FIND

Because diagnosis matters

Basic anatomy of a demonstration study

Catharina Boehme

December 2014

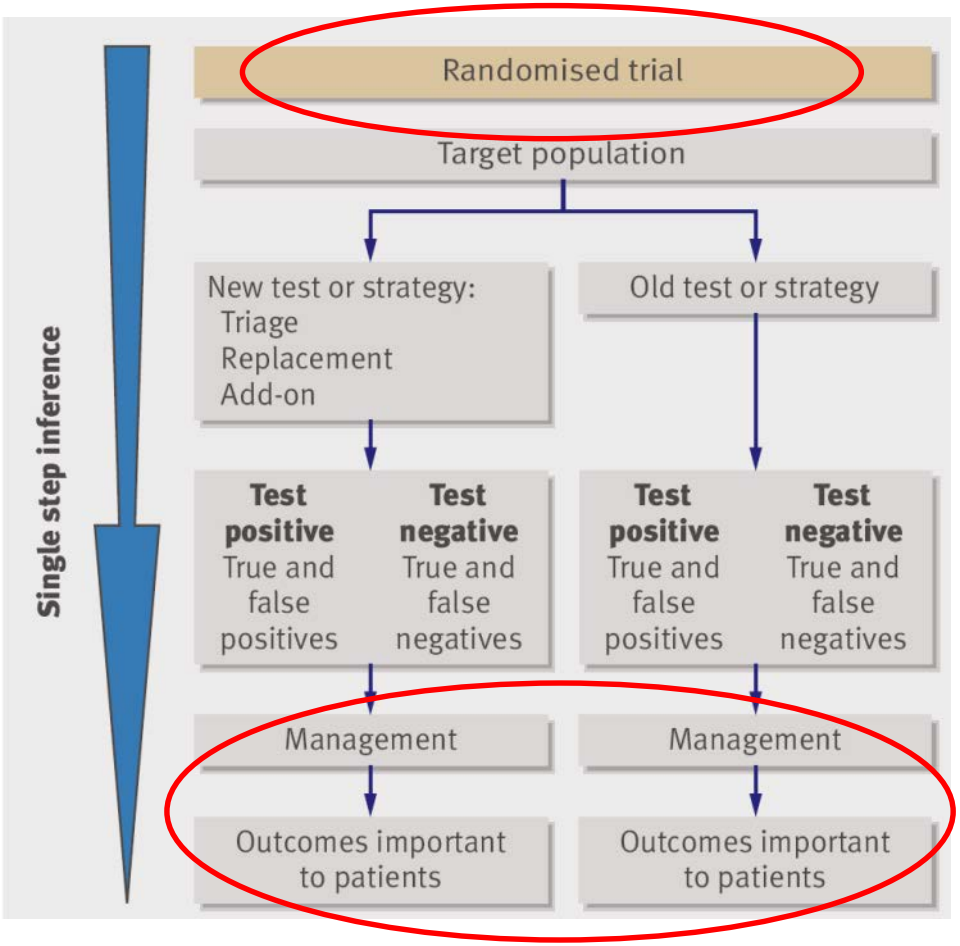
On the menu

- Demonstration studies: Additional value
- Considerations for preparation and study conduct
- Examples

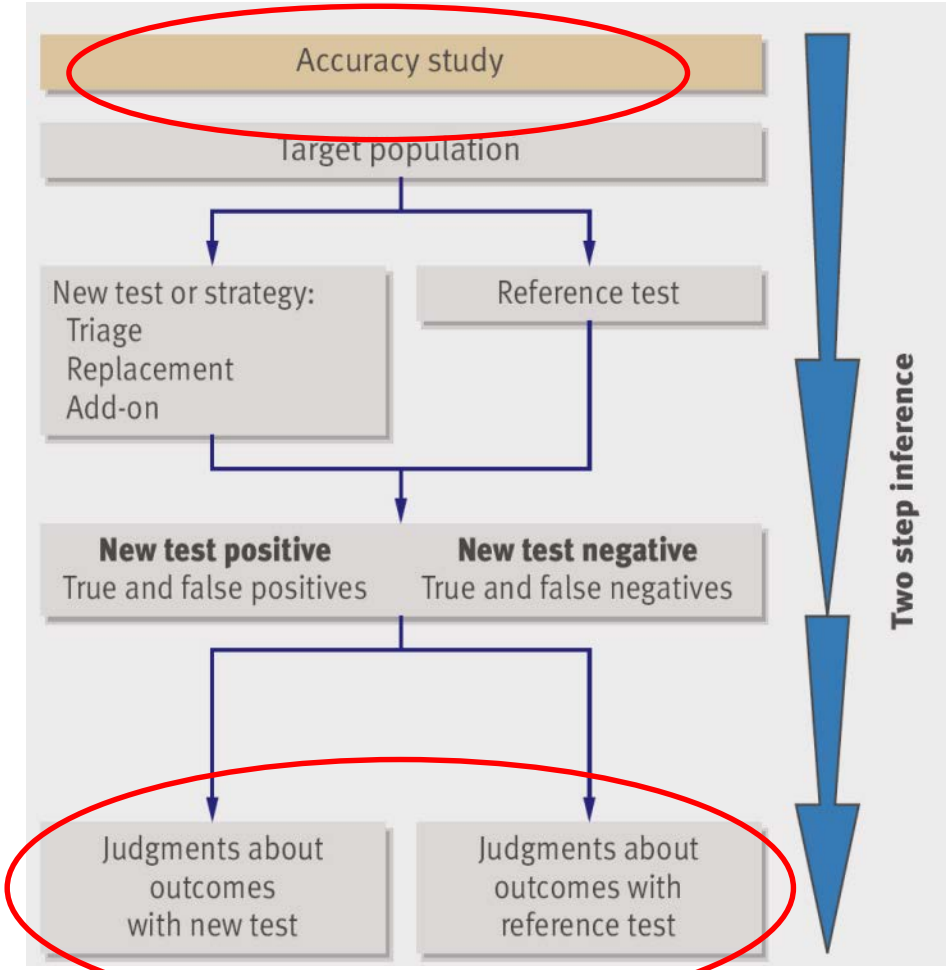
Leap of faith – Rolling out diagnostics based on evaluation studies?

Diagnostic study dilemma

What we want …



What we have …



Schunemann et al. BMJ, 2008

Demonstration studies: definition

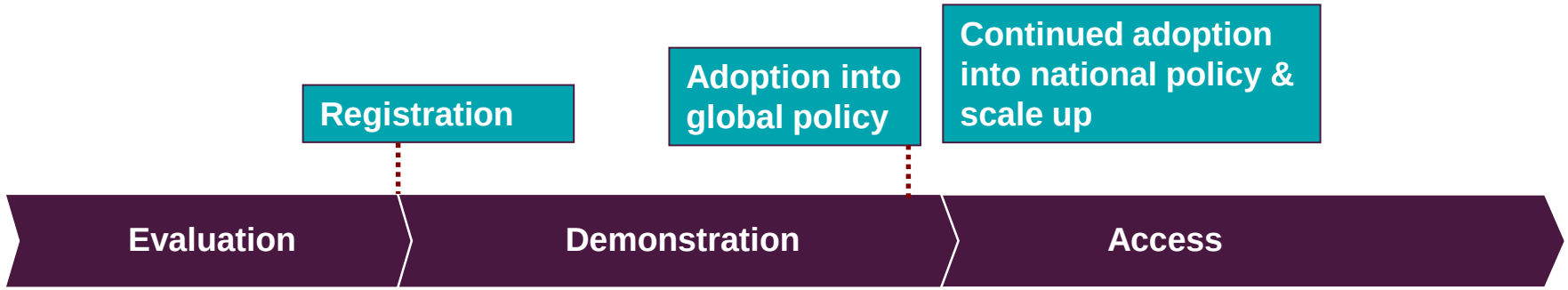
- Demonstration studies are large studies, in geographically and representative number of settings, and are intended to provide the evidence that new tests that perform well in controlled settings can also do so in uncontrolled settings, and can have an important medical and public health impact when implemented in programmatic settings.

- Carried out in the context of routine clinical services provision, either directly by the Ministry of Health (MOH), e.g., the National Disease Program or by other agencies working in collaboration with the MOH.

- The type of endpoints commonly studied include the feasibility of assay implementation, comparative costing between new and old technologies, and the impact on speed or accuracy of detection and subsequent patient management

- Require that treatment is based on new test

Evidence body required for successful uptake goes beyond accuracy data



Evaluation	Demonstration	Access
• Test performance • Ease of use • Surrogate patient important outcomes • Cost-comparisons	• Test performance in settings of intended use • Comparative effectiveness (patient important outcomes, case detection); • Cost of diagnostic process / treatment, incl. for patients • Operational requirements for implementation (infrastructure and human)	• (Comparative) validation for product selection to assess best fit • Algorithms, operational requirements • Test and resource utilization • Patient impact • Epidemiological impact • Economic impact • Health system impact

Is the test useful ?

	Evaluation	Demonstration
Reproducibility (precision)	*Usually done earlier*	-
Accuracy (compare test to «gold std»	X	-
Feasibility of implementation (robustness, costing)	X	X
Effects on clinical decisions and fit in clinical pathway	-	X
Effects on outcomes (and is it worth the costs)	*Surrogate marker data*	X

Determining Usefulness of a Diagnostic Test

Question	Possible Designs	Statistics for Results
What is the *feasibility* of implementing the test (costs, risks, user acceptability and feasibility (robustness, training requirements, etc.) of the test?	Assessment accompanying prospective studies	Mean cost, proportions experiencing adverse effects, proportions (operators) willing to use the test, failure rate, mean training duration

Determining Usefulness of a Diagnostic Test

Question	Possible Designs	Statistics for Results
How often do *test results affect clinical decisions*?	Diagnostic yield studies, studies of pre-& post test clinical decision making	Proportion abnormal, proportion with discordant results, proportion of tests leading to changes in clinical decisions; cost per abnormal result or per decision change

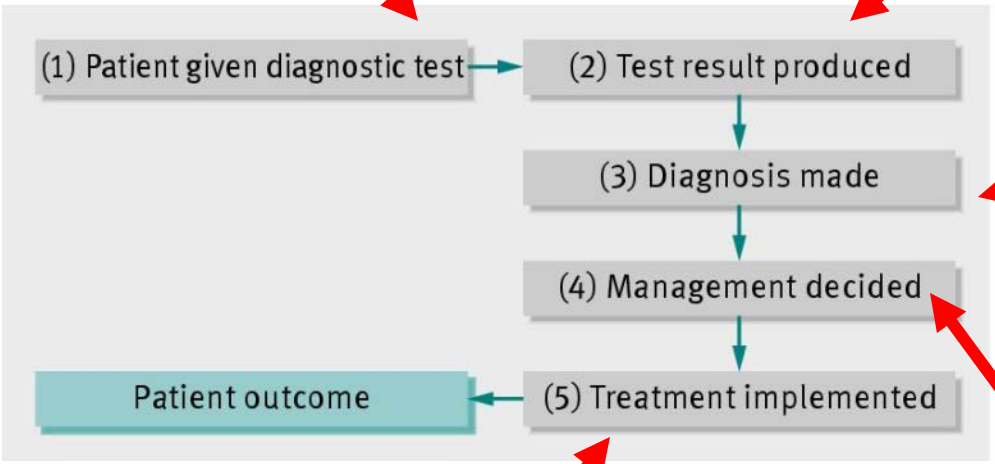
Determining Usefulness of a Diagnostic Test

Question	Possible Designs	Statistics for Results
Does doing the test *improve clinical outcome*, or *having adverse effects*?	Randomized studies in which the predictor variable is receiving the test & the outcome includes morbidity, mortality, or costs related either to the disease or to its treatment	Risk ratios, odd ratios, hazard ratios, number needed to treat, rates and ratios of desirable and undesirable outcomes

A good diagnostic alone does not make the patient better

Timing of test
Feasibility
Test process

Interpretability
Accuracy
Speed of results



Speed of diagnosis
Diagnostic yield
Diagnostic confidence

Speed of receiving treatment
Treatement efficacy
Adherence

Therapeutic yield
Therapeutic confidence

Preparing and conducting demonstration studies

Partner selection

Country criteria

- Middle / Low income
- Political commitment (MOU NTP or MOH) → management based on test results
- High expected medical need / benefit / cost saving → impact

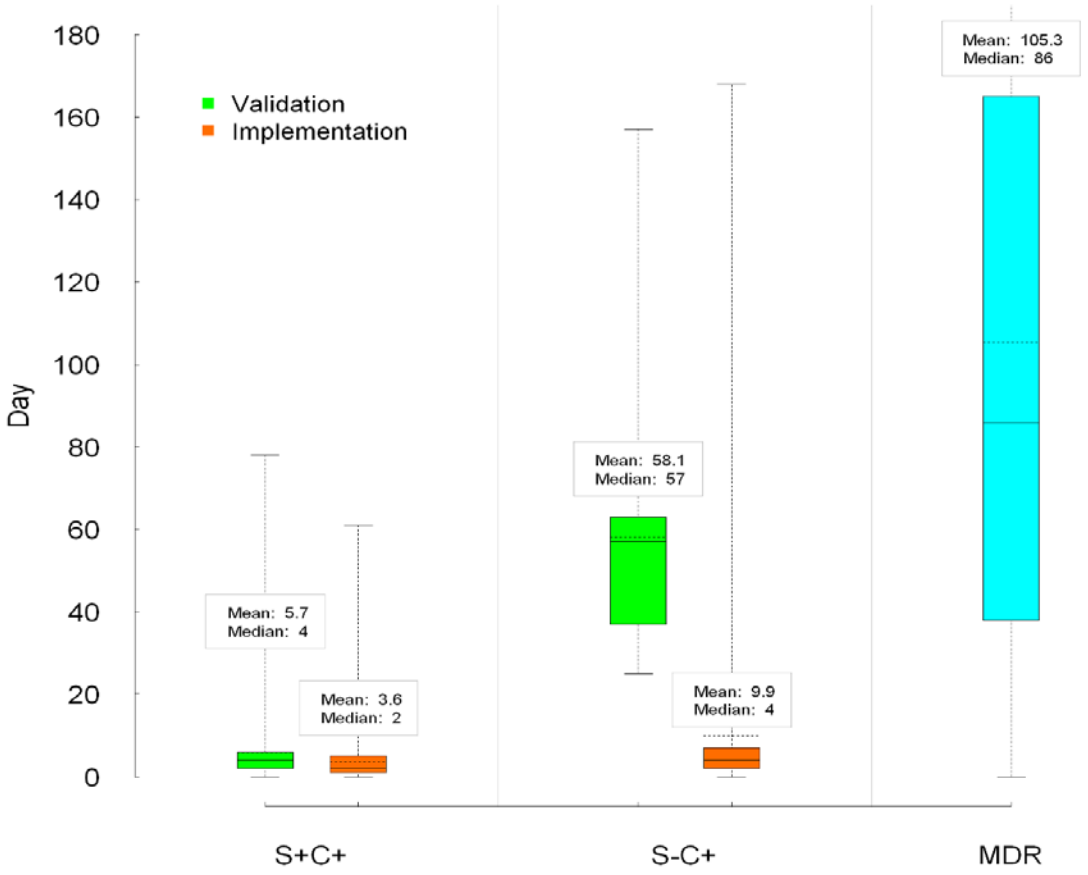
Trial site criteria

- Contributes to overall representativeness of data (TB/S- TB/MDR/HIV prevalence; strain types)
- Easily accessible from supervisory site
- Ease and speed of collecting data (local presence preferred)
- High quality reference tests available
- Representative sites for intended use and destined health care level:
  - Infrastructure (space, volume)
  - Logistics (supply chain, urban/rural)
  - HR (workload, skills)
  - Disease prevalence
  - other operational challenges (power supply, temperature)
- Meets quality standards
- Costs / Co-funding opportunities / opportunities to collaborate with implementing agencies

Demonstration study endpoints

Category	Details
Clinical performance	• Sensitivity/specificity/predictive values (stratified by site/smear/HIV) in intended settings of use
Operational performance	• Assess robustness of reagents & equipment; indeterminate rates; intermittent testing of (blinded) controls; T°/H log tags, dust, power irregularities, contamination rates; customer support interventions • Determine minimal training needs / performance dependence on skills/motivation/workload/user fatigue through proficiency testing tool & performance (stratification by user / over time) • Determine potential for scalability/batching • Assess user appraisal / requirements for implementation (such as waste management or storage) through user appraisal questionnaire and group discussions
Impact	• Time to detection of TB/DR; time to reporting compared to routine diagnostic algorithm; time to initiation of appropriate treatment; • Dropout rate prior to diagnosis; mortality during FU; hospitalization frequency & duration; time to return to work • Determine most suitable algorithms of use • Assess the cost (“per test” and “per patient” costs for the health system compared to routine) and cost-effectiveness compared to baseline (instead of and in addition to scenarios)

Example: Xpert demonstration study



Box plot showing the time to TB treatment during validation (treatment based on conventional methods) and implementation phases (treatment based on C-BAN and conventional methods). For the time to MDR treatment, treatment decisions during this study were only made on the basis of conventional DST (or LPA in South Africa).

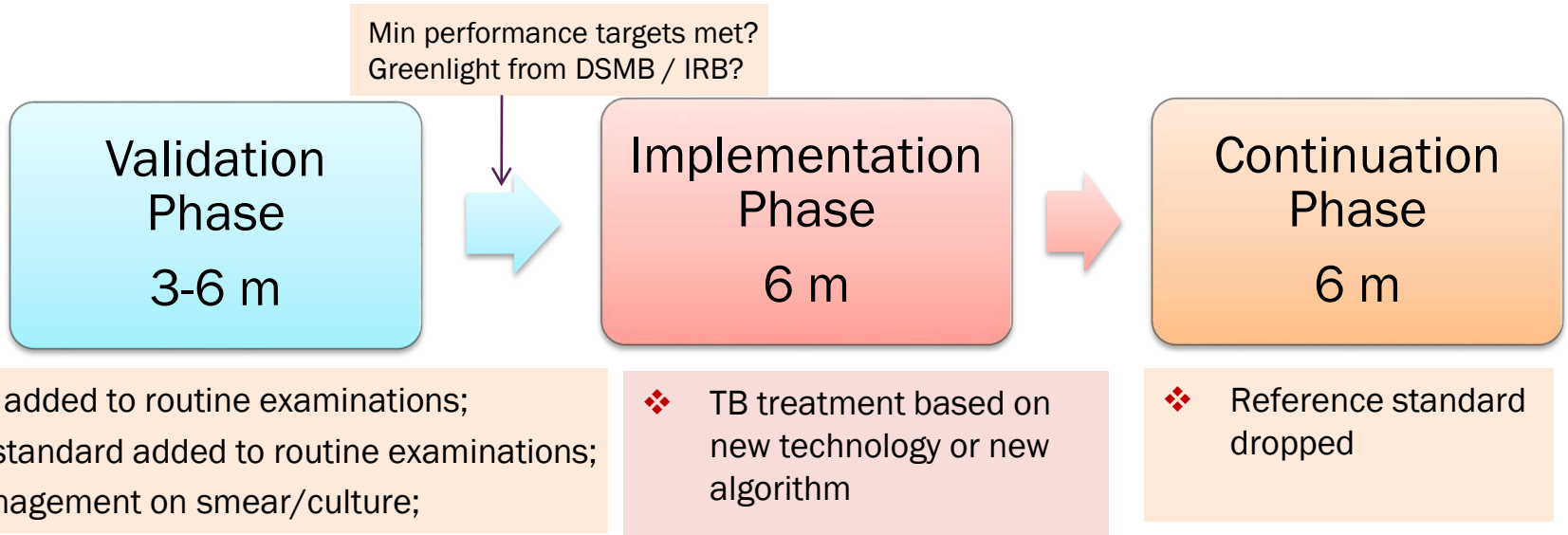
Example: Urine lateral flow LAM assay for diagnosing active TB in people living with HIV

Patient important outcomes

- Is there an association between a positive LF-LAM result and disease severity or mortality?
- Compared with current standard practice, what is the effect of LF-LAM on the following patient outcomes: time to treatment initiation; ART initiation; morbidity (for example, length of hospitalization, Karnofsky index); and mortality?
- What are the affordability, cost, and cost-effectiveness of LF-LAM implementation for TB diagnosis or screening compared with sputum microscopy or Xpert MTB/RIF?

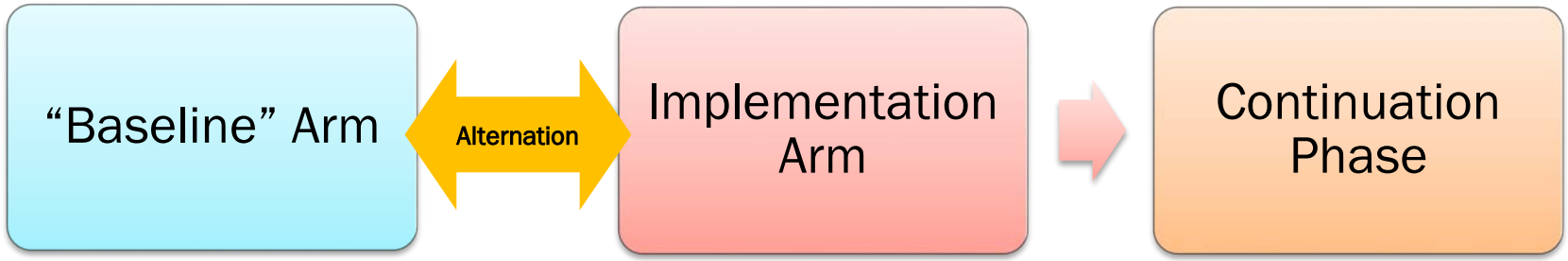
Designing a demonstration study

- Prospective, consecutive enrolment
- Inclusion criteria should not deviate from local guidelines (TB/MDR suspect);
- Analysis plan important (sample number; Tb tx; valid reference standard etc)
- Algorithm / Positioning reflect our thinking of how this new technology should be used



Design Variations (Cross-over; Baseline)

Xpert; LAMP

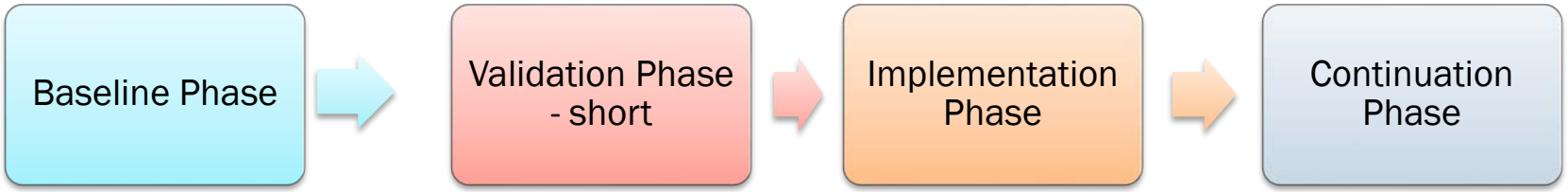


- Routine examination
- Culture / DST added as reference standard;
- Patient management on routine/reference;

- New test replacing routine

- Reference standard dropped

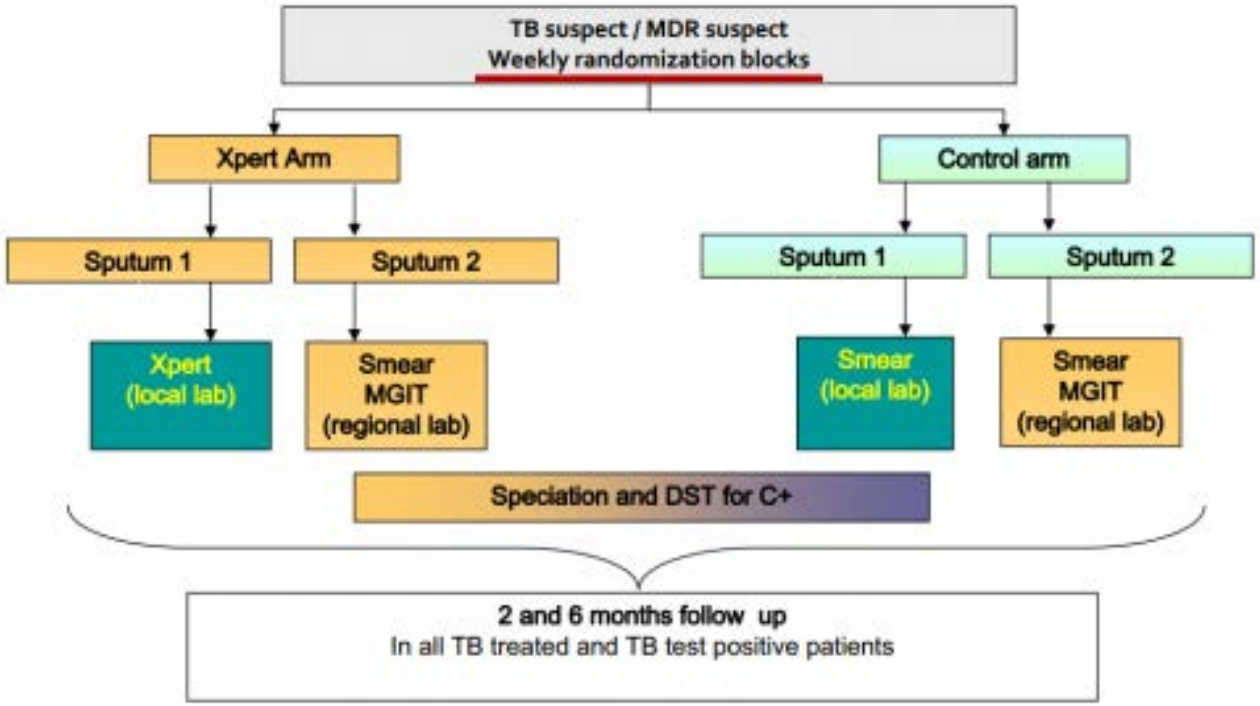
LED-FM; LPA (capacity of routine lab techs)



- Routine examination;
- Patient management on routine;
- Culture / DST added as reference standard;

Other designs may become more important:
True baseline assessment (no reference standard at enrolment) ; Cluster randomization; Comparison of different algorithms (add on versus replacement)

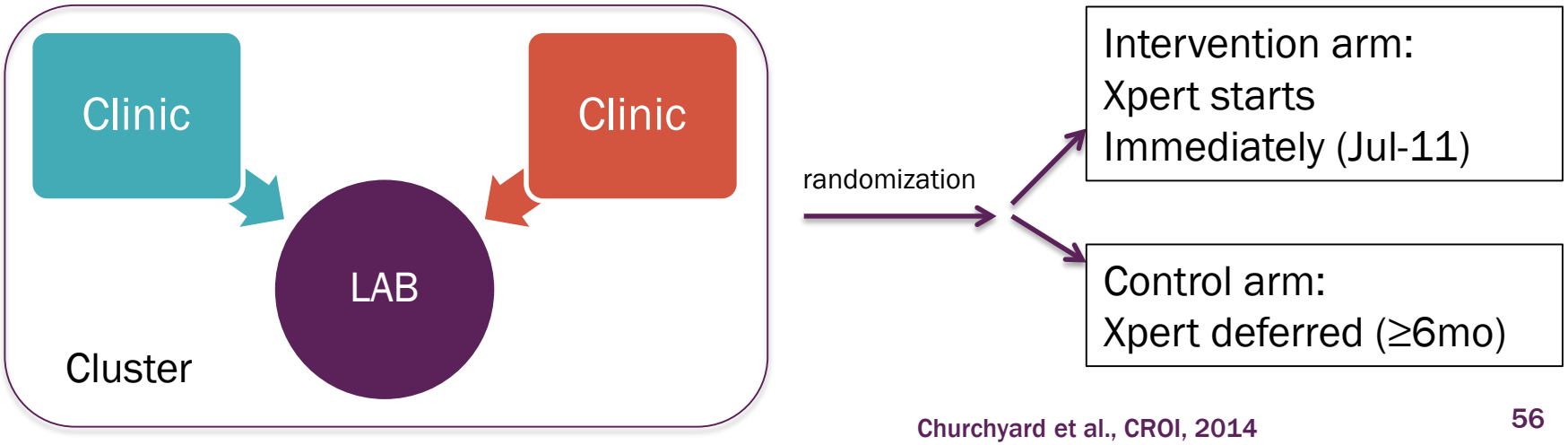
Example: Xpert MTB/RIF demonstration study, Cape Town, South Africa



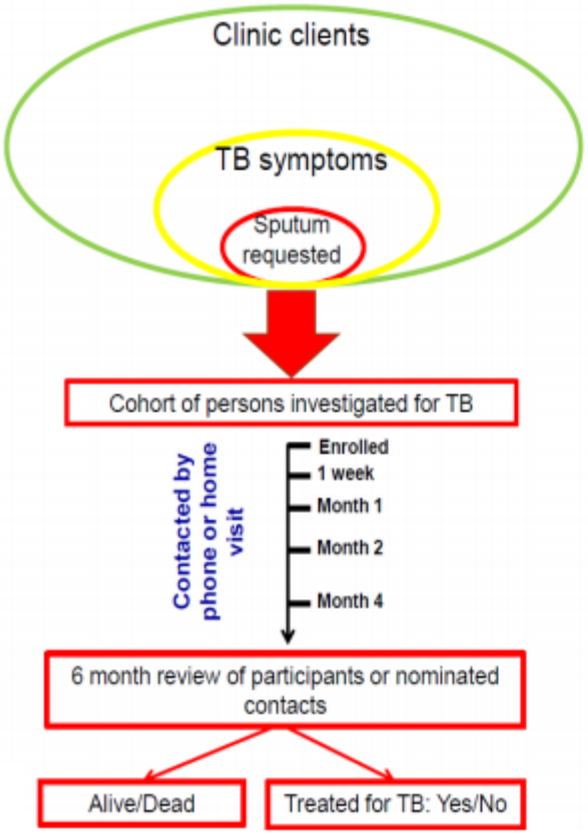
Example: E*valuating impact and cost-effectiveness of Xpert MTB/RIF in the routine roll-out in South Africa*

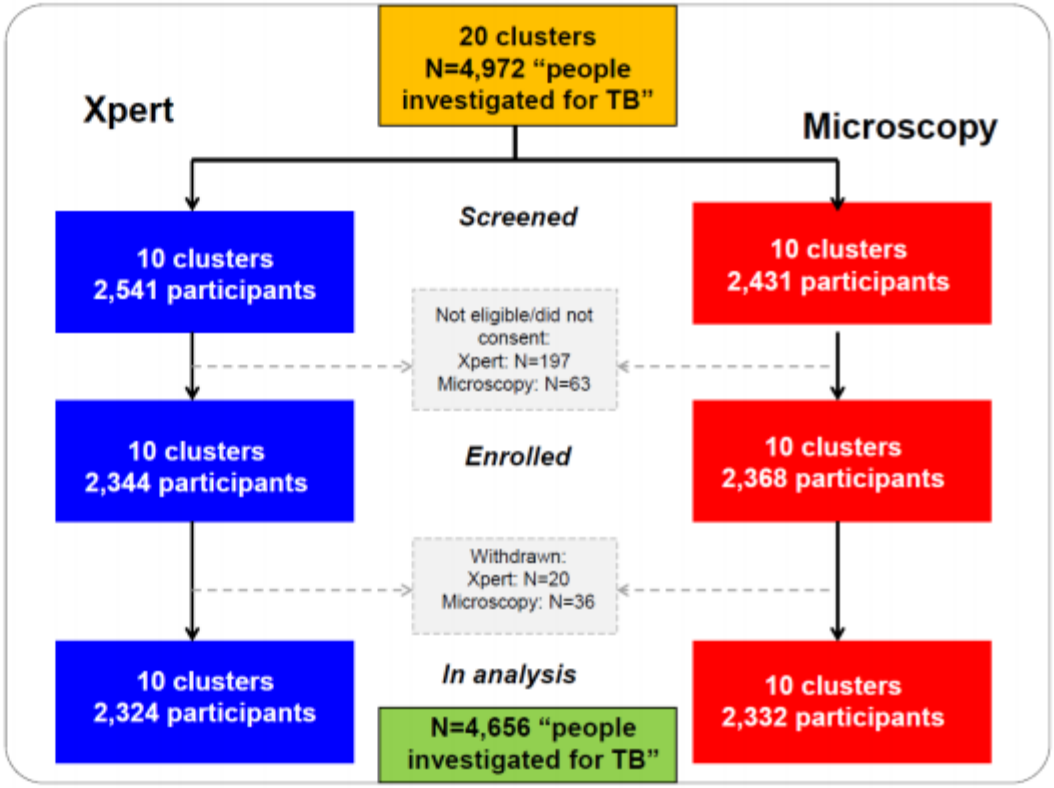
- Aim: to evaluate the effectiveness and cost effectiveness of Xpert MTB/RIF in the investigation of TB and TB drug resistance, and its impact on patient and programme outcomes and transmission at a population level.

- A two arm parallel cluster randomised trial design with laboratories as the unit of randomisation will be used to evaluate the effectiveness of Xpert; nested in the phased implementation of Xpert as per national plan for the rollout of Xpert. Xpert will be implemented at the start of the study in intervention clusters and will be deferred in the control arm for at least six months.



Churchyard et al., CROI, 2014





Churchyard et al., CROI, 2014

Example: Xpert roll-out in Brazil

Durovni et al., Xpert Global Implementers' Meeting, 2013

Primary endpoint: Impact of Xpert implementation on pulmonary TB case detection

Control data: From routine microscopy performed during baseline

Intervention: Introduction of Xpert as microscopy replacement

Follow up

- Time to treatment, drop out rate or other impact endpoints: At least passive follow up through treatment registers for all gold standard positive patients

- Discrepant patients and random controls (deficient gold standard)

- Sensitivity / specificity of baseline algorithm (smear + clinical judgment (CXR, symptoms)) : At least passive follow up for all through patient records

Ethical Considerations

- Patients will be treated on the basis of the result of the new test
  → Ethical approval will therefore be required for all participating sites/countries

- If no foreseeable risk to patients (Registered product; strong accuracy data; sufficient benefits (culture/DST)
  → A waiver of informed consent can be requested from relevant IRBs

- Data safety and monitoring board
  → Adverse events, assay modifications

A fine balance

1. Quality standards vs representativeness

More monitoring and supervision,

More endpoints (CRF length, follow up)

Higher clinical trial standard (study nurses, data managers)

=

Results less representative of average routine setting

Greater bias

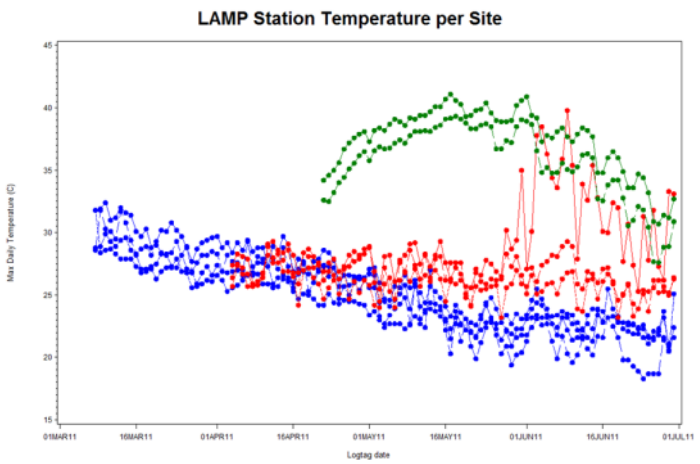
Less real-world scenario

2. Solidity of data vs urgency for change

Risk that findings do not translate to other settings / aspects overlooked

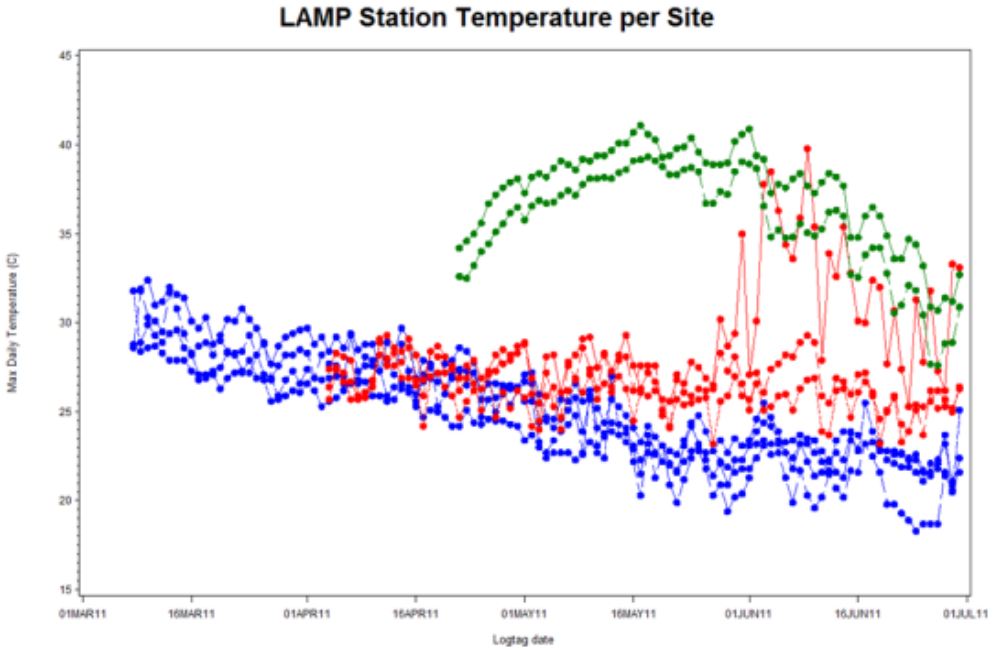
Holding up market introduction in view of large unmet diagnostic need

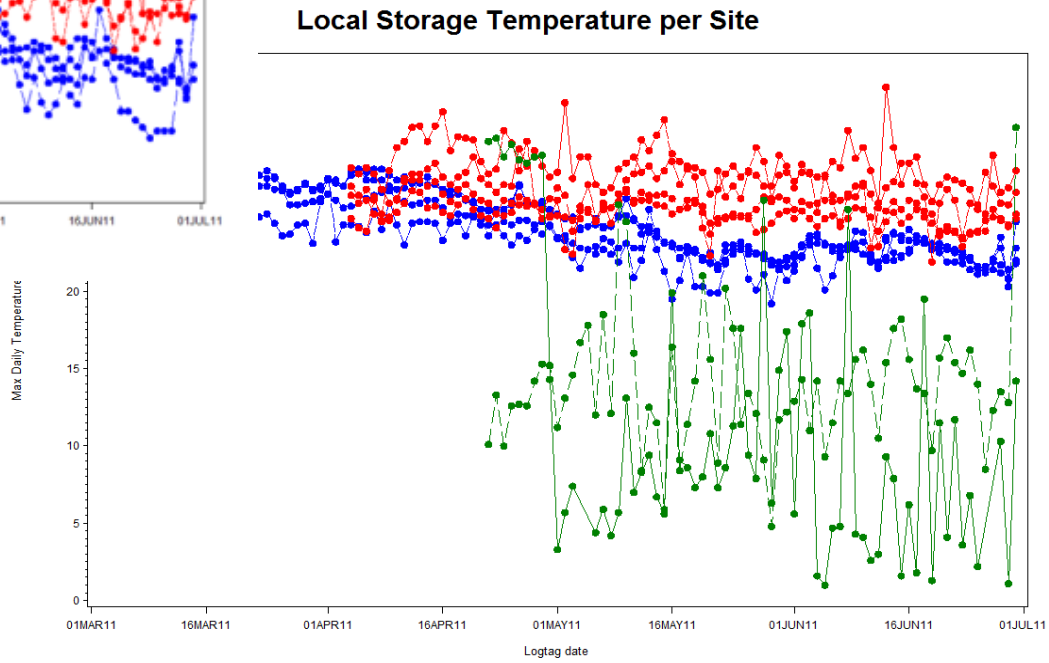
Focus on operational issues



- Training needs
- Influences of workload on test results
- Storage
- Waste managment
- Electrical supply and backup power
- Biosafety requirements
- Operating temperature/humidity
- Scalability

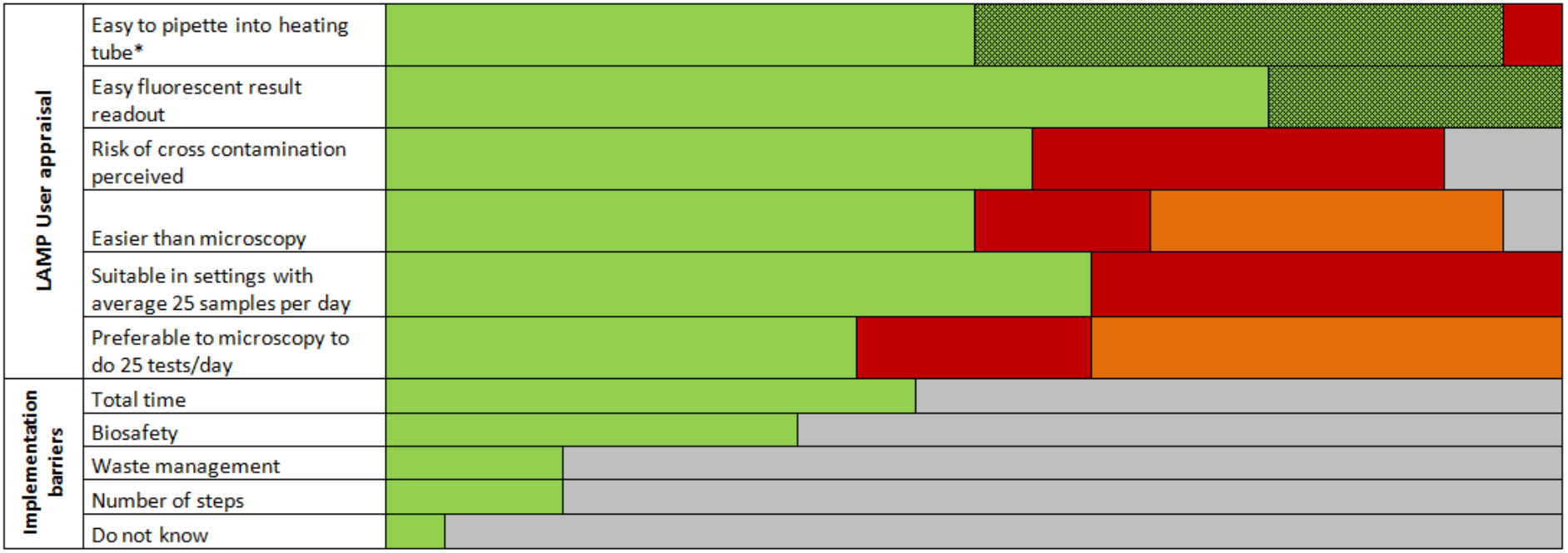
Example: LAMP demonstration study





Ease of use

- 16 / 20 users passed proficiency testing after 5 days training
- 2/4 after 10 days training and 2/4 after 15 days training
- Average hands on time 45 min



Design changes for POC test demonstration studies?

Diagnosis and Treatment: Hand in hand?

Community health worker applying a RDT and treating Malaria on the spot

What would we do with a TB POC test? – Diagnose & Treat or Screen & Refer?

- Patient values / preferences more important?
- Ease of use assessment and assessment of operational feasibility even more important
- Larger number of CHW to reach sample size and representativeness

Assessing the surrounding package