Optimism bias, inflated accuracy estimates, and contradicted findings in TB diagnostic research

Madhukar Pai, MD, PhD [madhukar.pai@mcgill.ca]
Jessica Minion, MD
McGill University, Montreal

Context

There is some evidence that:

- Initially stronger effects and subsequent contradictions are not infrequent in highly cited research of clinical interventions and their outcomes.
- Claims from highly cited observational studies persist and continue to be supported in the medical literature despite strong contradictory evidence from randomized trials.
- Newly discovered true (non-null) associations often have inflated effects compared with the true effect sizes.
- Publication bias is a major concern and may be more widespread than we think; some have also challenged the conventional publishing model.
- Even within published studies, selective reporting of positive outcomes in randomized trials as well as observational studies appears to be frequent.
- Lack of replication of research findings and over-interpretation of findings are other concerns, especially in some fields.
Context

- All of these likely result in "optimism bias"—unwarranted belief in the efficacy of new therapies, and overinterpretation of the applicability of findings.
- Optimism bias is more likely in industry-supported research.
- Optimism bias and conflicting study findings appear to be eroding the public's faith in research.
- Even among some researchers, there is concern that most published research findings may be false.
Publication bias and selective publication

SELECTIVE PUBLICATION OF ANTIDEPRESSANT TRIALS AND ITS INFLUENCE ON APPARENT EFFICACY


ABSTRACT

While almost all trials with “positive” results on antidepressants had been published, trials with “negative” results submitted to the US Food and Drug Administration, with few exceptions, remained either unpublished or were published with the results presented so that they would appear “positive.”

Non-replicated studies and publication bias – especially in genetic and biomarker studies
Even within published studies, selective reporting of outcomes

In RCTs

iciency in published studies, selective reporting of outcomes, and the nature of these studies on the same question, and importantly the ratio of true to false evidence underlying a relationship proposed in each study, tends to vary in its influence on effect size estimates. While there is a general consensus on the conclusion that most positive findings are overestimates, the generalizability of the conclusions and the variability in the effect size estimates is cause for concern. While there is a growing number of studies examining the impact of publication bias on the validity of research findings, especially in epidemiologic studies, the variability in the effect size estimates is cause for concern.

In observational studies

Selection in Reported Epidemiologic Risks: An Empirical Assessment

Abstract

Background

Epidemiologic studies reporting the association of a drug or a treatment with an outcome of interest, are often plausible, especially if the observed relationship is found in multiple studies with different designs and settings. It is more likely for a treatment effect to be overstated when the evidence is weak. However, it is reassuring to note that the variability in the effect size estimates is cause for concern. While there is a growing number of studies examining the impact of publication bias on the validity of research findings, especially in epidemiologic studies, the variability in the effect size estimates is cause for concern.

In RCTs

In observational studies

Why Most Published Research Findings Are False

John P.A. Ioannidis

Summary

Recent advances in science and medicine have led to the publication of numerous research findings. These findings are often based on small studies with limited power, and they are often reported in a way that emphasizes their significance. However, this emphasis can lead to a distorted view of the evidence, and it can result in the publication of false positive findings. Using a simple model of the process of research, it can be shown that the proportion of false positive findings is likely to increase with the number of studies published.

Several methodological strategies have been proposed to mitigate this problem, including the use of meta-analysis, the use of replication studies, and the use of pre-registration of studies. However, these strategies have limitations, and they do not provide a complete solution to the problem of false positive findings.

It can be proven that most claimed research findings are false. The probability of a claim being true depends on the sample size and the effect size of the study. In general, the larger the sample size and the smaller the effect size, the higher the probability of a claim being true. This is because larger studies have more power to detect true effects, and smaller studies are more prone to false positive findings.

Conclusions

The high rate of false positive findings in published research is a significant concern, and it is important to address this issue. Researchers, funding agencies, and journals should consider implementing strategies to reduce the rate of false positive findings. These strategies should focus on increasing the power of studies, increasing the sample size, and reducing the publication bias.
As researchers in tuberculosis we asked the question:

“is there evidence for ‘optimism bias’ in TB diagnostic research?”

We present several case studies to answer this question

Several new diagnostics are in the pipeline
But do they work? Will optimism bias prove to be a big issue?
Case study 1:
How much evidence is sufficient for commercialization?

Promising new Point of Care test: LAM antigen detection

Rapid diagnosis of tuberculosis by detection of mycobacterial lipoarabinomannan in urine

- Sensitivity 93%
- Specificity 95%
Early data lead to rapid commercialization and marketing of a urine LAM assay

Detection of mycobacterial lipoarabinomannan with an antigen-capture ELISA in unprocessed urine of Tanzanian patients with suspected tuberculosis

C. Boehme, E. Melkova, F. Nina, S. Geis, T. Loscher, L. Maboko, V. Kaulich, M. Heeseker

231 patients with suspected pulmonary TB and 103 healthy volunteers were screened with standard TB tests and with the new LAM-ELISA.

Of 132 patients with positive sputum culture, 106 were positive using the LAM-ELISA (sensitivity 80.3%).

To define the specificity of the assay, urine samples from 103 healthy volunteers were also screened using LAM-ELISA. All but one had an optical density below the cut-off (specificity 99%).

Marketed in 2007/08 by Inverness Medical Innovations
Subsequent evidence from field studies in India, S Africa, Zimbabwe, Tanzania

- Rapid commercialization on the basis of early data may be problematic (especially case-control studies that can exaggerate accuracy estimates).
- Thorough field evaluation in diverse settings (e.g. varying HIV prevalence) should have been done.
- This case study raises an interesting question: at what point in time after a test is introduced should meta-analyses be done?
Case study 2:

How should we design and analyze diagnostic studies?

Serologic (antibody) tests for TB

- Attractive ... especially if point of care (POC) option
- >80 antigenic targets evaluated and several commercial assays developed
- All existing serologic tests have failed to demonstrate adequate accuracy
Sensitivity varies from 0 – 100%

![Graph of sensitivity varying from 0 to 100%](Image)

Steingart, CVI 2009

WHO evaluation of 19 commercial serological tests for TB

![ROC curve of commercial rapid tests for the diagnosis of pulmonary tuberculosis (all patients, n=256)](Image)

TDR/WHO Report 2008
Why do these tests fail in field studies?

A large % were case-control studies
Confirmed TB cases Vs. Healthy controls (often from low-incidence countries)

Spectrum bias (a form of selection bias)

- Population used for evaluating the test:
  - Extreme contrast
    - Case-control design
  - Normal contrast (Indicated population)
    - Consecutively recruited patients in whom the disease is suspected

- Extreme contrast (spectrum bias) can result in overestimation of test accuracy
We find this in TB as well:
Example: PCR tests for TB meningitis

Case-control studies had a two-fold higher diagnostic odds ratios than cross-sectional studies.
LED microscopy for sputum examination

- Cross Sectional Studies
  - Sensitivity 72.6%
    - (69.2, 75.8)
  - Specificity 96.9%
    - (92.1, 98.8)

- Case Control Studies
  - Sensitivity 88.7%
    - (81.4, 93.4)
  - Specificity 98.6%
    - (97.3, 99.3)

Analysis of diagnostic studies

- It is not uncommon to see researchers:
  1. Excluding patients or controls with no definitive diagnoses (“diagnostic myopia bias”)
  2. Excluding indeterminate or inconclusive results
  3. Performing post-hoc “discrepant” analysis to move numbers within 2 x 2 tables

- Such analyses often result in spuriously inflated accuracy estimates
Example: exclusion of indeterminates can inflate accuracy estimates

If indeterminates are included:
- Sens = 74%

If indeterminates are excluded:
- Sens = 86%

Abstract

The performance of the tuberculosis specific interferon gamma release assay (IGRA) has not been well-characterized in patients with pulmonary tuberculosis, and in indeterminate reactors. This study evaluated the utility of IGRA in a US and HIV-positive population and the effect of HIV infection and CD4 cell count on test performance.

Methods/Principal Findings: 145 patients with pulmonary tuberculosis (TB) were enrolled in the HIV and CD4 subset. Overall, 71 (49%) were HIV positive. The majority of patients were men (80%), white (85%), and without diabetes (96%) and had a mean CD4 cell count of 507 cells/μl. The rate of concordance with IGRA and the QuantiFERON-TB Gold test (QFT) was 85% (125/145). Using the QFT as the gold standard, the sensitivity for IGRA in HIV-negative patients was 60% (23/38); in HIV-positive patients it was 83% (14/17; OR: 3.6, 95% CI: 1.0–12.8; p = 0.06). Among HIV-positive QFT-positive patients the sensitivity of IGRA was 86% (14/16). The comparison of IGRA and QFT showed a negative predictive value of 85% (21/25) and positive predictive value of 71% (14/20). Conclusion/Significance: Sensitivity of the QFT for diagnosing active TB infection was reasonable when excluding indeterminate results and in HIV-negative patients. However, since the rate of indeterminate results was relatively high (13%), the IGRA for detecting active TB infection is not a substitute for the QFT. A higher proportion of patients with TB infection may be missed, especially in HIV-positive patients. IGRA results should not be used for alternative screening for active TB infection. IGRA is not recommended for the initial diagnosis of active TB infection.
Lessons

- Early case-control studies are often used to promote and market tests
- But a large proportion of tests fail, once they are used in real world settings (e.g., large number of failed commercial serological tests)
- Case-control studies exaggerate accuracy estimates, especially if the two-gate approach is used
- Certain data analytic approaches can also inflate accuracy estimates
- Diagnostic studies can begin as case-control studies, but need to move beyond that to prospective studies in clinically indicated populations
- Even accuracy data may be insufficient to decide on clinical impact
- Regulatory agencies should demand prospective data and not just rely on case-control accuracy studies

Case study 3:

Where should TB tests be evaluated and which populations are appropriate?
It is not uncommon to see TB test evaluations where:

- Cases come from a high-incidence country and controls from a low-incidence country
- Tests work well in a low-incidence country and fall apart in a high-incidence country
- Tests that work well in immunocompetent persons fail in populations with high HIV prevalence

Example: cases from Zambia and controls from England

<table>
<thead>
<tr>
<th></th>
<th>HIV- (n = 11)</th>
<th>HIV+ (n = 30)</th>
<th>HIV- (n = 54)</th>
<th>HIV+ (n = 21)</th>
<th>Healthy British adults (n = 46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuberculosis patients</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESA6 peptides</td>
<td>1.1 (100)</td>
<td>34 (87)</td>
<td>28 (52)</td>
<td>6 (29)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CFP10 peptides</td>
<td>8 (73)</td>
<td>25 (66)</td>
<td>34 (63)</td>
<td>7 (33)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Combined ESA6/CFP10 peptides</td>
<td>11 (100)</td>
<td>35 (90)</td>
<td>37* (69)*</td>
<td>9* (43)*</td>
<td>0 (0)</td>
</tr>
<tr>
<td>ESAT6 antigen</td>
<td>9 (82)</td>
<td>18 (46)</td>
<td>9 (17)</td>
<td>4 (19)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>PPD</td>
<td>11 (100)</td>
<td>28 (72)</td>
<td>43 (83)</td>
<td>6 (29)</td>
<td>33 (72)</td>
</tr>
<tr>
<td>TST</td>
<td>-</td>
<td>-</td>
<td>28/35** (80)</td>
<td>5/14** (36)</td>
<td>-</td>
</tr>
</tbody>
</table>

**P** value for difference 0.064, **P** value for difference 0.0057.
Lack of discrimination in TB endemic settings: example

Mycobacterium tuberculosis-specific ESAT-6 and CFP-10 antibody responses

Tanzanian cases

Danish controls

Figure 3. Dotplot showing the optical density (OD) values obtained from 195 patients with active tuberculosis disease who resided in northern Tanzania (121 TB and 74 healthy) and 145 Calmette-Guerin-vaccinated, Danish resident volunteers with no known risk factors for tuberculosis (EX-CG). The dotted line indicates the cutoff value calculated as the mean OD + 3 SDs for the 32 healthy Danish residents.

Hoff et al. Clinical Infectious Diseases 2007; 45:575–82

Variation in performance in high vs low endemic countries: example

T-cell interferon-γ release assays for the rapid immunodiagnosis of tuberculosis: clinical utility in high-burden vs. low-burden settings

Keertan Dheeraj1*, Richard van Zyl Smit1, Motosim Badri1 and Madhukar Pai1

High incidence countries

Low incidence countries

34
HIV can prove to be the acid test for any test! Example of MycoDot

MycoDot was hailed to be a breakthrough because it was a simple dipstick test.

Commercialized and marketed by Mossman Associates (with support of PATH)

Package insert: sensitivity of 70% and specificity of 95%

But when the test was evaluated in countries with high HIV prevalence, the performance was disastrous

Evaluation of the MycoDot™ test in patients with suspected tuberculosis in a field setting in Tanzania

G. A. Senk, R. A. Witten, G. M. Shemusya, Y. A. Boyo

Evaluation of a commercial immuno-diagnostic kit incorporating liposomal Q3 in the serodiagnosis of pulmonary tuberculosis in Ghana

Sens in HIV+ = 26% Sens in HIV+ = 25%

Despite these results, the test is still available on the market!
Lessons

- TB evaluation studies must be done in high TB incidence countries, especially in high HIV prevalent settings.

- Performance outcomes from low incidence countries may be deceptive and not reflect the performance in high incidence settings where the challenges include:
  - HIV
  - Severe TB
  - High background prevalence of TB infection
  - Widespread BCG vaccination
  - Malnutrition
  - Other diseases that can affect performance (e.g. worm infestations)

- If tests perform well in TB/HIV endemic countries, then they are likely to hold up well!

Case study 4:

Who should conduct TB diagnostic studies?
Industry involvement in drug trials and its impact on study outcomes and conclusions

Scope and Impact of Financial Conflicts of Interest in Biomedical Research
A Systematic Review

BMJ 2002

Pharmaceutical industry sponsorship and research outcome and quality: systematic review
BMJ 2003

Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized trials

BMJ 2003

Association between competing interests and authors’ conclusions: epidemiological study of randomized clinical trials published in the BMJ

CMAJ 2004

Industry involvement in diagnostic studies?

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


Table 2. Characteristics of the studies included (N = 50).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Frequency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td></td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>45 (90)</td>
</tr>
<tr>
<td>AIDS</td>
<td>10 (20)</td>
</tr>
<tr>
<td>HIV</td>
<td>27 (54)</td>
</tr>
<tr>
<td>Study origin</td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>18</td>
</tr>
<tr>
<td>Asia</td>
<td>20</td>
</tr>
<tr>
<td>Australia and Oceania</td>
<td>8 (16)</td>
</tr>
<tr>
<td>Europe</td>
<td>27</td>
</tr>
<tr>
<td>North America</td>
<td>19</td>
</tr>
<tr>
<td>South America</td>
<td>8</td>
</tr>
<tr>
<td>Number of patients per study</td>
<td></td>
</tr>
<tr>
<td>Under 1000</td>
<td>20 (40)</td>
</tr>
<tr>
<td>1000 to 5000</td>
<td>30 (60)</td>
</tr>
<tr>
<td>ICDR                        5000</td>
<td>5 (10)</td>
</tr>
<tr>
<td>Number of studies with conflict of interest</td>
<td>30 (60)</td>
</tr>
<tr>
<td>Year of publication</td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td>42 (84)</td>
</tr>
<tr>
<td>2005</td>
<td>18 (36)</td>
</tr>
<tr>
<td>2006</td>
<td>20 (40)</td>
</tr>
<tr>
<td>Number of journals where individual studies were published</td>
<td>40</td>
</tr>
</tbody>
</table>

About 40% of TB, HIV, Malaria diagnostic studies had industry involvement or known conflict of interest
Industry involvement in TB diagnostic studies: example from IGRA literature

Of the 38 studies in the meta-analysis, 21 (55%) had some sort of industry involvement or support, such as sponsorship, donation of test kits, participation in advisory boards, involvement of test developers, or ownership of patents.

Industry involvement in TB, HIV, Malaria studies and likely impact: McGill-TDR/WHO study

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry involvement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1.0</td>
<td>Reference group</td>
</tr>
<tr>
<td>Yes</td>
<td>4.28</td>
<td>1.83 - 10.02</td>
</tr>
<tr>
<td>NR</td>
<td>5.11</td>
<td>1.77 - 14.74</td>
</tr>
</tbody>
</table>
Industry involvement in TB studies and likely impact: commercial IGRAs

- We searched for cost-effectiveness studies on commercial IFN-gamma release assays

- We found a total of 10 studies

- Of these 6 studies had industry involvement of some sort
  - 2 of 6 had CEO of a test making company as author!

- Of the 6 studies with industry involvement: ALL concluded in favor of the commercial test and claimed superior cost-effectiveness

- Of the 4 independent studies, two were in favor of the test, and two were cautious and recommended a more selective use of the test
**Lessons**

- **When test developers do the studies, test performance is always good; performance is less optimal when others try to replicate the results**
  - may be suppression of unfavourable data
  - may just be a learning curve issue (test developers, by definition, understand the test better and know how to make it work!)

- While industry is critical for test development and commercialization, test evaluations should, ideally, be done independent of industry support

- At the very least, industry involvement should be clearly disclosed in all publications and presentations

- Industry and test developers should definitely not be involved in guideline and policy development
  - At least 17 countries have guidelines and statements on IGRA
  - Vast majority of these guidelines had no disclosures on conflicts of interest
Case study 5:

Can we trust the package insert?

Commercial package inserts always provide data on test accuracy: can we trust them?

According to the company, this test has 89% sensitivity in active TB

<table>
<thead>
<tr>
<th>Study</th>
<th>Quantiferon Gold CS</th>
<th>Quantiferon Gold SF</th>
<th>Ctrl Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>36</td>
<td>25</td>
<td>89% (113/126)</td>
</tr>
<tr>
<td>USA</td>
<td>27</td>
<td>31</td>
<td>83% (132/160)</td>
</tr>
<tr>
<td>Europe</td>
<td>36</td>
<td>25</td>
<td>79% (60/77)</td>
</tr>
<tr>
<td>Overall</td>
<td>99% (614/620)</td>
<td>89% (113/126)</td>
<td>79% (60/77)</td>
</tr>
</tbody>
</table>

Notes:
- Pos = Positive, Neg = Negative, Int = Indeterminate
- In the U.S. study of 113 active TB patients, 1 in 10 results were missing for 4 and invalid for 3.
Updated meta-analyses on sensitivity of QuantiFERON-TB Gold In Tube

Pooled estimate was about 79%

More examples...

<table>
<thead>
<tr>
<th>Test</th>
<th>Package insert sens</th>
<th>Package insert spec</th>
<th>Meta-analysis sens</th>
<th>Meta-analysis spec</th>
</tr>
</thead>
<tbody>
<tr>
<td>FASTPlaque-Response</td>
<td>96 – 100%</td>
<td>99 – 100%</td>
<td>95%</td>
<td>95%</td>
</tr>
<tr>
<td>Anda-TB IgG</td>
<td>85 - 90%</td>
<td>85 - 100%</td>
<td>60 - 75%</td>
<td>~90%</td>
</tr>
<tr>
<td>MycoDot</td>
<td>70%</td>
<td>95%</td>
<td>26% - 76%</td>
<td>84% - 97%</td>
</tr>
<tr>
<td>Clearview TB ELISA</td>
<td>81% (HIV+)</td>
<td>93 – 98%</td>
<td>56% (HIV+)</td>
<td>95%</td>
</tr>
<tr>
<td>GenoType MDTBDrplus</td>
<td>99%</td>
<td>99%</td>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td>Gen-Probe MTD</td>
<td>97% (S+)</td>
<td>100% (S+)</td>
<td>97% (S+)</td>
<td>96% (S+)</td>
</tr>
<tr>
<td></td>
<td>72 (S-)</td>
<td>99% (S-)</td>
<td>76% (S-)</td>
<td>95% (S-)</td>
</tr>
</tbody>
</table>
Lessons

- Company package inserts often present optimistic estimates based on small in-house evaluations that are usually sponsored by the companies.
- Lab professionals and clinicians must be critical of advertised estimates of accuracy and performance.
- Even when contradictory data are published, companies may not revise their package inserts or advertisements.
- There is very little post-marketing surveillance of diagnostics and devices.
- Regulatory agencies may not require companies to revise their package inserts.
- Poorly performing tests may, in fact, never get pulled off the market.

Case study 6:

Should we expect tests to be transferable and replicable?

Transferability: technologies that work well in the hands of developers will not necessarily work well everywhere.

Eg. MODS, phage assays.
MODS: developed in Peru – performs excellent

Sensitivity better than LJ (98 vs. 84%)

Fast turnaround time (1 week vs. 6 weeks+)

Implemented in India – performs poorly

Sensitivity 80%

Issues with contamination

Issues with reliability

Diagnostic accuracy of the microscopic observation drug susceptibility assay: a pilot study from India

J. S. Michael,* R. Dabry,* B. Kabeleho,* A. Lottha, J. Vijayakumar, D. Mathubu, K. R. Iysets, M. Phal

Simple, phage-based (FASTPlaque) technology to determine rifampicin resistance of Mycobacterium tuberculosis directly from sputum

H. A. Ghani,* A. Tolfis,* T. Sereetsi, A. R. Mole

*Wits Lambronics, University Hospital, Pretoria, South Africa

**National Laboratory of Tropical Medicine, Pasteur Institute, Paris, France

FASTPlaque assay – performed well when done by industry

100% sens

100% spec

Implemented in Kenya – performs poorly

Despite upgrading the lab:

Low accuracy (31% sens; 95% spec)

Issues with contamination (nearly impossible to interpret)

Evaluation of FASTPlaque® to diagnose smear-negative tuberculosis in a peripheral clinic in Kenya

M. Boutell,* L. Lungrad,* K. Varaklin,* A. Bukwy,* P. M. Kithinji,* J. Osoro,* G. P. O. Leach,* I. Osoro,*

1. Tuberculosis Unit, National Institute for Medical Research, Dar es Salaam, Tanzania

2. Department of Laboratory Medicine, University of Dar es Salaam, Tanzania

3. Tuberculosis Control Programme, Tanzania National Tuberculosis Programme, Ministry of Health and Social Welfare, Dar es Salaam, Tanzania

4. Lung and TB Research Institute, University of California Los Angeles, Los Angeles, California, USA

5. Centre for the Study of Drug Resistance in Tuberculosis, London School of Hygiene and Tropical Medicine, London, UK

6. University of Sydney, Sydney, Australia

OBJECTIVES: To evaluate the performance and feasibility of FASTPlaque™ to diagnose smear-negative tuberculosis (TB) among rural children and children in urban slums.

METHODS: A total of 204 specimens were obtained from smear-negative children with a history of TB and enrolled in the study. Each patient was tested with both smear microscopy and FASTPlaque to determine sputum bacilli. The FASTPlaque was performed by fast tracking and drug sensitivity testing.
Replication

- There are many examples of novel tests for TB that show great promise, but do not get replicated

- Or subsequent results are disappointing and commercialization is abandoned

- Or test may be quite good, but impossible to develop and manufacture in a cost-effective way

- Results in a graveyard of inexplicably abandoned diagnostics

Example: MPB64 skin patch test (Sequella Inc.)

Early data in 1998:

- Sensitivity: 98%
- Specificity: 100%

In 2010, still not commercially available — plans have been abandoned
Lessons

- Many novel tests and tools are bound to fail
- We need to appreciate the “failure rate” of new tests and interventions
- Replication, in diverse settings, is required, before proceeding with commercialization and clinical use
- Transferability of technologies must receive attention; tests need to be robust if they have to work well in all settings
- Tests that work well in the hands of developers may not work well in field settings, especially in resource-limited countries
- Single studies are never sufficient for policy and guideline development; we need more extensive evidence
- Even accuracy data are not sufficient for evidence-based policies
Will Xpert MTB/RIF survive “optimism bias”?

- Validation data and early demonstration data look very good
- Not much “real world” experience in resource-limited and routine programmatic settings
- Impact of “point-of-treatment” use is not demonstrated