Diagnostic RCTs: Methodological Elements of Recent Case Finding RCTs

John Metcalfe, MD, MPH
Assistant Professor, Division of Pulmonary and Critical Care
University of California, San Francisco
PhD candidate, Division of Epidemiology
University of California, Berkeley
John.Metcalfe@ucsf.edu

Would that I could discover truth as easily as I can uncover falsehood.

—Cisero (44 B.C.)

Overview

- Introduction
- Corbett et al., Lancet 2010
  - Cluster Randomization
- Churchyard et al., Thorax 2011
  - Intention to Treat
- Cuevas et al., PLOS Med 2011
  - Equivalence (Non-Inferiority) Trials
- Boeheme et al., Lancet 2011
  - Blinding

Overview

- Introduction
- **Corbett et al., Lancet 2010**
  - **Cluster Randomization**
- Churchyard et al., Thorax 2011
  - Intention to Treat
- Cuevas et al., PLOS Med 2011
  - Equivalence (Non-Inferiority) Trials
- Boeheme et al., Lancet 2011
  - Blinding

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;

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

Trial Synopsis:
Corbett et al. Lancet 2010

- Population: Adults, cough ≥ 2 weeks, Harare
- Intervention: **Mobile van**
- Comparator: … vs. **door-to-door** visits for community-wide active case finding
- Outcome: Cumulative SS+ TB detected per 1000 subjects
  - Change in prevalence of Cx+ TB in randomly selected HHs

Trial Synopsis: *Corbett et al. Lancet 2010*

- **Results:**
  - ITT analysis
  - 46 study clusters, ~115,000 individuals
  - Mobile van 255/5466 vs. 137/4711 door-to-door
    - RR 1.48, 95% CI 1.11–1.96, p<.01 favoring mobile van

Cluster Randomization I

- Randomization at the **group** level
- Appropriate if exposure is a group level exposure (e.g., public health or social intervention)
- Feasibility
- Minimizes “contamination”

Cluster Randomization II

- Reduced statistical power
  - Members of a group (cluster) resemble each other more than they resemble members of other groups (clusters) = ***FEWER independent observations***
  - Intraclass Correlation Coefficient (ICC)
    - Ratio of between-cluster variance to total variance
      - ICC=1 : complete (perfect) correlation
      - ICC=0 : no correlation (independent obs)
      - Typically 0.001 - 0.05

$$\rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2}$$

Effective Sample Size (ESS)

- Design Effect
  - Penalty for using a clustered design
  - Ratio of variance under cluster randomization to variance assoc with a simple random sample of same number
  - DE=1+(m-1) ρ

ESS= mk/DE

k = 46 clusters
m = ~2500 people/cluster
Total n = 115,000

ICC(ρ)	DE	Ne
0	1.0	115,000
0.001	3.5	33,000
0.01	26	4,400
0.1	251	460
0.5	1251	92
1	2500	46

Cluster Randomization III: Accounting for clustering effect

- *"The practice of ignoring clustering in the design stage of a trial can lead to type 2 error, while ignoring it at the analysis phase inevitably leads to type I error*." – Campbell 2007
  - Design Phase: Standard sample size calcs = underpowered for CRT's
  - Analysis Phase: Not accounting for clustering gives falsely tight CI's

Overview

- Introduction
- Corbett et al., Lancet 2010
  - Cluster Randomization
- **Churchyard et al., Thorax 2011**
  - **Intention to Treat**
- Cuevas et al., PLOS Med 2011
  - Equivalence (Non-Inferiority) Trials
- Boeheme et al., Lancet 2011
  - Blinding

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]

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.

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]

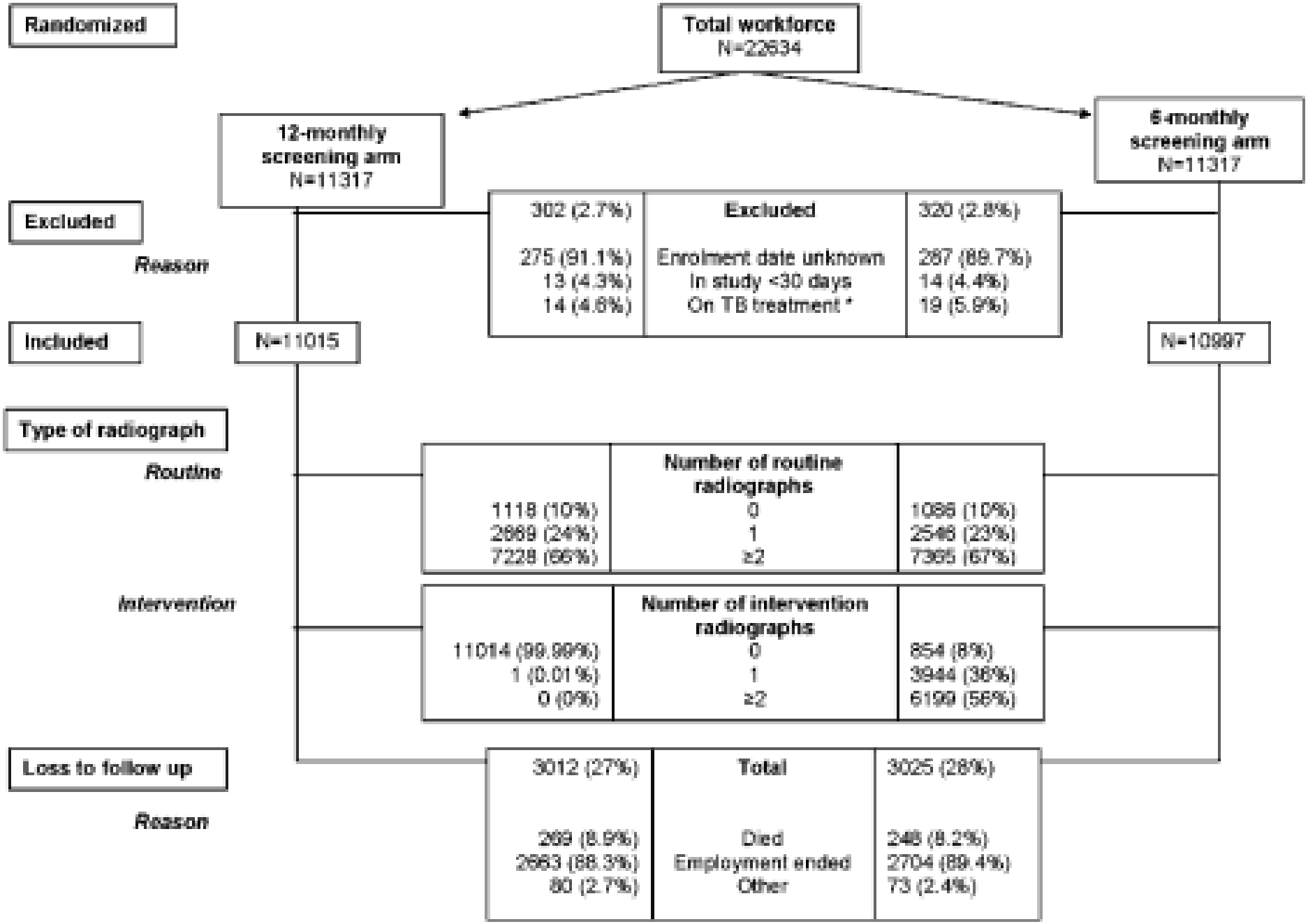
Radiological screening has been used in the gold mining industry for decades[18]; both 6- and

Synopsis: Churchyard et al., Thorax 2011

- **Population**: 22,634 South African miners followed for median 2 years
- **Intervention**: 6-monthly vs. ….
- **Comparator**: standard 12-monthly radiological screening for active TB case-finding among gold miners
- **Outcome**: Proportion of total cases diagnosed during study period
  - Secondary: Extent of disease, mortality

Churchyard et al., Thorax 2011

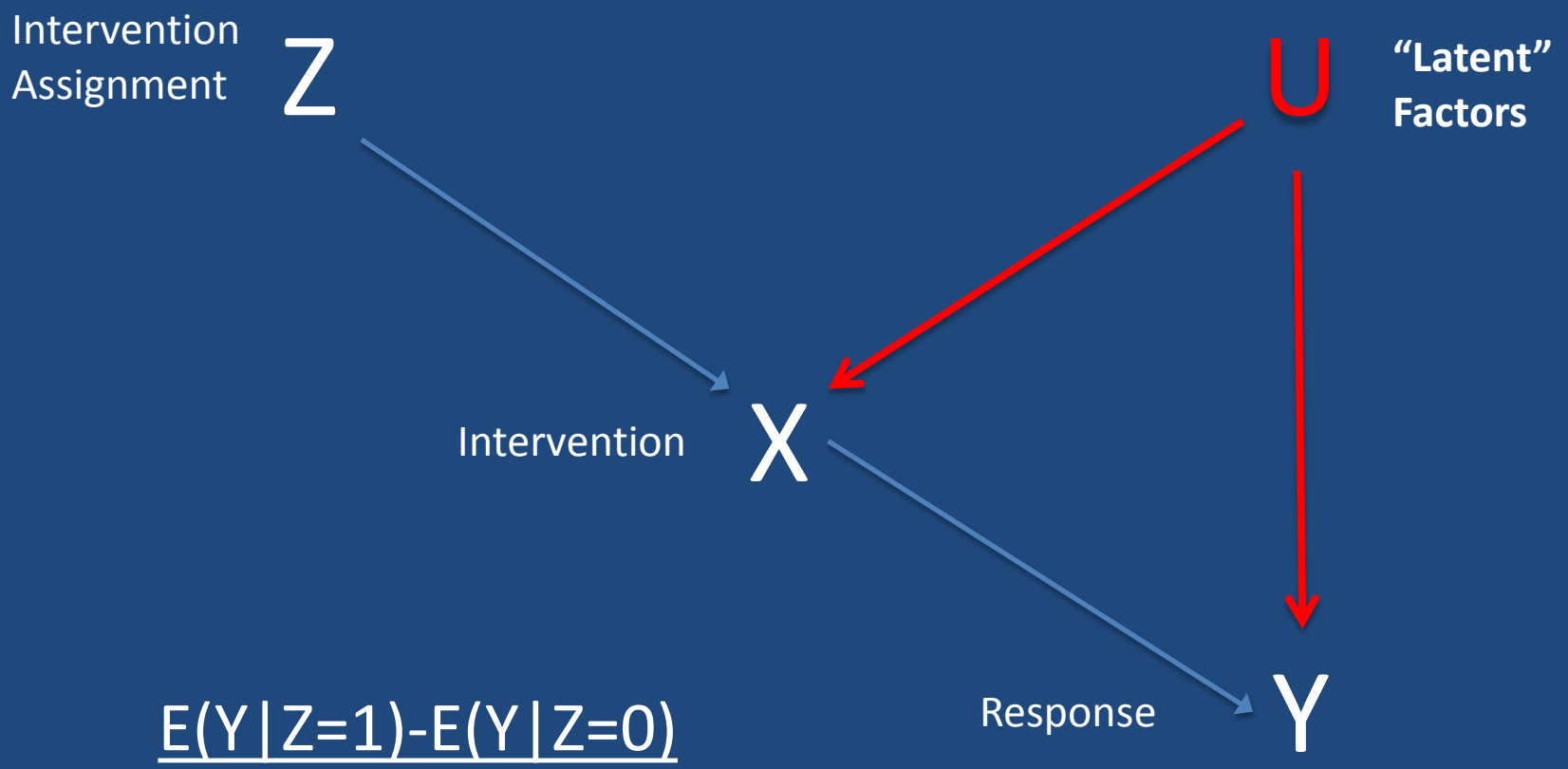
- Results: Similar proportion detected in both screening arms
  - 6-monthly: 29.4% (197/679); 12-monthly: 28.3% (179/632)


Randomized
Total workforce
N=22634
12-monthly screening arm
N=11317
6-monthly screening arm
N=11317
Excluded
Reason
302 (2.7%)
Excluded
320 (2.8%)
275 (91.1%)
Enrolment date unknown
287 (89.7%)
13 (4.3%)
In study <30 days
14 (4.4%)
14 (4.6%)
On TB treatment *
19 (5.9%)
Included
N=11015
N=10997
Type of radiograph
Routine
Number of routine radiographs
1118 (10%)
0
1086 (10%)
2689 (24%)
1
2548 (23%)
7228 (66%)
≥2
7365 (67%)
Intervention
Number of intervention radiographs
11014 (99.99%)
0
854 (8%)
1 (0.01%)
1
3944 (36%)
0 (0%)
≥2
6199 (56%)
Loss to follow up
3012 (27%)
Total
3025 (28%)
Reason
269 (8.9%)
Died
248 (8.2%)
2663 (88.3%)
Employment ended
2704 (89.4%)
80 (2.7%)
Other
73 (2.4%)

Intent-to-Treat

- Analyze subjects as they were randomized, regardless of adherence, withdrawals or cross-overs
  - "as treated" analysis: based on treatment actually received
  - "per protocol" analysis: includes ONLY those who actually followed protocol (eg, non-adherers dropped)
- Ensures benefits of randomization maintained
- Considered a conservative estimate of the treatment effect
- Best approximation of treatment *effectiveness*

Intent-to-Treat



$$\frac{E(Y|Z=1)-E(Y|Z=0)}{E(X|Z=1)-E(X|Z=0)}$$

Modified from Pearl, *Causality* 2009

Overview

- Introduction
- Corbett et al., Lancet 2010
  - Cluster Randomization
- Churchyard et al., Thorax 2011
  - Intention to Treat
- **Cuevas et al., PLOS Med 2011**
  - **Equivalence (Non-Inferiority) Trials**
- Boeheme et al., Lancet 2011
  - Blinding

A Multi-Country Non-Inferiority Cluster Randomized Trial of Frontloaded Smear Microscopy for the Diagnosis of Pulmonary Tuberculosis

Luis Eduardo Cuevas[1,2]*, Mohammed Ahmed Yassin[1], Najla Al-Sonboli[3], Lovett Lawson[4], Isabel Arbide[5], Nasher Al-Aghbari[6], Jeevan Bahadur Sherchand[7], Amin Al-Absi[6], Emmanuel Nnamdi Emenyonu[4], Yared Merid[8], Mosis Ifenyi Okobi[9], Juliana Olubunmi Onuoha[4], Melkamsew Aschalew[8], Abraham Aseffa[10], Greg Harper[1], Rachel Mary Anderson de Cuevas[1], Kristin Kremer[11], Dick van Soolingen[11], Carl-Michael Nathanson[2], Jean Joly[2], Brian Faragher[1], Stephen Bertel Squire[1], Andrew Ramsay[2]

PLOS MEDICINE

Trial Synopsis - *Cuevas et al.*

- **Population:** Adults with cough ≥ 2 weeks living in 4 low-income countries
- **Intervention:** Front-loaded (spot-spot-morning (SSM))
- **Comparator:** … vs. standard (spot-morning-spot (SMS)) sputum smear collection
- **Outcome:** Diagnostic accuracy
  - culture (x1) gold standard

Trial Results – *Cuevas et al.*

- Analysis
  - Block-randomization by week
  - Non-inferiority interval: **0 to –5%**
- 6,467 TB suspects with culture performed
  - 222 randomized weeks (114 standard, 108 front-loaded)
  - 1,224 (18.5%) SS+
  - 1,561 (24.1%) culture-positive
- Sensitivity
  - **SSM** (**70.2%**, 95% CI 66.5% – 73.9%) **non-inferior** to **SMS** (**65.9%**, 95% CI 62.3%–69.5%)
    - Difference: 4.3% (95% CI 0.6% to 9.0%)
  - **SS** (**63.6%**, 95% CI 59.7%–67.5%) also non-inferior to **SM** (**64.8%**, 95% CI 61.3%–68.3%)
    - Difference: −1.2% (95% CI −3.9% to 6.4%)

Why do a Non-inferiority/Equivalence Trial?

- New treatment/intervention is expected to match efficacy of the standard

  BUT may have advantages with regards to:

  - Safety
  - Convenience
  - Cost
  - Side effects

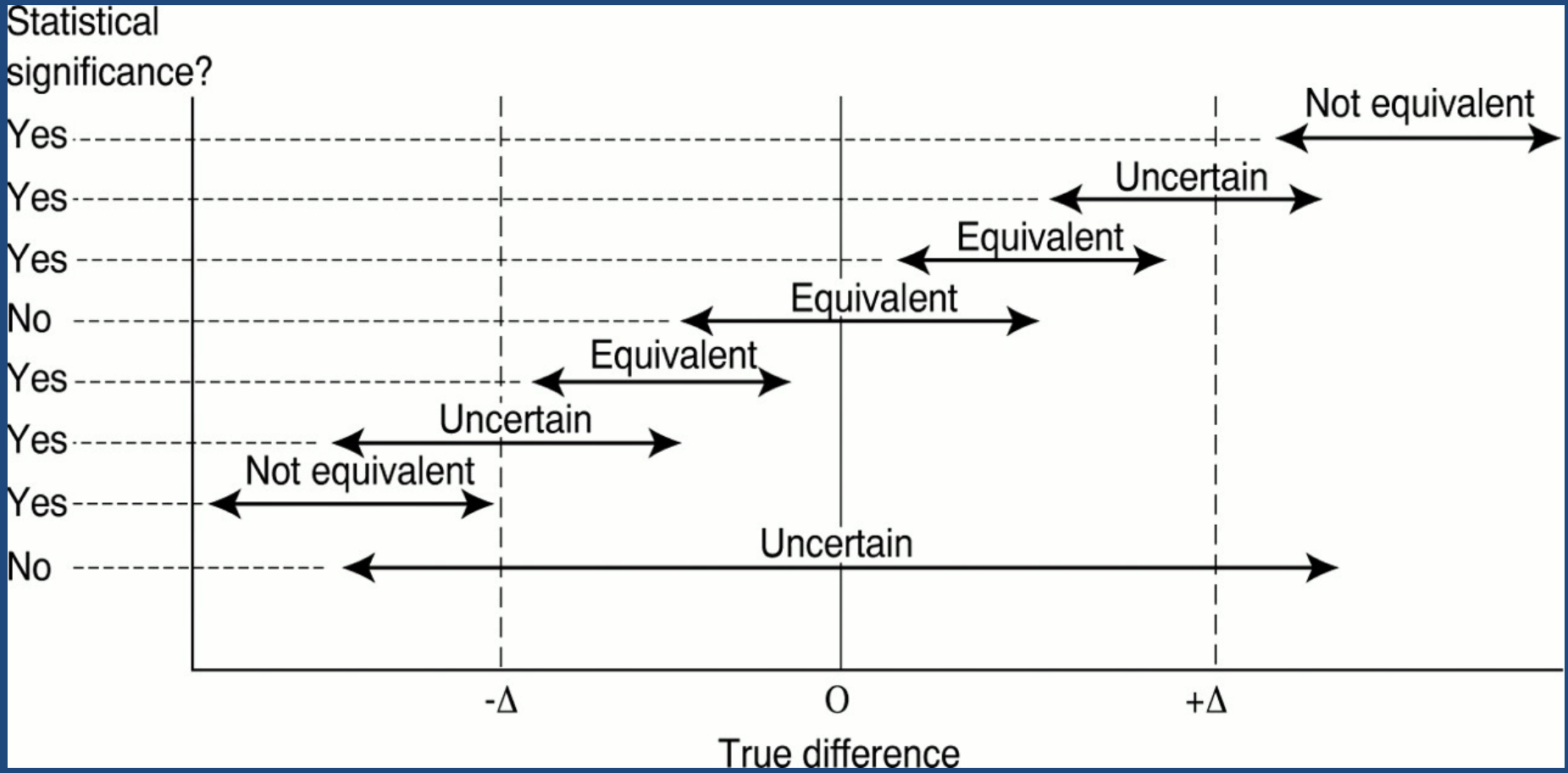
- Not ethical to use placebo as a comparison group

Piaggio, 2006

Definitions

- **Superiority trials**
  - **Seek to show that new treatment is different from (better than) an existing treatment**
  - $H_{0:}\ \mu_{test} = \mu_{act}$
  - $H_{a:}\ \mu_{test} > \mu_{act}$
- ***Equivalence trials***
  - Seek to show that new treatment differs from an existing treatment by no more than a pre-specified amount ***(equivalence margin, -/+Δ)***
  - $H_{0:}\ \mu_{test} \neq \mu_{act}$
  - $H_{a:}\ \mu_{test} = \mu_{act}$ or $-\Delta < \mu_{test} - \mu_{act} < \Delta$
- ***Non-inferiority trials***
  - Seek to show that new treatment is no worse than an existing treatment by a pre-specified amount ***(noninferiority margin, -Δ)***
  - $H_{0:}\ \mu_{test} < \mu_{act}$
  - $H_{a:}\ \mu_{test} \geq \mu_{act}$ or $\mu_{test} - \mu_{act} \geq -\Delta$

Christensen, 2007



Jones BMJ 1996

Margin of Equivalence

- "Margin of Equivalence" (Δ) must be established *a priori*
  - A difference of means
  - A difference of proportions
  - Risk ratios (relative risk)
- Generally smaller than in superiority trials
- Examples:
  - No larger than the smallest difference between standard txt and placebo demonstrated in RCT's (Temple, 2000)
  - The smallest value that would be a clinically important effect (Piaggio, 2006)
  - No more than ½ the value used in a superiority trial (Christenson, 2007)

Natural Incentives Towards Internal Validity

Common sources of bias in a clinical trial tend to obscure differences between treatment groups

- Nonadherence
- Drop-outs
- Missing data
- Cross-overs
- Use of concomitant therapy potentially affecting study outcome

Both intention-to-treat and per-protocol analyses should be presented

Piaggio, 2006

Non-inferiority Trials: Summary

- Make the correct hypothesis
  - Absence of evidence (of a difference) must not be confused with evidence of absence (of a difference)
  - The observation of a lack of a difference between 2 treatments cannot automatically be used as evidence of equivalence.
- A margin of equivalence should be justified *a priori* and used for sample size calculation
- Loss to follow up, withdrawals, or compliance issues can bias results towards accepting equivalence
- Perform both ITT & PP analyses
- Follow reporting guidelines

Overview

- Introduction
- Corbett et al., Lancet 2010
  - Cluster Randomization
- Churchyard et al., Thorax 2011
  - Intention to Treat
- Cuevas et al., PLOS Med 2011
  - Equivalence (Non-Inferiority) Trials
- **Boeheme et al., Lancet 2011**
  - **Blinding**

Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study

Catharina C Boehme, Mark P Nicol, Pamela Nabeta, Joy S Michael, Eduardo Gotuzzo, Rasim Tahirli, Ma Tarcela Gler, Robert Blakemore, William Worodria, Christen Gray, Laurence Huang, Tatiana Caceres, Rafail Mehdiyev, Lawrence Raymond, Andrew Whitelaw, Kalaiselvan Sagadevan, Heather Alexander, Heidi Albert, Frank Cobelens, Helen Cox, David Alland, Mark D Perkins

Summary

Background The Xpert MTB/RIF test (Cepheid, Sunnyvale, CA, USA) can detect tuberculosis and its multidrug-resistant form with very high sensitivity and specificity in controlled studies, but no performance data exist from district and subdistrict health facilities in tuberculosis-endemic countries. We aimed to assess operational feasibility, accuracy, and effectiveness of implementation in such settings.

Methods We assessed adults (≥18 years) with suspected tuberculosis or multidrug-resistant tuberculosis consecutively presenting with cough lasting at least 2 weeks to urban health centres in South Africa, Peru, and India, drug-resistance screening facilities in Azerbaijan and the Philippines, and an emergency room in Uganda. Patients were excluded from the main analyses if their second sputum sample was collected more than 1 week after the first sample, or if no valid reference standard or MTB/RIF test was available. We compared one-off direct MTB/RIF testing in nine microscopy laboratories adjacent to study sites with 2–3 sputum smears and 1–3 cultures, dependent on site, and drug-susceptibility testing. We assessed indicators of robustness including indeterminate rate and between-site performance,

Lancet 2011; 377: 1495–1505

Published Online
April 19, 2011
DOI:10.1016/S0140-6736(11)60438-8

See **Comment** page 1467

Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland (C C Boehme MD, P Nabeta MD, C Gray MPH, H Alexander PhD, H Albert PhD, M D Perkins MD); **National Health Laboratory Service, Groote Schuur Hospital,**

Trial Synopsis:
Boehme et al., Cape Town Site

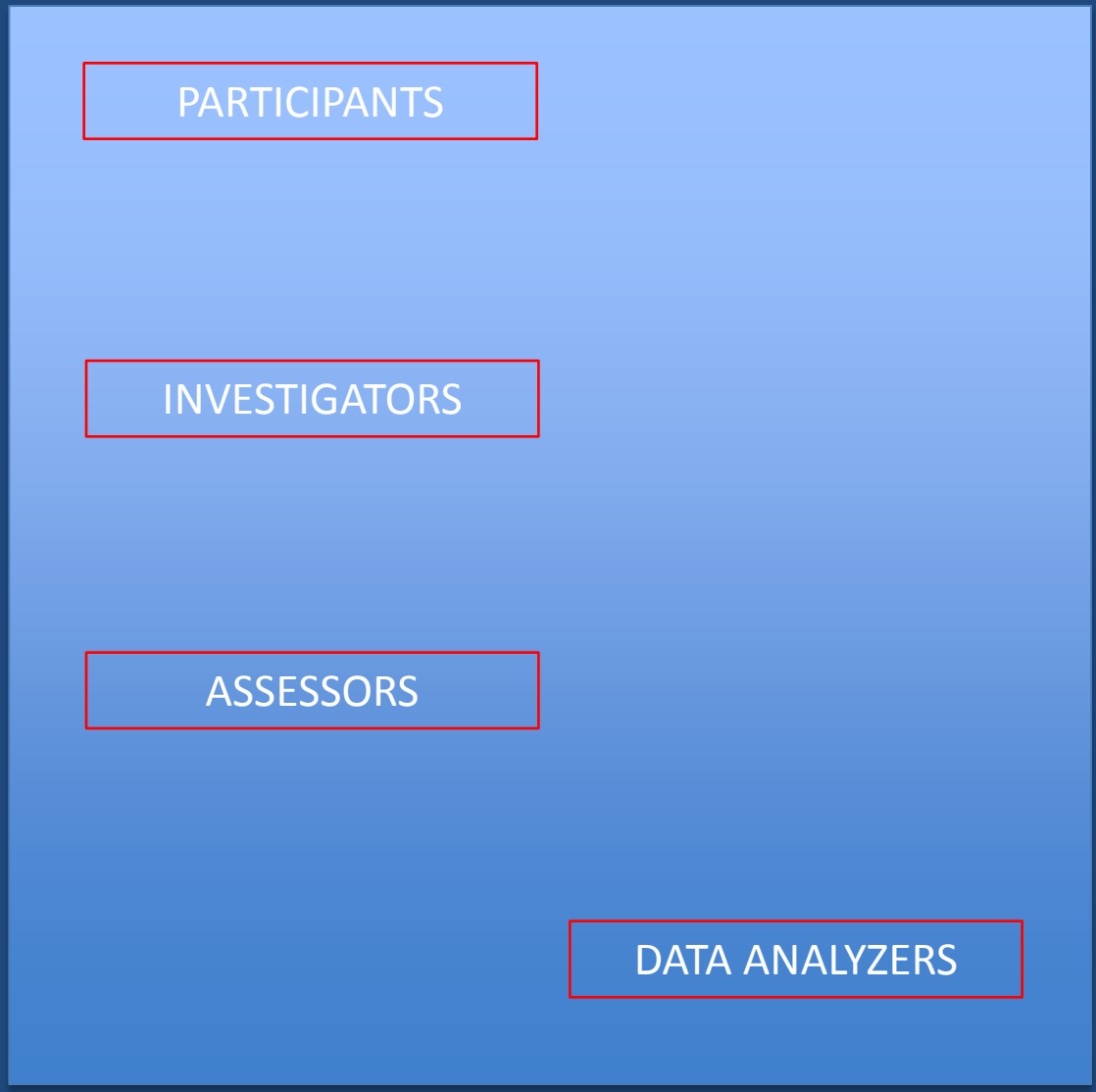
- **Population:** Adults, cough ≥ 2 weeks, health center or provincial hospital, Cape Town
- **Intervention:** Weekly alternation of Xpert MTB/RIF x 1
- **Comparator:** … vs. routine smear microscopy/culture (2 FM smears and 1 MGIT culture)
- **Outcome:** Diagnostic accuracy
  - culture as gold standard

Blinding

"Masking, which was not necessary in South Africa due to study design, was accomplished at the other sites by having different staff do smear microscopy and MTB/RIF testing."

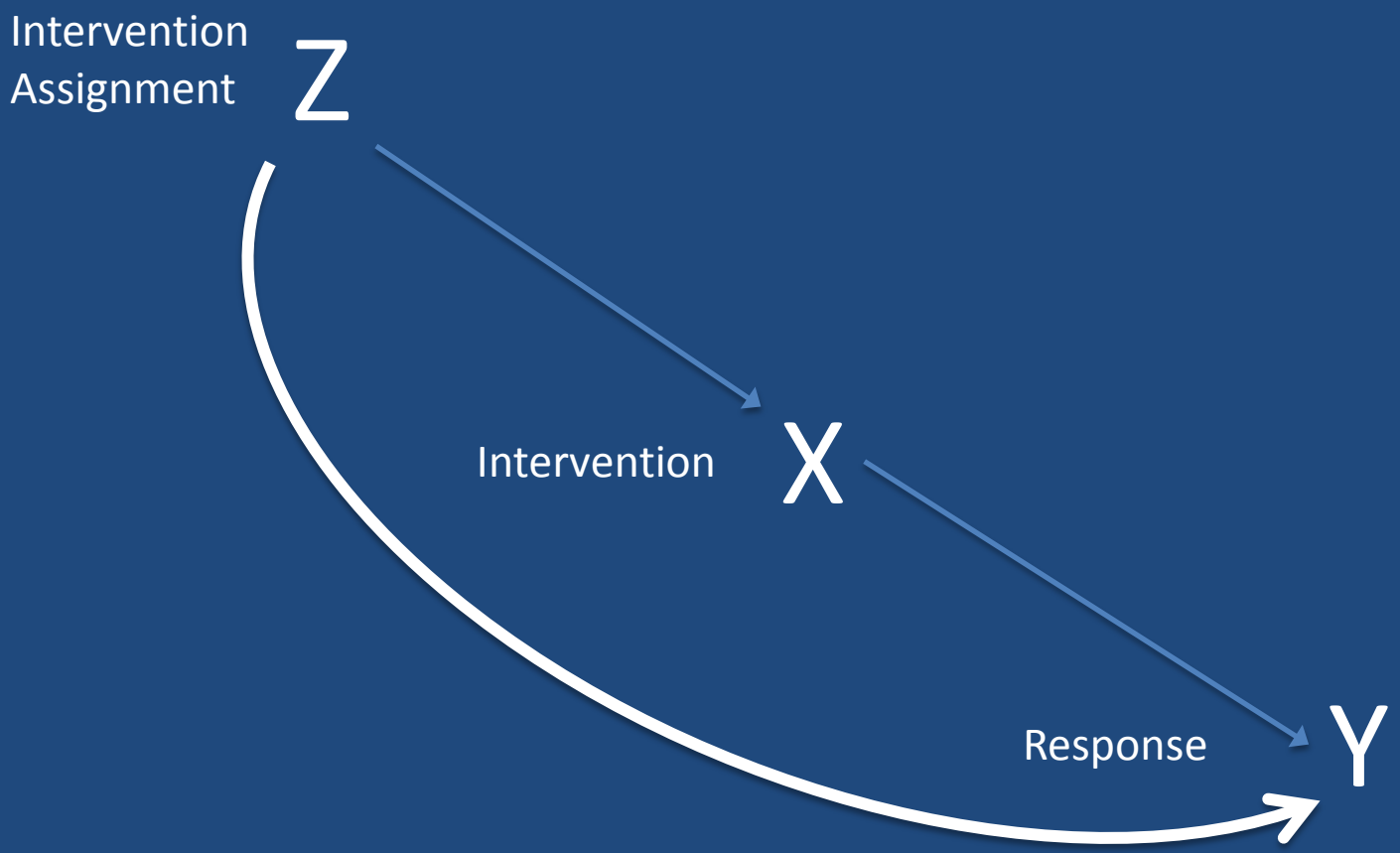
Levels of Blinding

- Single Blinding
- Double Blinding
- Triple Blinding
- Quadruple Blinding



Schulz, 2002

Blinding



Modified from Pearl, *Causality* 2009

Conclusions

Diagnostic RCTs: Methodological Lessons Learned

- For TB, patient-important outcomes??
- More comprehensive view of test evaluation
- Rare
  - Expensive
  - Time-intensive
  - Sample size requirements
  - Ethical limitations
  - Not always necessary

Acknowledgments

- Jack Colford, MD, PhD, Professor of Epidemiology, UC Berkeley