemiological impact

- 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?

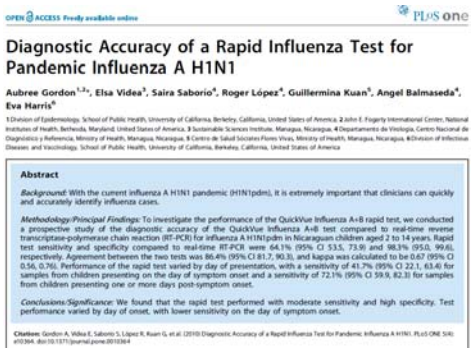
28

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)

29

Test accuracy studies



Diagnostic Accuracy of a Rapid Influenza Test for Pandemic Influenza A H1N1

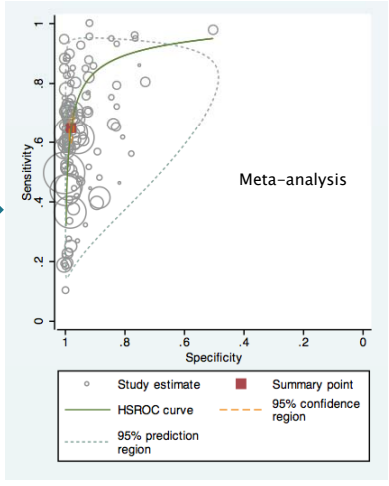
Aubree Gordon, Elsa Videa, Saira Saborio, Roger López, Guillermina Kuan, Angel Balmaseda, Eva Harris

Abstract

Background: With the current influenza A H1N1 pandemic (H1N1pdm), it is extremely important that clinicians can quickly and accurately identify influenza cases.

Methodology/Principal Findings: To investigate the performance of the QuickVue Influenza A+B rapid test, we conducted a prospective study of the diagnostic accuracy of the QuickVue Influenza A+B test compared to real-time reverse transcriptase-polymerase chain reaction (RT-PCR) for influenza A H1N1pdm in Nicaraguan children aged 2 to 14 years. Rapid test sensitivity and specificity compared to real-time RT-PCR were 64.1% (95% CI 53.5, 73.9) and 98.3% (95.0, 99.6), respectively. Agreement between the two tests was 86.4% (95% CI 81.7, 90.3), and kappa was calculated to be 0.67 (95% CI 0.56, 0.76). Performance of the rapid test varied by day of presentation, with a sensitivity of 41.7% (95% CI 22.1, 63.4) for samples from children presenting on the day of symptom onset and a sensitivity of 72.1% (95% CI 59.9, 82.3) for samples from children presenting one or more days post-symptom onset.

Conclusions/Significance: We found that the rapid test performed with moderate sensitivity and high specificity. Test performance varied by day of onset, with lower sensitivity on the day of symptom onset.



Chartrand C et al.

30

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, Prospective, Controlled Trial

Aleta B. Bonner, DVM, MD*; Kathy W. Monroe, MD*; Lynya I. Talley, PhDS; Ann E. Klauner, MD, MPH*; and David W. Kimberlin, MD‡

ABSTRACT. *Objective.* To determine the impact of the rapid diagnosis of influenza on physician decision-making and patient management, including laboratory tests and radiographs ordered, patient charges associated with these tests, antibiotics/antivirals prescribed, and length of time to patient discharge from the emergency department.

Methods. Patients aged 2 months to 21 years presenting to an urban children's teaching hospital emergency department were screened for fever and cough, coryza, myalgias, headache, and/or malaise. After obtaining informed consent, patients were randomized to 1 of 2 groups: 1) physician receives (physician aware of) the rapid influenza test result; or 2) physician does not receive (physician unaware of) the result. For patients in the physician aware group, nasopharyngeal swabs were obtained, immediately tested with the FluOIA test for influenza A and B, and the result was placed on the chart before patient evaluation by the attending physician. For the physician unaware group, nasopharyngeal swabs were obtained, stored according to manufacturer's directions, and tested within 24 hours. Results for the physician unaware group were not disclosed to the treating physicians at any time. The 2 resultant influenza-positive groups (aware and unaware) were compared for laboratory and radiograph studies and their associated patient charges, antibiotic/antiviral prescriptions, and length of stay in the emergency department.

Results. A total of 418 patients were enrolled, and 391 completed the study. Of these, 202 tested positive for influenza. Comparison of the 96 influenza-positive patients whose physician was aware of the result with the 106 influenza-positive patients whose physician was unaware of the result revealed significant reductions among the former group in: 1) numbers of complete blood counts, blood cultures, urinalyses, urine cultures, and chest radiographs performed; 2) charges associated with these tests; 3) antibiotics prescribed; and 4) length of stay in the emergency department. The number of influenza-positive patients who received prescriptions for antiviral drugs was significantly higher among those whose physician was aware of the result.

Conclusions. Physician awareness of a rapid diagnosis of influenza in the pediatric emergency department significantly reduced the number of laboratory tests and radiographs ordered and their associated charges, decreased antibiotic use, increased antiviral use, and decreased length of time to discharge. *Pediatrics* 2003;112:363–367; *pediatric, influenza, physician decision-making, patient management.*

Influenza virus types A and B are common respiratory pathogens in the pediatric population. Depending on age, attack rates may be 1.5 to 3 times higher than for adults, with school-aged children having the highest attack rates.[1,2] A retrospective cohort study of children under 15 years of age demonstrated outpatient visits attributable to influenza ranging from 6 to 15 per 100 children.[3] Infection with influenza virus leads to a significant increase in primary care visits, and also increases in emergency department utilization during wintertime epidemics.[3]

Rapid diagnostic test kits for influenza types A and B are currently available for outpatient use and have proven to be both sensitive and specific.[4–7] Few studies have been performed which analyze the impact of rapid diagnostic testing for influenza and subsequent effect on patient management.[8–11] To date, there are no prospective, randomized studies analyzing use of rapid influenza testing and effect on patient management in the pediatric emergency department. Rapid diagnostic tests are not currently routinely incorporated in the work-up of infants and children with fever and vague symptoms, or with fever and no documented source.[12] Use of rapid tests in the pediatric emergency department which are sensitive and specific for influenza could potentially decrease performance of other more invasive tests, thereby reducing associated patient charges, reducing patient length of stay in the emergency depart-

Pediatrics 2003;112:363–367

Impact of Rapid Diagnosis on Management of Adults Hospitalized With Influenza

ARCHIVES EXPRESS

Ann R. Falsey, MD; Yoshihiko Murata, MD, PhD; Edward E. Walsh, MD

Background: Rapid influenza testing decreases antibiotic and ancillary test use in febrile children, yet its effect on the care of hospitalized adults is unexplored. We compared the clinical management of patients with influenza whose rapid antigen test result was positive (Ag+) with the management of those whose rapid antigen test result was negative or the test was not performed (Ag0).

Methods: Medical record review was performed on patients with influenza hospitalized during 4 winters (1999-2003). Hospital policy mandated influenza testing (antigen or culture) for all patients with acute cardiopulmonary diseases admitted from November 15 through April 15. A subset of patients participated in an epidemiological study and had reverse-transcriptase polymerase chain reaction or serologic testing performed. Clinical data from Ag+ and Ag0 patients were compared.

Results: Of 166 patients with available records, 86 were Ag+ and 80 were Ag0. Antibiotic use (74 [86%] of 86 patients vs 79 [99%] of 80 patients; P = .002) was less and antibiotic discontinuance (12 [14%] of 86 patients vs 2 [2%] of 80 patients; P = .01) was greater in Ag+ compared with Ag0 patients. No significant differences in antibiotic days, length of hospital stay, or antibiotic complications were noted. Antiviral use (63 [73%] of 86 patients vs 0 [0%] of 80 patients; P < .001) was greater in Ag+ than Ag0 patients. Antigen status was independently associated with withholding or discontinuing antibiotics in multivariate analysis. Of 44 Ag+ patients deemed low risk for bacterial infection, 27 continued to receive antibiotics despite positive influenza test results. These patients more commonly had pulmonary disease and had significantly more abnormal lung examination results (P = .005) compared with those in whom antibiotics were withheld or discontinued.

Conclusions: Rapid influenza testing leads to reductions in antibiotic use in hospitalized adults. Better tools to rule out concomitant bacterial infection are needed to optimize the impact of viral testing.

Arch Intern Med. 2007;167:354-360

"Impact" outcomes include:

- Change in clinical decisions
- Reduction in antibiotic use
- Increased antiviral use
- Decreased length of time to discharge
- Reduction in lab investigations, etc

31

In TB, since we have mostly accuracy data:

example from WHO EGM on tests for drug-resistant TB

Test, # Studies (participants)	Design	Limitations	Directness	Inconsistency	Imprecise or sparse data	Publication Bias	Evidence Quality
MODS, 9 (1474)	CS & CC	Low	No evidence −1	Low	Low	Possible	Moderate
NRA, 19 (2304)	CS & CC	Low	No evidence −1	Low	Low	Possible	Moderate
CRI, 31 (2498)	CS & CC	Low	No evidence −1	Low	Low	Possible	Moderate
TLA, 3 (439)	CS & CC	Low	No evidence −1	Low	High −1	Possible	Low
Phage, 12 (2935)	CS & CC	Moderate/High −1	No evidence −1	Moderate/High −1	Low	Probable	Very low
LPA, 12 (4937)	CS & CC	Low	No evidence −1	Low	Low	Possible	Moderate

- Regardless of study quality, precision, consistency … accuracy studies will never lead to High Quality Evidence

32

There are 45+ systematic reviews on TB tests, but almost all focus on sensitivity and specificity (accuracy)

33

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

34

Clinical impact is therefore a key part of demonstration studies and evidence for scale-up

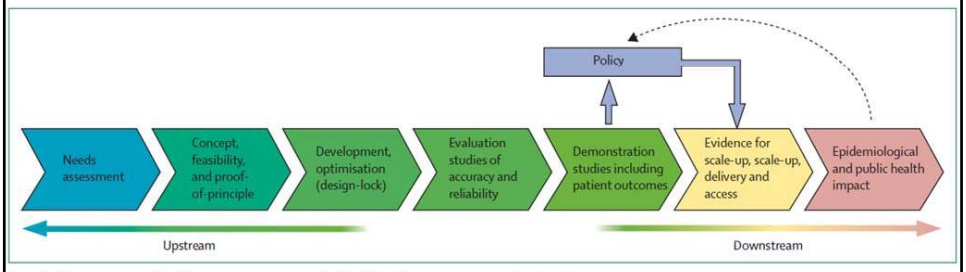


Figure 4: Schematic showing the pathway to tuberculosis diagnostics, from concept to delivery
Source: Stop TB Partnership's New Diagnostics Working Group. Pathways to better diagnostics for tuberculosis: a blueprint for the development of TB diagnostics (2009),[180] and reproduced with permission from 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

36