DIAGNOSTIC RCTS

Madhukar Pai
Associate Professor
McGill University, Montreal

madhukar.pai@mcgill.ca

VERY FEW DIAGNOSTIC RCTS IN TB

- Whatever little we have, is all new!!
  - A few on active case detection
  - A trial on same-day smear diagnosis (TDR)
  - A trial on Xpert MTB/RIF in South Africa
  - Ongoing trials on Xpert MTB/RIF
  - Cluster-randomized stepped wedge designs for phased implementation

Comparison of two active case-finding strategies for community-based diagnosis of symptomatic smear-positive tuberculosis and control of infectious tuberculosis in Harare, Zimbabwe (DETECTB): a cluster-randomised trial

Elizabeth L Corbett, Tsitsi Bandason, Trinh Duong, Ethel Dauya, Beauty Makamure, Gavin J Churchyard, Brian G Williams, Shungu S Munyati, Anthony E Butterworth, Peter R Mason, Stanley Mungofa, Richard J Hayes

Lancet 2010; 376: 1244–53
Published Online October 4, 2010
DOI:10.1016/S0140-6736(10)61425-0

See Comment page 1205

Clinical Research Unit (E L Corbett PhD, Prof A E Butterworth PhD) **and Infectious Disease Epidemiology Unit** (T Duong MSc, Prof R J Hayes DSc), **London School of Hygiene and Tropical Medicine, London, UK; Biomedical Research and Training Institute, Harare, Zimbabwe** (E L Corbett, T Bandason MSc, E Dauya BCom, B Makamure BTech, S S Munyati MSc, Prof A E Butterworth, Prof P R Mason PhD); **Aurum Institute, Johannesburg, South Africa** (Prof G J Churchyard PhD); **South African Centre for Epidemiology and Mathematical Analysis (SACEMA), University of Stellenbosch, South Africa** (Prof B G Williams PhD); **Department of Medical Laboratory Sciences, College of Health Sciences, University of**

Summary

Background Control of tuberculosis in settings with high HIV prevalence is a pressing public health priority. We tested two active case-finding strategies to target long periods of infectiousness before diagnosis, which is typical of HIV-negative tuberculosis and is a key driver of transmission.

Methods Clusters of neighbourhoods in the high-density residential suburbs of Harare, Zimbabwe, were randomised to receive six rounds of active case finding at 6-monthly intervals by either mobile van or door-to-door visits. Randomisation was done by selection of discs of two colours from an opaque bag, with one disc to represent every cluster, and one colour allocated to each intervention group before selection began. In both groups, adult (≥16 years) residents volunteering chronic cough (≥2 weeks) had two sputum specimens collected for fluorescence microscopy. Community health workers and cluster residents were not masked to intervention allocation, but investigators and laboratory staff were masked to allocation until final analysis. The primary outcome was the cumulative yield of smear-positive tuberculosis per 1000 adult residents, compared between intervention groups; analysis was by intention to treat. The secondary outcome was change in prevalence of culture-positive tuberculosis from before intervention to before round six of intervention in 12% of randomly selected households from the two intervention groups combined; analysis was based on participants who provided sputum in the two prevalence surveys. This trial is registered, number ISRCTN84352452.

Findings 46 study clusters were identified and randomly allocated equally between intervention groups, with 55 741 adults in the mobile van group and 54 691 in the door-to-door group at baseline. HIV prevalence was 21% (1916/9060) and in the 6 months before intervention the smear-positive case notification rate was 2·8 per 1000 adults per year. The trial was completed as planned with no adverse events. The mobile van detected 255 smear-positive patients from 5466 participants submitting sputum compared with 137 of 4711 participants identified through door-to-door visits (adjusted risk ratio 1·48, 95% CI 1·11–1·96, p=0·0087). The overall prevalence of culture-positive tuberculosis declined from 6·5 per 1000 adults (95% CI 5·1–8·3) to 3·7 per 1000 adults (2·6–5·0; adjusted risk ratio 0·59, 95% CI 0·40–0·89, p=0·0112).

Interpretation Wide implementation of active case finding, particularly with a mobile van approach, could have rapid effects on tuberculosis transmission and disease.

Lancet 2010

Twelve-monthly versus six-monthly radiological screening for active case-finding of tuberculosis: a randomised controlled trial

Gavin J Churchyard,[1,2] Katherine Fielding,[2] Surita Roux,[1] Elizabeth L Corbett,[3] Richard E Chaisson,[4] Kevin M De Cock,[3] Richard J Hayes,[2] Alison D Grant[3]

[1]Aurum Institute for Health Research, Johannesburg, South Africa
[2]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK
[3]Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK
[4]Center for TB Research, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence to
Professor G J Churchyard, Aurum Institute for Health Research, PostNet Suite 300, Private Bag X30500, Houghton 2041, South Africa; gchurchyard@auruminstitute.org

Received 18 March 2010
Accepted 7 October 2010

ABSTRACT

Background The incidence of tuberculosis has increased among South African gold miners despite comprehensive control programmes, including a radiological screening programme. No data are available as to the optimal frequency of screening. The aim of this study was to compare 6-monthly and 12-monthly radiological screening for active tuberculosis case-finding.

Methods Employees of a gold mining company were randomly assigned to the control arm (screening at baseline, 12 and 24 months) or the intervention arm (additional 'intervention' radiographs at 6 and 18 months after baseline). Study outcomes included proportion of tuberculosis cases detected by screening, proportion smear-positive, extent of disease and mortality.

Results 22 634 miners were randomised. Compared with 12-monthly screening, 6-monthly screening detected more tuberculosis suspects but not more cases, partly due to greater attrition between screening and further investigation after 'intervention' compared with routine radiographs. Tuberculosis cases detected in the 6-monthly versus the 12-monthly screening arm had less extensive disease ($p=0.05$) and a lower tuberculosis-specific mortality (death on tuberculosis treatment) (2.1 and 2.8 per 1000 person-years respectively, HR 0.73, 95% CI 0.50 to 1.08, $p=0.1$), which was most marked in the first 2 months of treatment (HR 0.48, 95% CI 0.23 to 0.98, $p=0.04$) when death from tuberculosis is most likely.

Discussion In settings with a high prevalence of HIV and tuberculosis despite standard tuberculosis control measures, more frequent case-finding may reduce the extent of disease, tuberculosis mortality and tuberculosis transmission through earlier detection of active tuberculosis cases. To be effective, however, all tuberculosis suspects identified through screening must be investigated for tuberculosis.

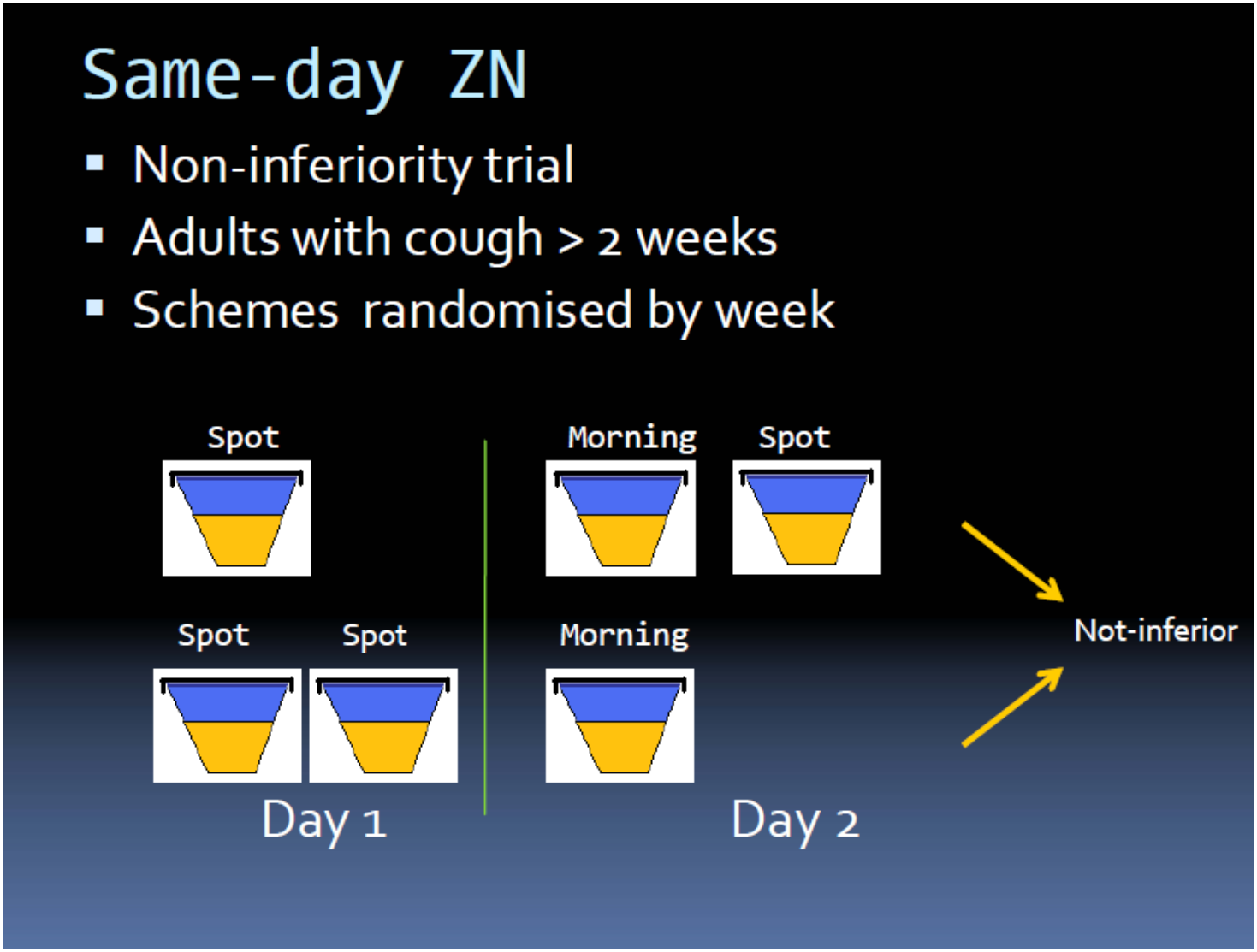
programmes[6–9] that include active case-finding using radiological screening as well as passive case-finding and treatment with fixed-dose combination tables taken under direct observation for the entire treatment period, tuberculosis rates among miners rose during the 1990s to over 3000 per 100 000 per year by 1999.[8] Silica dust exposure and silicosis are both risk factors for tuberculosis.[10–17] Silicosis occurs commonly among gold miners, so that miners now have a high prevalence of two of the most powerful risk factors for developing tuberculosis disease following infection (silicosis and HIV), and their combined effect is multiplicative.[16]

Radiological screening has been used in the gold mining industry for decades[18]; both 6- and 12-monthly radiological screening were used in different companies, with no data concerning the most effective screening frequency, particularly in the context of high HIV prevalence. Earlier detection of tuberculosis could reduce tuberculosis-specific morbidity, particularly chronic lung sequelae, mortality and the duration of infectiousness. In support of this, observational studies have demonstrated significantly less extensive radiological disease[19] and lower case fatality rates among tuberculosis cases detected by the radiological screening programme compared with those who self-presented with symptoms.[7] However, no trial has previously investigated the optimal frequency of active tuberculosis case-finding using radiological screening.

The aim of this trial was to compare the individual level effect of 6-monthly versus 12-monthly radiological screening on the proportion of tuberculosis cases detected by screening, the proportion who were smear-positive, the extent of the disease and mortality.

Thorax 2010

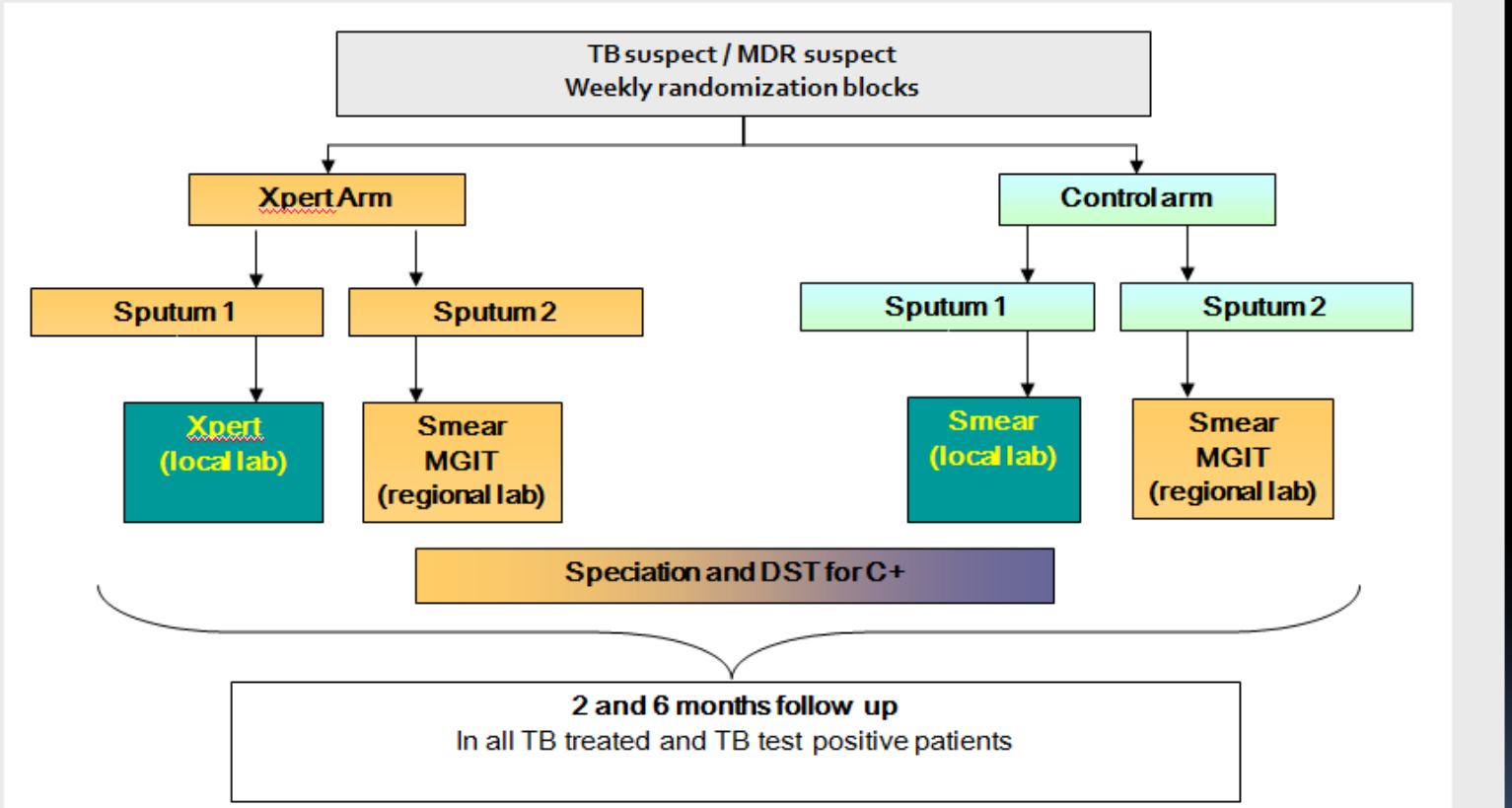
NON-INFERIORITY RCT ON SAME-DAY MICROSCOPY



Cuevas L et al. PLoS Med (in press)

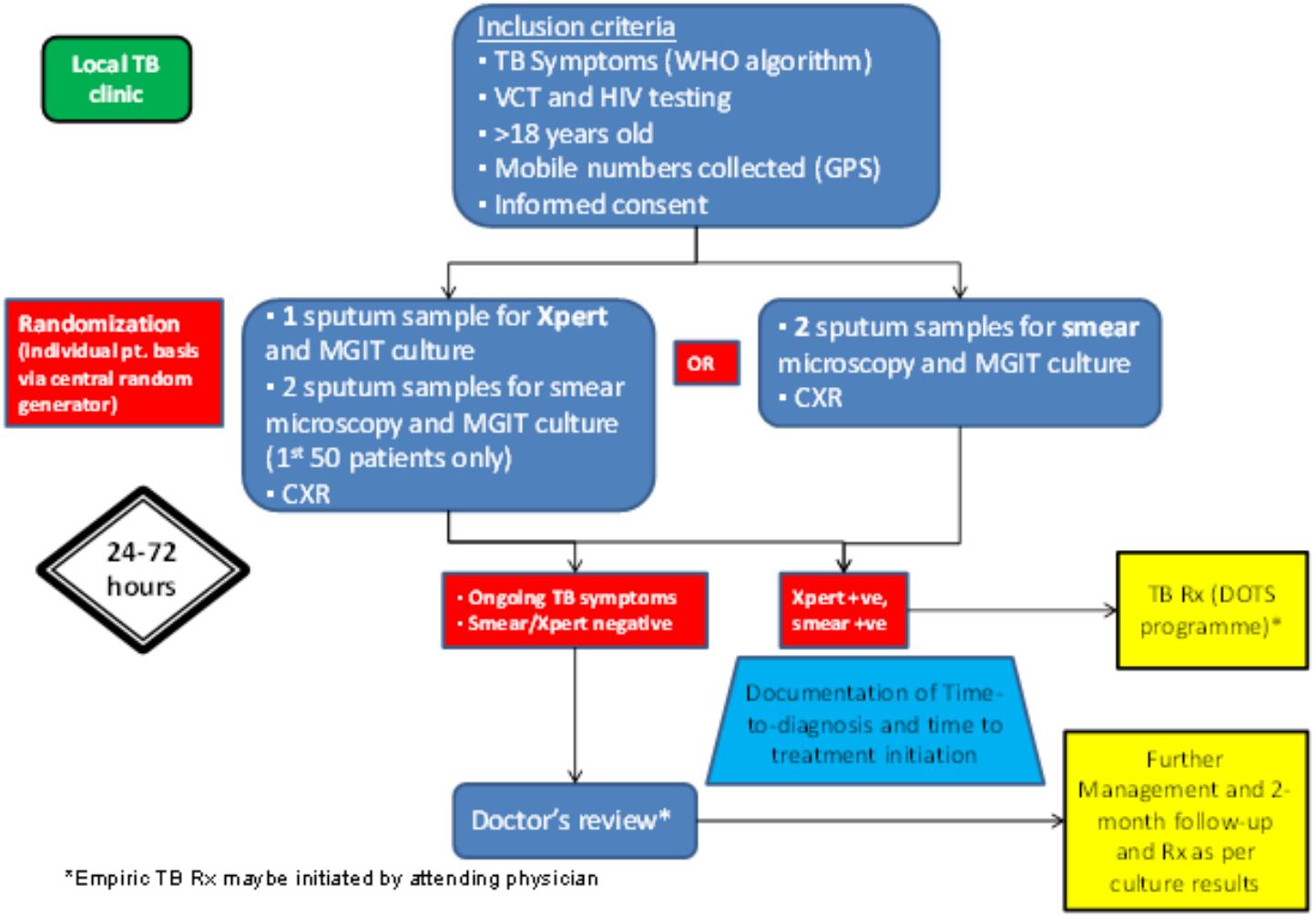
FIND-SUPPORTED DEMONSTRATION STUDY

Cape Town Xpert MTB/RIF Demo Study Design: Phase I



Nicol M et al. Union Conference, Berlin 2010

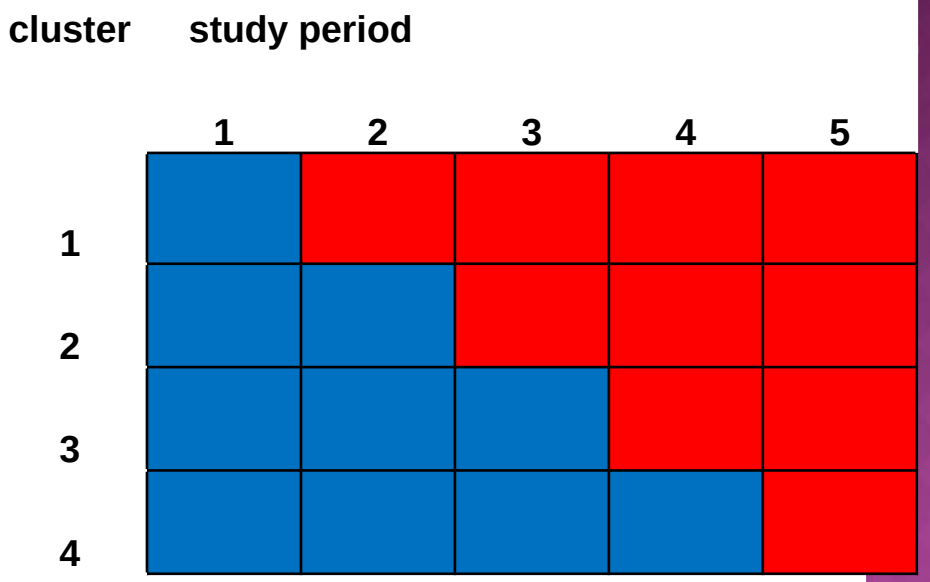
A RANDOMISED CONTROLLED TRIAL OF POINT-OF-TREATMENT GENEXPERT MTB/RIF ASSAY FOR THE DIAGNOSIS OF TB AT PRIMARY CARE CLINICS IN HIGH HIV PREVALENCE RESOURCE LIMITED SETTINGS



TB NEAT consortium. Cape Town, Zimbabwe and Lusaka [Dheda K et al].

STEPPED WEDGE DESIGN

- Unidirectional cross-over
- All transfer from control to intervention, but at different times
- Randomization of order of transfer
  - More than one cluster may transfer at once
  - Restrained in case of small number of clusters
- Need even more time than in regular cross-over design

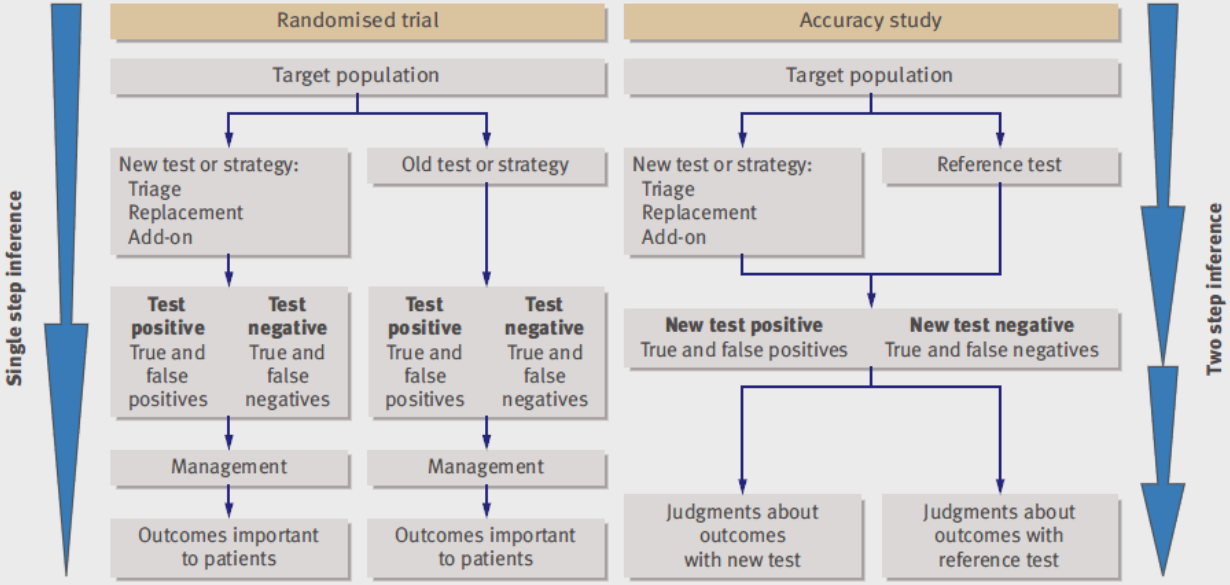


EXAMPLE STEPPED WEDGE DESIGN

- Roll-out of GeneXpert in Brazil
  - Rio de Janeiro and Manaus
- 14 clusters = defined areas in which all health care units send in samples to same laboratory for diagnosis of TB
- GeneXpert used instead of smear microscopy (and culture)
- Transfer from smear to GeneXpert in 7 steps
- Main effects studied:
  - Diagnosis and treatment registration of bacteriologically confirmed pulmonary TB, overall and for HIV+ individuals
  - Proportion of TB patients diagnosed with MDR
  - Time to appropriate treatment after diagnosis for (MDR) TB patients
  - Cost-effectiveness
  - (Treatment outcomes)

METHODOLOGICAL ISSUES IN DX RCTS

GRADE: FOR HIGH QUALITY EVIDENCE, IMPACT ON PATIENT-IMPORTANT OUTCOMES NEEDS TO BE DEMONSTRATED



Example
Randomised control trials (RCTs) explored a diagnostic strategy guided by the use of B type natriuretic peptide (BNP)—designed to aid diagnosis of heart failure—compared with no use of BNP in patients presenting to the emergency department with acute dyspnoea.[8 9] As it turned out, the group randomised to receive BNP spent a shorter time in the hospital at lower cost, with no increased mortality or morbidity

Example
Consistent evidence from well designed studies shows fewer false negative results with non-contrast helical computed tomography (CT) than with intravenous pyelography (IVP) in the diagnosis of suspected acute urolithiasis.[10] However, the stones in the ureter that CT detects but IVP "misses" are smaller, and hence are likely to pass more easily. As RCTs evaluating the outcomes in patients treated for smaller stones are not available, the extent to which reduction in cases that are missed (false negatives) and follow-up of incidental findings unrelated to renal calculi with CT have important health benefits remains uncertain[11]

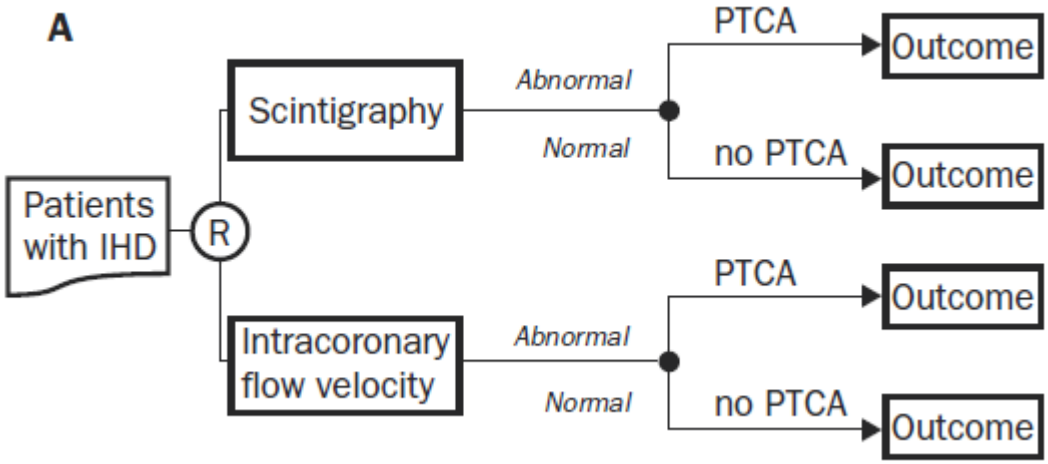
Two generic ways in which a test or diagnostic strategy can be evaluated. On the left, patients are randomised to a new test or strategy or to an old test or strategy. Those with a positive test result (cases detected) are randomised (or were previously randomised) to receive the best available management (second step of randomisation for management not shown). Investigators evaluate and compare patient-important outcomes in all patients in both groups.[6] On the right, patients receive both a new test and a reference test (old or comparator test or strategy). Investigators can then calculate the accuracy of the test compared with the reference test (first step). To make judgments about importance to patients of this information, patients with a positive test (or strategy) in either group are (or have been in previous studies) submitted to treatment or no treatment; investigators then evaluate and compare patient-important outcomes in all patients in both groups (second step)

Schunemann H et al. BMJ 2008

Viewpoint

Randomised comparisons of medical tests: sometimes invalid, not always efficient

Patrick M M Bossuyt, Jeroen G Lijmer, Ben W J Mol



- Not just the test, but test-treatment combination that is evaluated
- There must be a clear, pre-specified link between test results and subsequent interventions

Running trials without a protocol for translating the test results to clinical management decisions is like putting pharmaceuticals to trial without prespecifying the preferred dosage, optimum route of administration, the need for monitoring, or the way to deal with side-effects. Designing such a drug trial would be hardly acceptable these days. Why then should we not apply the same stringent criteria to trials of test-treatment combinations?

Lancet 2000

Distraction From Randomization in Diagnostic Research

CORNELIS J. BIESHEUVEL, PHD, DIEDERICK E. GROBBEE, MD, PHD,
AND KAREL G.M. MOONS, PHD

The ultimate goal of medical care, including diagnostic testing, is to improve patient outcome. Accordingly, it has been advocated widely that when establishing a test's diagnostic accuracy, the impact of the test on patient outcome subsequently must be quantified. When studying patient outcome in medical research, the use of randomized comparisons comes into perspective. In our view, randomized studies often are not necessary to validly estimate the effect of the diagnostic test on patient outcome. Results of cross-sectional diagnostic studies, combined with results from therapeutic studies, often will suffice.
Ann Epidemiol 2006;16:540–544.

Diagnostic questions are not etiologic questions, but randomization transforms a diagnostic study into an etiologic or intervention study, which may not be necessary in many instances.

DIAGNOSTIC RCT: IS IT REALLY DIAGNOSTIC?

When performing a randomized trial to determine the impact of a diagnostic test or strategy on patient outcome, an initially *diagnostic* research question is transformed into *therapeutic* research question (with the goal of establishing causality) with corresponding consequences for the design of the study. A disadvantage of a randomized approach to directly quantify the contribution of a diagnostic test and treatment on patient outcome is that it often addresses diagnosis and treatment as a single combined strategy, a "package deal." This makes it impossible to determine afterwards whether a positive effect on patient outcome was attributed solely to the improved diagnosis by using the test under study or to the chosen (new) treatment strategies.

Moons KGM. In: Grobbee & Hoes. Clinical Epidemiology. 2009

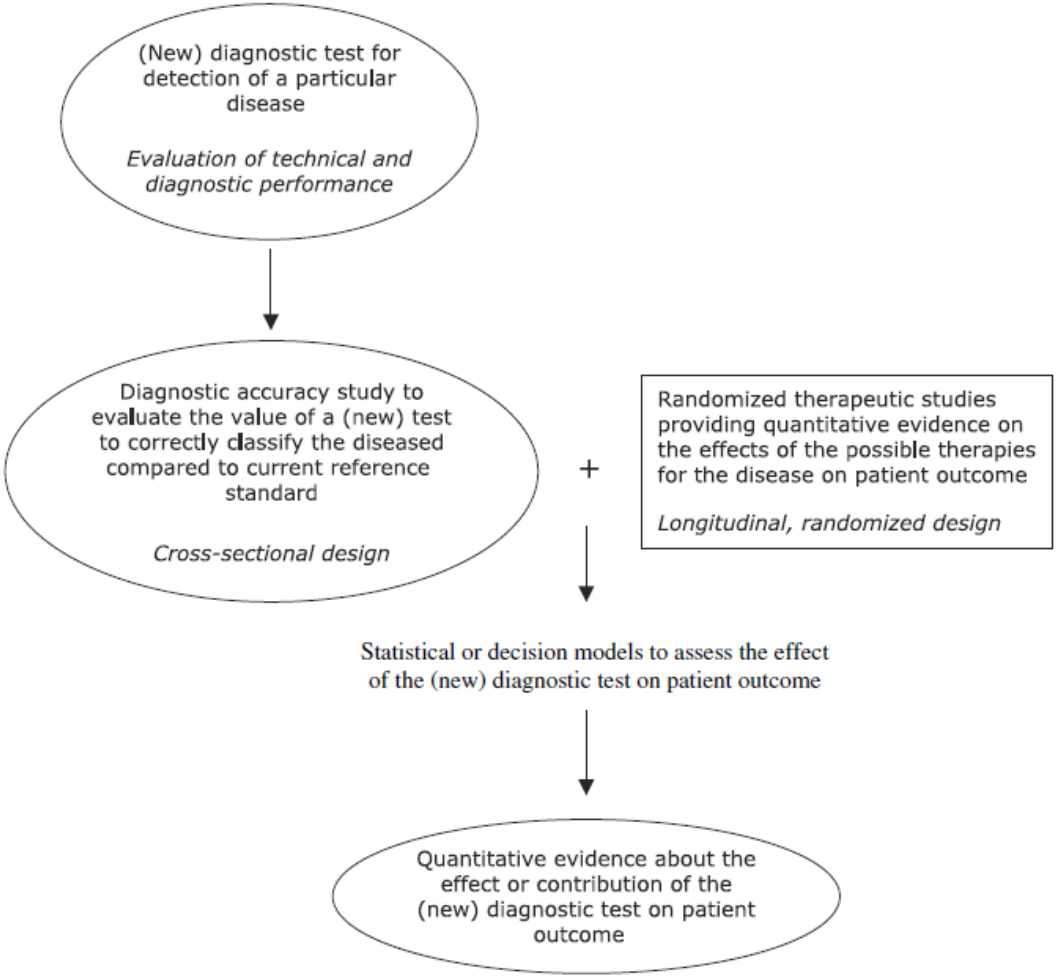


FIGURE 1. In the first phases of evaluating a new test, technical feasibility and diagnostic performance (is the test able to discriminate diseased from nondiseased?) are assessed (1–4). Subsequently, the value of the test to detect a particular disease is obtained by a cross-sectional study in which the diagnostic accuracy of the new test is compared with the current reference test. The information obtained by this study can be combined with information obtained from randomized controlled trials that evaluated the effect of a therapeutic strategy for that disease on patient outcome. By combining results of both studies, one can easily quantify the value of the diagnostic test to improve patient outcome.

Ann Epidemiol 2006;16:540–544.

Annals of Internal Medicine | Academia and Clinic

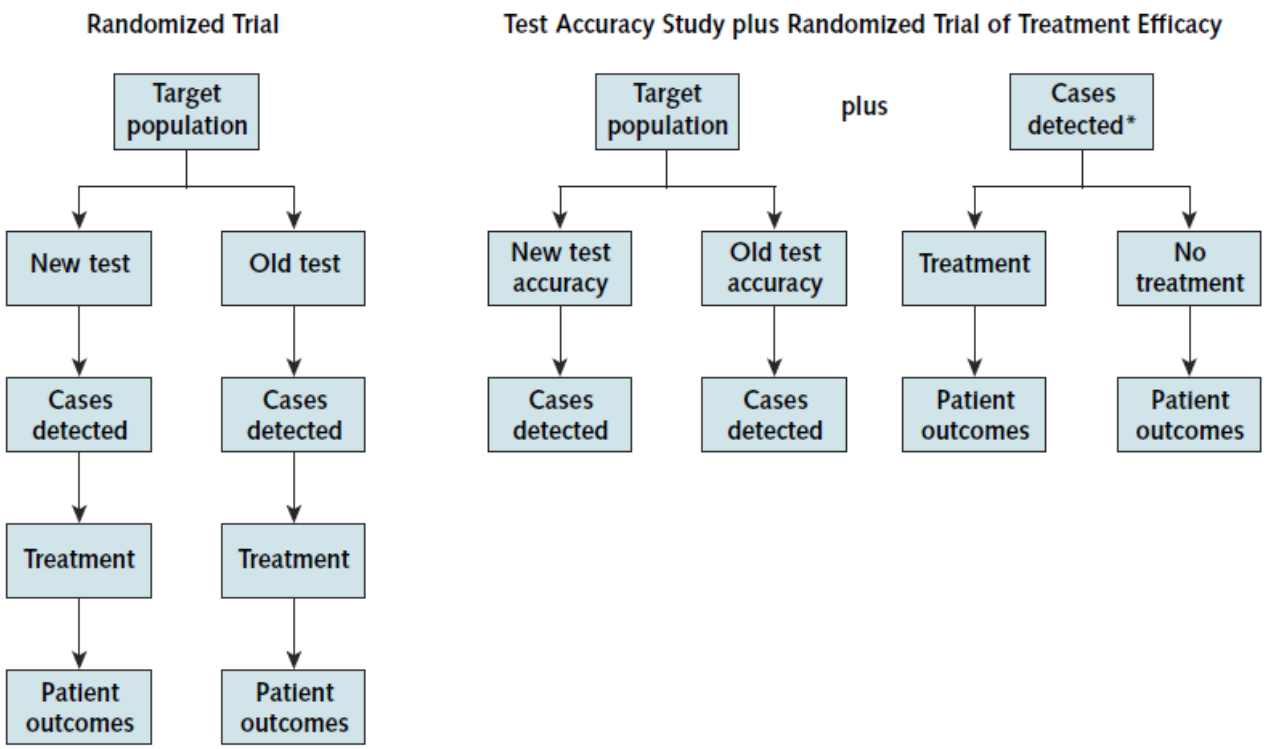
When Is Measuring Sensitivity and Specificity Sufficient To Evaluate a Diagnostic Test, and When Do We Need Randomized Trials?

Sarah J. Lord, MBBS, MS; Les Irwig, MBBCh, PhD; and R. John Simes, MBBS, MS, MD

The clinical value of using a new diagnostic test depends on whether it improves patient outcomes beyond the outcomes achieved using an old diagnostic test. When can studies of diagnostic test accuracy provide sufficient information to infer clinical value, and when do clinicians need to wait for results from randomized trials? The authors argue that accuracy studies suffice if a new diagnostic test is safer or more specific than, but of similar sensitivity to, an old test. However, if a new test is more sensitive than an old test, it leads to the detection of extra cases of disease. Results from treatment trials that enrolled only patients detected by the old test may not apply to these extra cases. Clinicians need to wait for results from randomized trials assessing treatment efficacy in cases detected by the new diagnostic test, unless they can be satisfied that the new test detects the same spectrum and subtype of disease as the old test or that treatment response is similar across the spectrum of disease.

Ann Intern Med. 2006;144:850-855. www.annals.org
For author affiliations, see end of text.

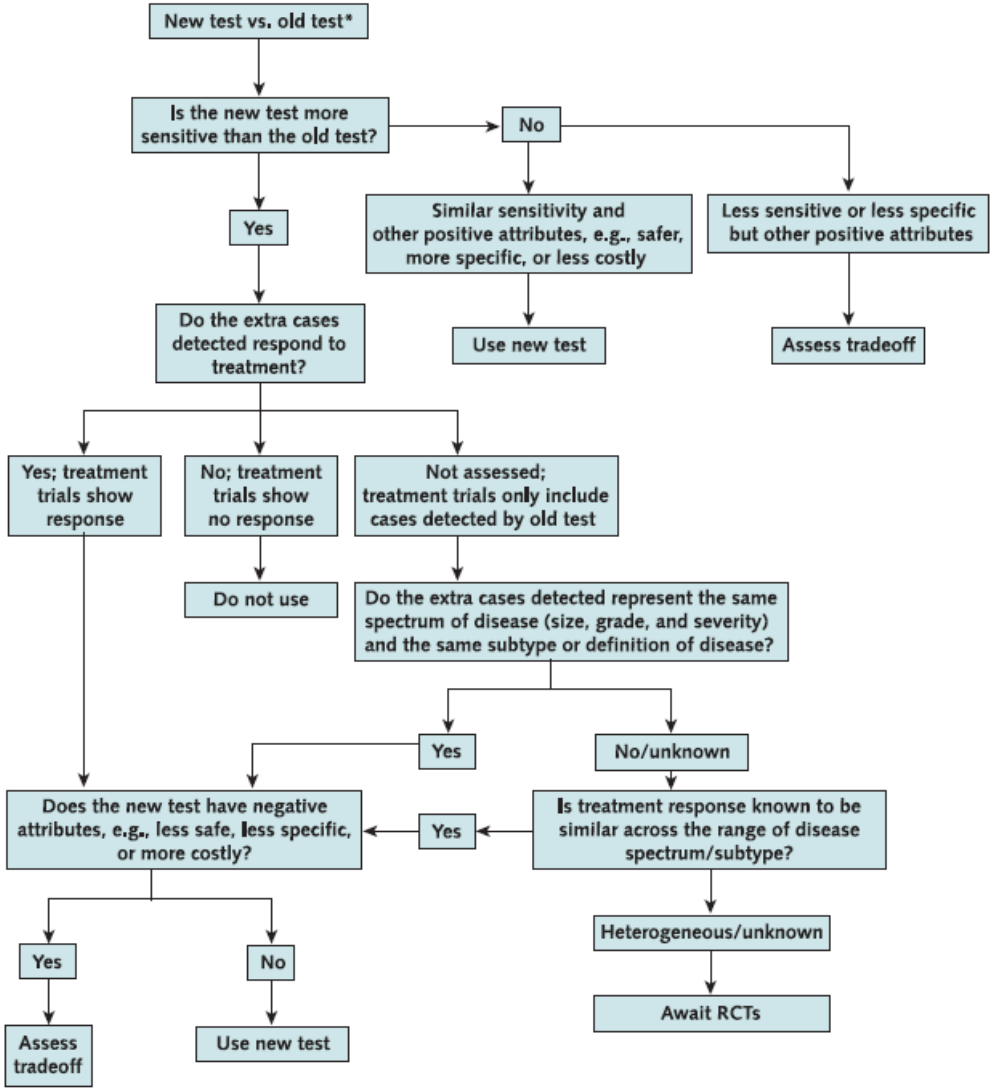
Figure 1. Trial evidence versus linked evidence of test accuracy and treatment efficacy.



*Cases detected by the new and old test may not show similar response to treatment.

Lord, Annals Int Med, 2006

RCT ALWAYS NEEDED?



Lord, Annals Int Med, 2006

Table. **New Diagnostic Test Assessment Framework and Examples***

Test and Indication	Proposed Benefits of New Diagnostic Test	Does the Evidence of Effective Treatment Apply to the Cases Detected by the New Test?	Can Studies of Test Accuracy Suffice?
Example 1: Doppler ultrasonography vs. venography for detection of deep venous thrombosis	Safer, less costly	Yes; no change to the definition or spectrum of disease.	Yes; the value of the new test may be inferred from an assessment of its relative safety and cost.
Example 2: new vs. standard FOBT for early detection of colorectal cancer	More specific	Yes; no change to the definition or spectrum of disease.	Yes; the value of the new test may be inferred from an assessment of its relative safety and cost and the benefits of avoiding a false-positive result.
Example 3: supine and prone positioning for CT colonography vs. supine positioning alone for detection of adenomatous colorectal polyps	More sensitive	Yes; no change to the definition or spectrum of disease because cases detected by both tests are subsequently confirmed by colonoscopy.	Yes; the value of the new test may be inferred from an assessment of its relative sensitivity, given no substantial loss in specificity and existing trial evidence that treatment improves patient outcomes.
Example 4: NMR spectroscopy vs. plasma lipoprotein levels for the detection of hyperlipidemia	More sensitive	Unknown; it is unknown whether NMR detects patients who would benefit more or less from lipid-lowering agents than those whose disorder was detected using the existing test.	No; trial evidence is required to determine the value of the new test.
Example 5: MRI vs. mammography for earlier detection of invasive breast cancer in women at high risk for the disease	More sensitive	No; it is uncertain if any benefits of treatment at the earlier stage of disease outweigh the harm of overdetection of cancer that would never have presented clinically.	No; trial evidence is required to determine the impact of testing on patient outcomes or, at least, the interval breast cancer rate.
Example 6: PET, MRI, and EEG vs. MRI and EEG to detect an epileptogenic focus in patients with medically refractory epilepsy who are being considered for surgery	More sensitive	Uncertain; it is uncertain whether patients with functional lesions detected by PET for whom existing standard imaging yields negative or inconclusive results will show the same treatment response to surgery as patients with structural lesions that can be detected with standard imaging alone.	Judgment is needed about whether clinicians require randomized trials to assess treatment response in the extra cases detected by PET or whether they can rely on existing trials conducted in patients with disease detected by MRI and EEG and observational evidence about treatment response in cases detected by PET.

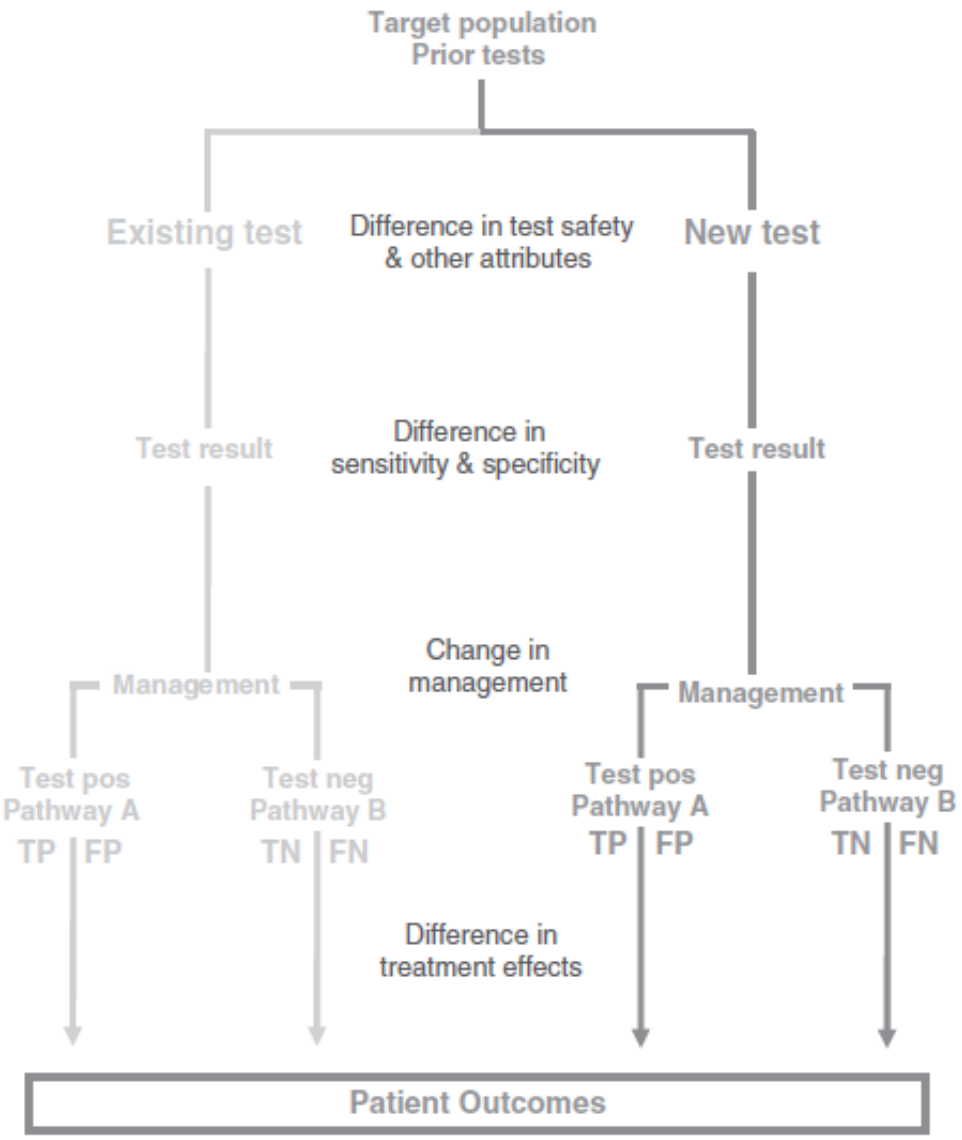
* CT = computed tomography; EEG = electroencephalography; FOBT = fecal occult blood test; MRI = magnetic resonance imaging; NMR = nuclear medicine resonance; PET = positron emission tomography.

Using the Principles of Randomized Controlled Trial Design to Guide Test Evaluation

Sarah J. Lord, MBBS, MS, Les Irwig, MBBCh, PhD, Patrick M. M. Bossuyt, PhD

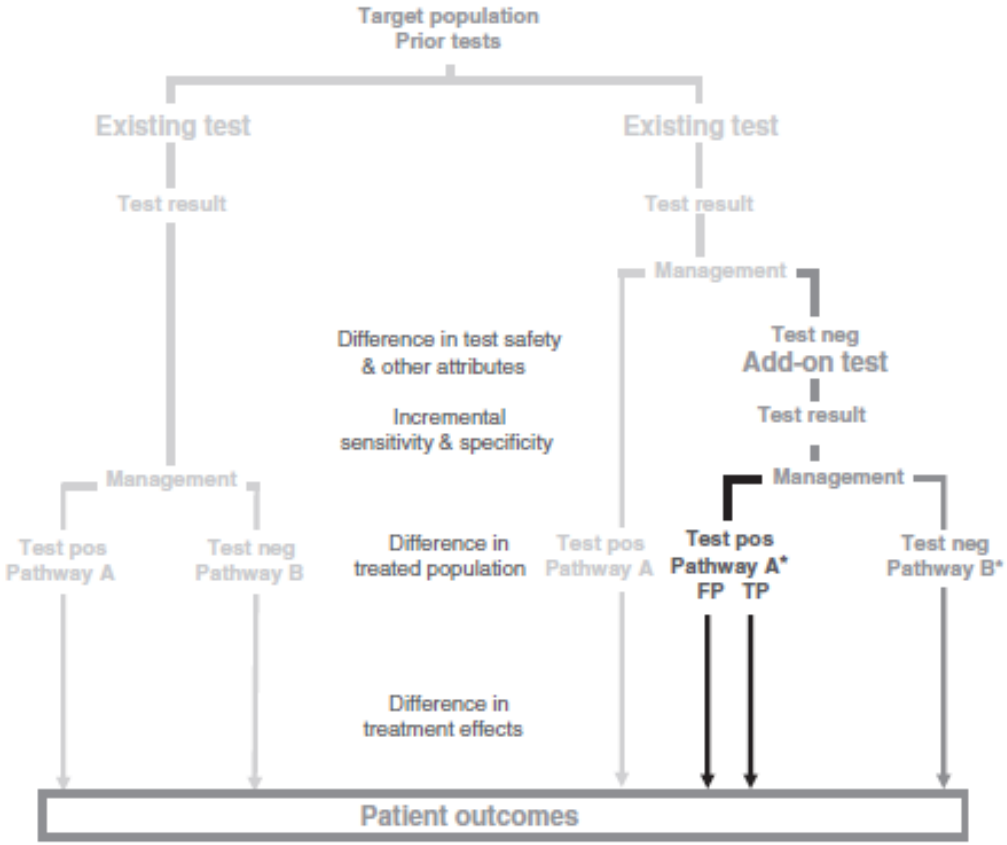
The decision to use a new test should be based on evidence that it will improve patient outcomes or produce other benefits without adversely affecting patients. In principle, long-term randomized controlled trials (RCTs) of test-plus-treatment strategies offer ideal evidence of the benefits of introducing a new test relative to current best practice. However, long-term RCTs may not always be necessary. The authors advocate using the hypothetical RCT as a conceptual framework to identify what types of comparative evidence are needed for test evaluation. Evaluation begins by stating the major claims for the new test and determining whether it will be used as a replacement, add-on, or triage test to achieve these claims. A flow diagram of this hypothetical RCT is constructed to show the essential design elements, including population, prior tests, new test and existing test strategies, and primary and secondary outcomes. Critical steps in the pathway between testing and patient outcomes, such as differences in test accuracy, changes in treatment, or avoidance of other tests, are displayed for each test strategy. All differences between the tests at these critical steps are identified and prioritized to determine the most important questions for evaluation. Long-term RCTs will not be necessary if it is valid to use other sources of evidence to address these questions. Validity will depend on issues such as the spectrum of patients identified by the old and new test strategies. ***Key words:*** *diagnostic techniques and procedures/standards; sensitivity and specificity; randomized controlled trials as topic; outcome assessment (health care).* ***(Med Decis Making 2009;29:E1–E12)***

a. The replacement test



b. The add-on test

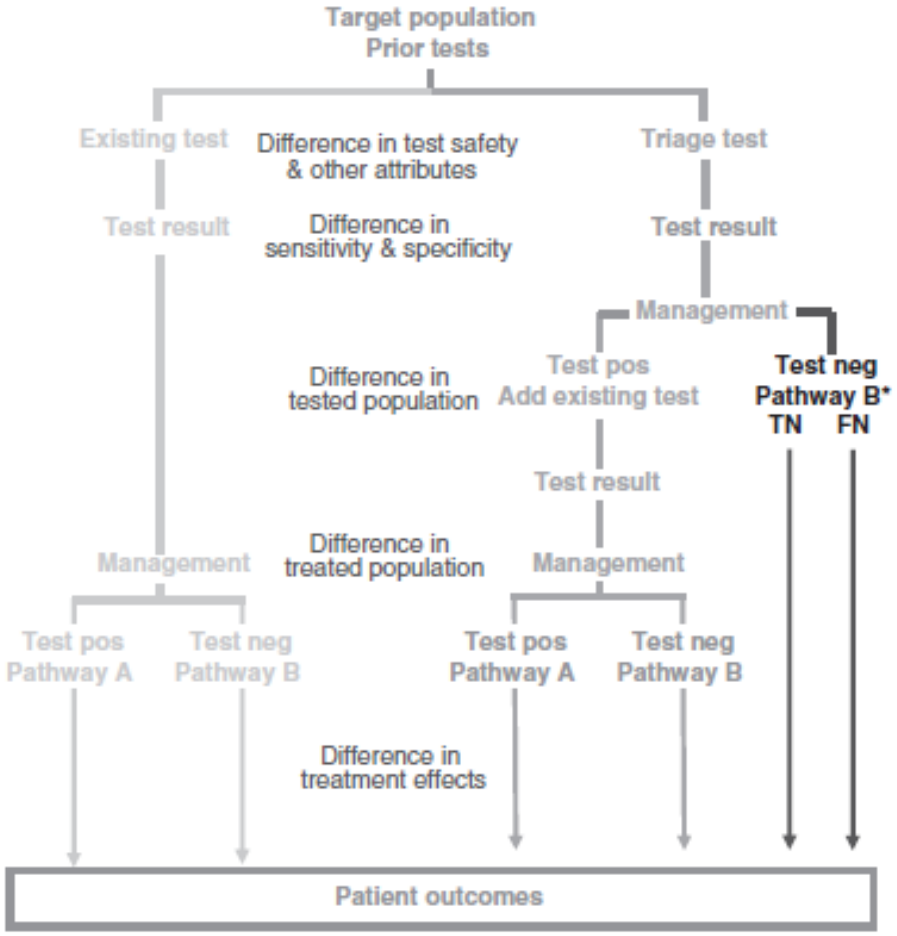
Difference in test-treatment pathway using add-on test shown in black



TP = true positive, FP = false positive
Pathway A* includes patients testing positive on the add-on test but negative on the existing test who would not have been assigned to treatment A using the existing test strategy.

c. The triage test

Difference in test-treatment pathway using triage test shown in black



SOME ISSUES..

- Actions after each possible test results must follow a clear unambiguous protocol
- Sample sizes may be large
- Diversity and complexity of diagnostic process leads to infinite number of possible trials
- Ethical questions
  - OK to randomize to an experimental test?
    - E.g. Not FDA approved
  - Once a test is WHO-approved and is shown to be superior to conventional test, OK to deny to half of trial participants?

Mariska Leeflang
m.m.leeflang@amc.uva.nl

BIAS IN RCTS

1. Randomization
   a. Valid randomization
   b. Concealment of allocation
2. Blinding
3. Sufficiently long follow-up
4. Analyses