Diagnostic Prediction Models & Incremental value

Samuel G. Schumacher
McGill University

Outline

1. What is **incremental value** and why is it important
2. "*Binary test case*": **3 binary tests**
3. "*Multivariable model case*": **Dx prediction models**
4. **Imperfect reference standard**

What is it and why is it important

Accuracy: 2 tests

Accuracy: 2 tests

Reference Standard	True Positives (100%)	True Negatives (100%)

Accuracy: 2 tests

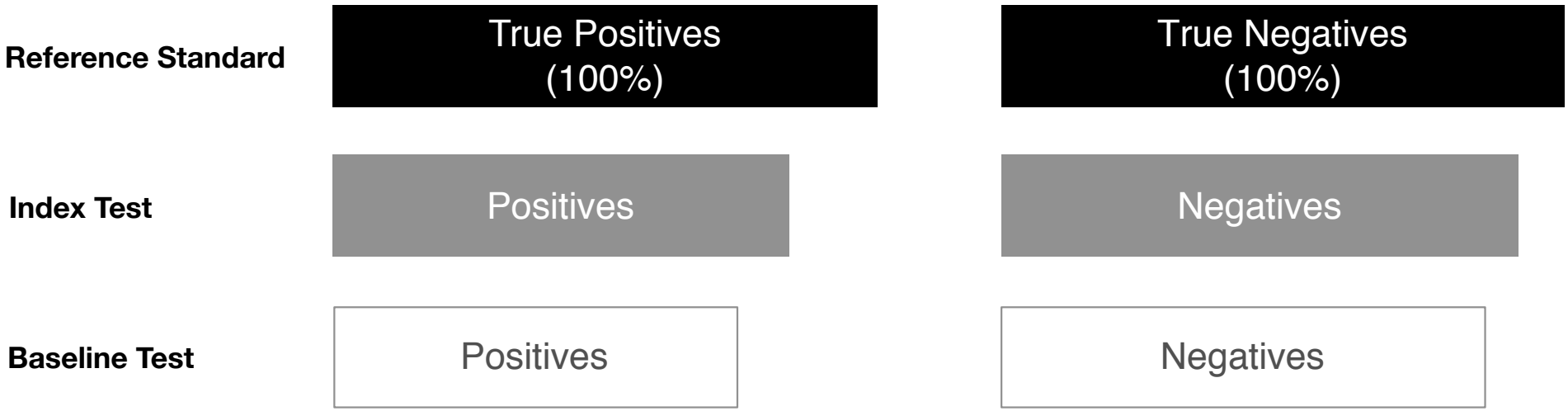
Reference Standard	True Positives (100%)	True Negatives (100%)
Index Test	Positives (% Sensitivity)	Negatives (% Specificity)

Accuracy: 2 tests

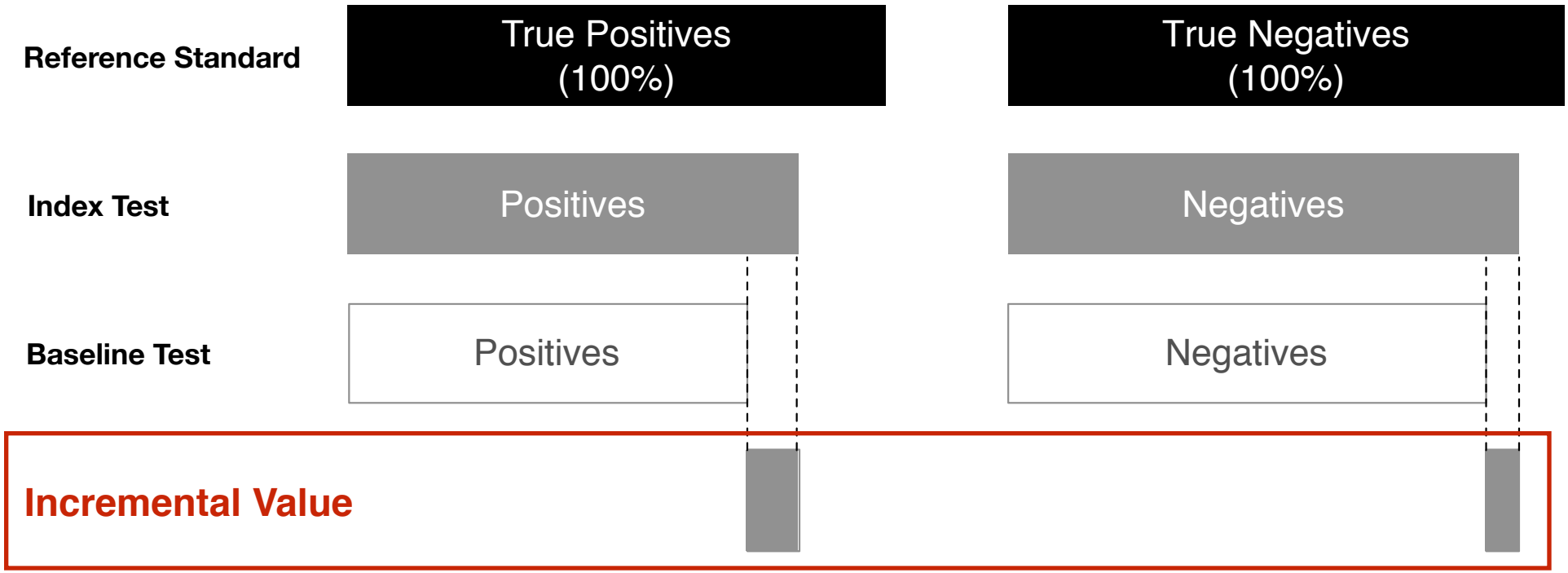
Reference Standard	True Positives (100%)	True Negatives (100%)
Index Test	Positives (% Sensitivity)	Negatives (% Specificity)

BUT: what matters is the **value beyond** what is already available (1st step towards impact)

Incremental value: 3 tests



Incremental value: 3 tests



Two basic paradigms

1. **"binary test case"**
   - a single test drives clinical decision making

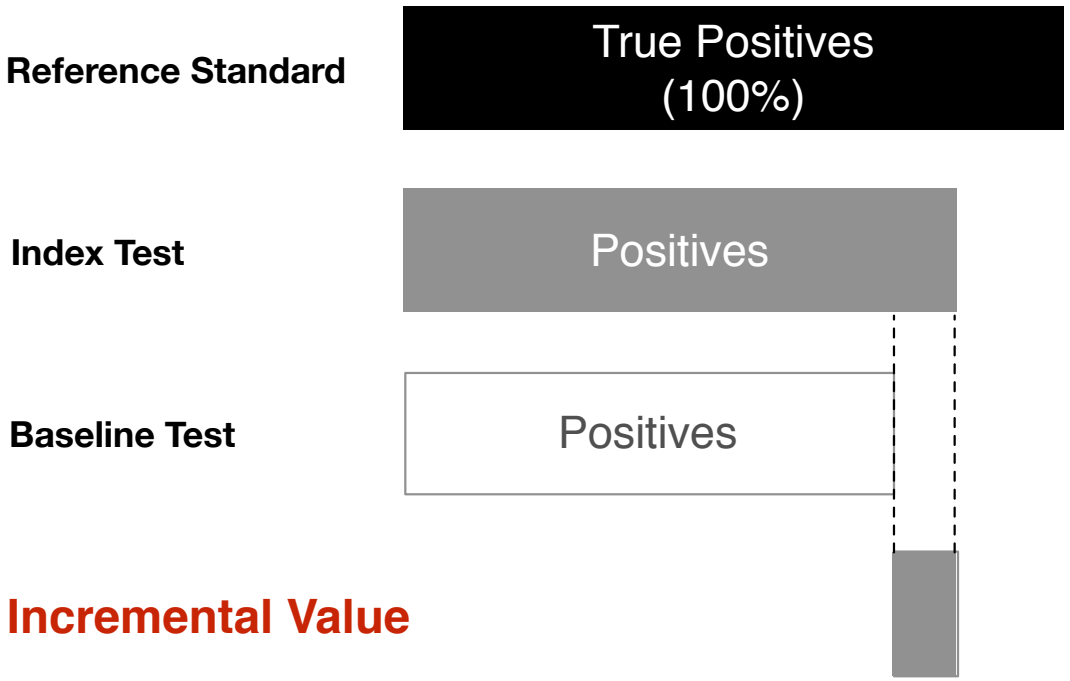
2. **"multivariable model case"**
   - patient demographics
   - clinical characteristics
   - signs and symptoms
   - perhaps multiple tests / biomarkers
   - perhaps tests that give quantitative results

“*Binary test case*”:

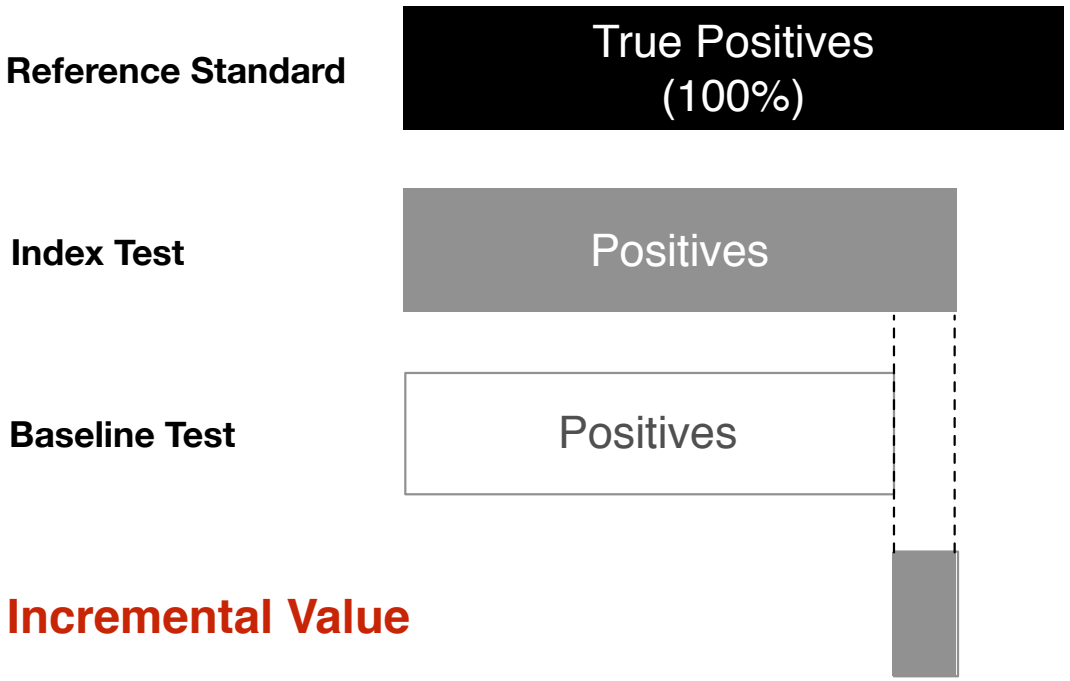
3 binary tests

Method

Method

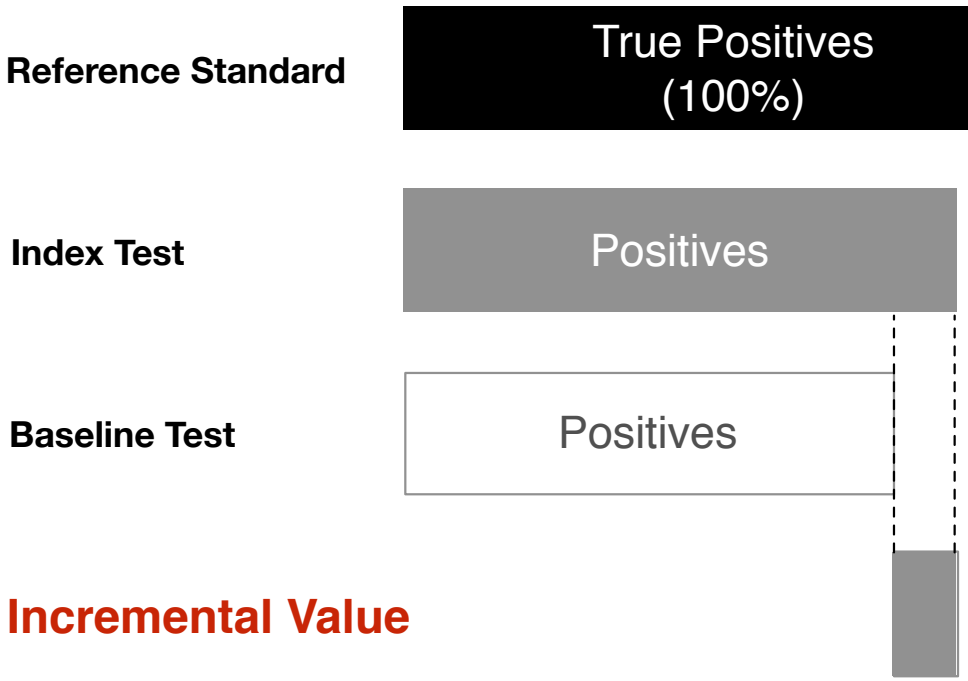

Reference Standard
True Positives
(100%)
Index Test
Positives
Baseline Test
Positives
Incremental Value

Method



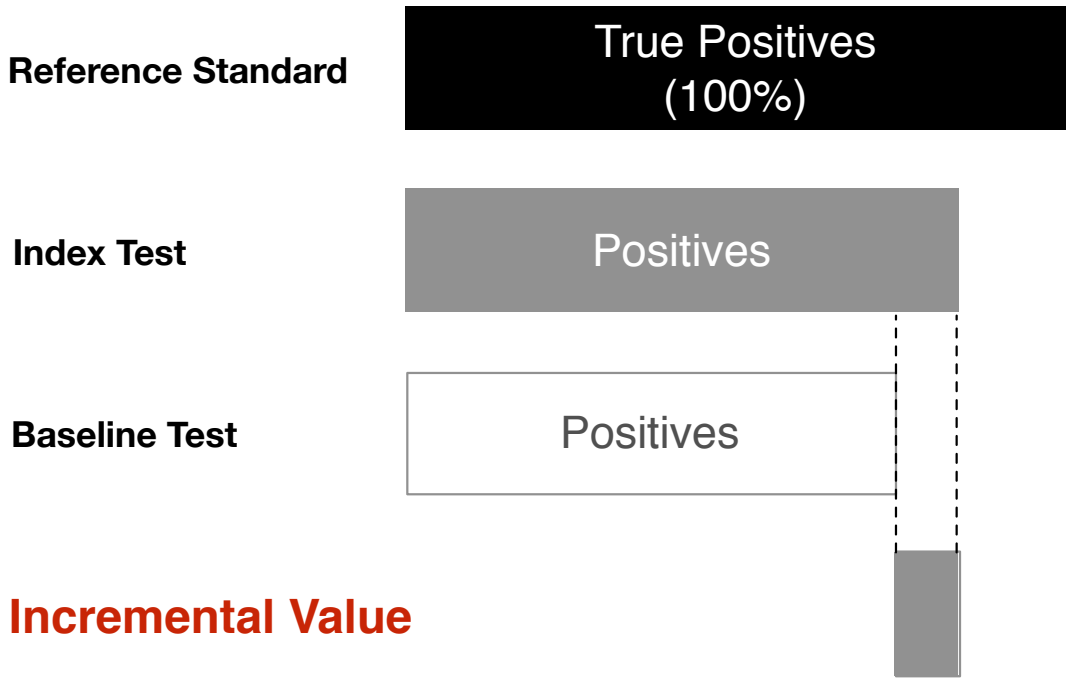
	n	%
Reference Standard	100	100%

Method



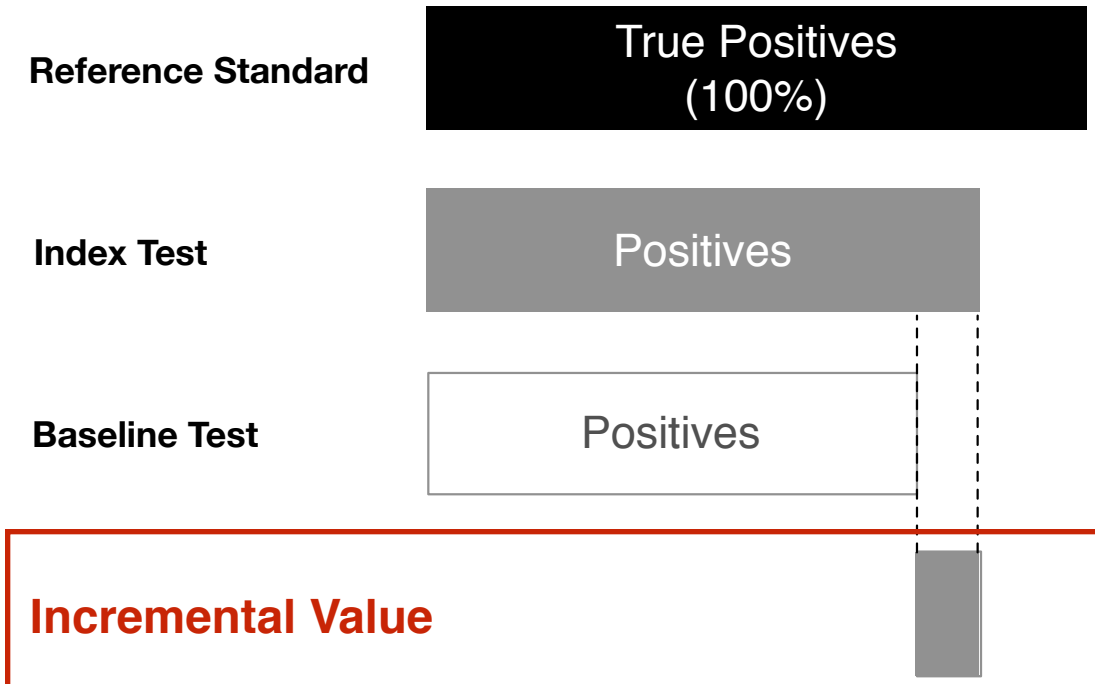
	n	%
Reference Standard	100	100%
Index Test	70	70%

Method



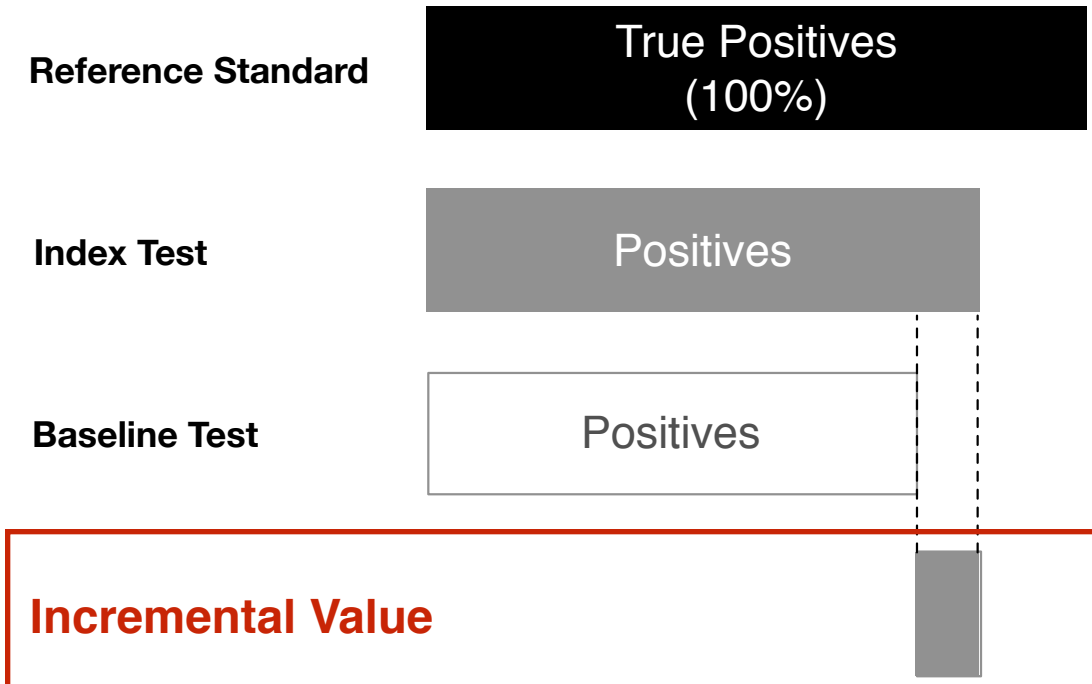
	n	%
Reference Standard	100	100%
Index Test	70	70%
Baseline Test	60	60%

Method



	n	%
Reference Standard	100	100%
Index Test	70	70%
Baseline Test	60	60%
Incremental Value	10	**10%**

Method



	n	%
Reference Standard	100	100%
Index Test	70	70%
Baseline Test	60	60%
Incremental Value	10	**10%**

➡ Difference in Sensitivity **or** ($Positives_{Index}$ - $Positives_{Baseline}$) / $Positives_{Reference}$

Serial sputum examinations

INT J TUBERC LUNG DIS 11(5):485–495

REVIEW ARTICLE

Yield of serial sputum specimen examinations in the diagnosis of pulmonary tuberculosis: a systematic review

S. R. Mase,*† A. Ramsay,‡ V. Ng,§ M. Henry,¶ P. C. Hopewell,*† J. Cunningham,‡ R. Urbanczik,# M. D. Perkins,** M. A. Aziz,†† M. Pai‡‡

* Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, San Francisco, † Francis J Curry National Tuberculosis Center, University of California, San Francisco, California, USA; ‡ United Nations Children's Fund/United Nations Development Programme/World Bank/World Health Organization (WHO) Special Programme for Research and Training in Tropical Diseases (TDR), WHO, Geneva, Switzerland; § Albany Medical College, Albany, New York, ¶ County of Sacramento Department of Health and Human Services, Sacramento, California, USA; # WHO Tuberculosis Laboratory Consultants Group, Schoemberg, Germany; ** Foundation for Innovative New Diagnostics (FIND), Geneva, †† Stop TB Department, WHO, Geneva, Switzerland; ‡‡ Department of Epidemiology & Biostatistics, McGill University, Montreal, Quebec, Canada

SUMMARY

Current international tuberculosis (TB) guidelines recommend the microscopic examination of three sputum specimens for acid-fast bacilli in the evaluation of persons suspected of having pulmonary TB. We conducted a systematic review of studies that quantified the diagnostic yield of each of three sputum specimens. By searching multiple databases and sources, we identified a total of 37 eligible studies. The incremental yield in smear-positive results (in studies using all smear-positive cases as the denominator) and the increase in sensitivity (in studies that used all culture-positive cases as the denominator) of the third specimen were the main outcomes of interest. Although heterogeneity in study methods and results presented challenges for data synthesis, subgroup analyses suggest that the average incremental yield and/or the increase in sensitivity of examining a third specimen ranged between 2% and 5%. Reducing the recommended number of specimens examined from three to two (particularly to two specimens collected on the same day) could benefit TB control programs, and potentially increase case detection for several reasons. A number of operational research issues need to be addressed. Studies examining the most effective and efficient means to utilize current technologies for microscopic examination of sputum would be most useful if they followed an internationally coordinated and standardized approach, both to strengthen the country-specific evidence base and to permit comparison among studies.

KEY WORDS: tuberculosis; smear microscopy; incremental yield; acid-fast bacillus; serial sputum specimens

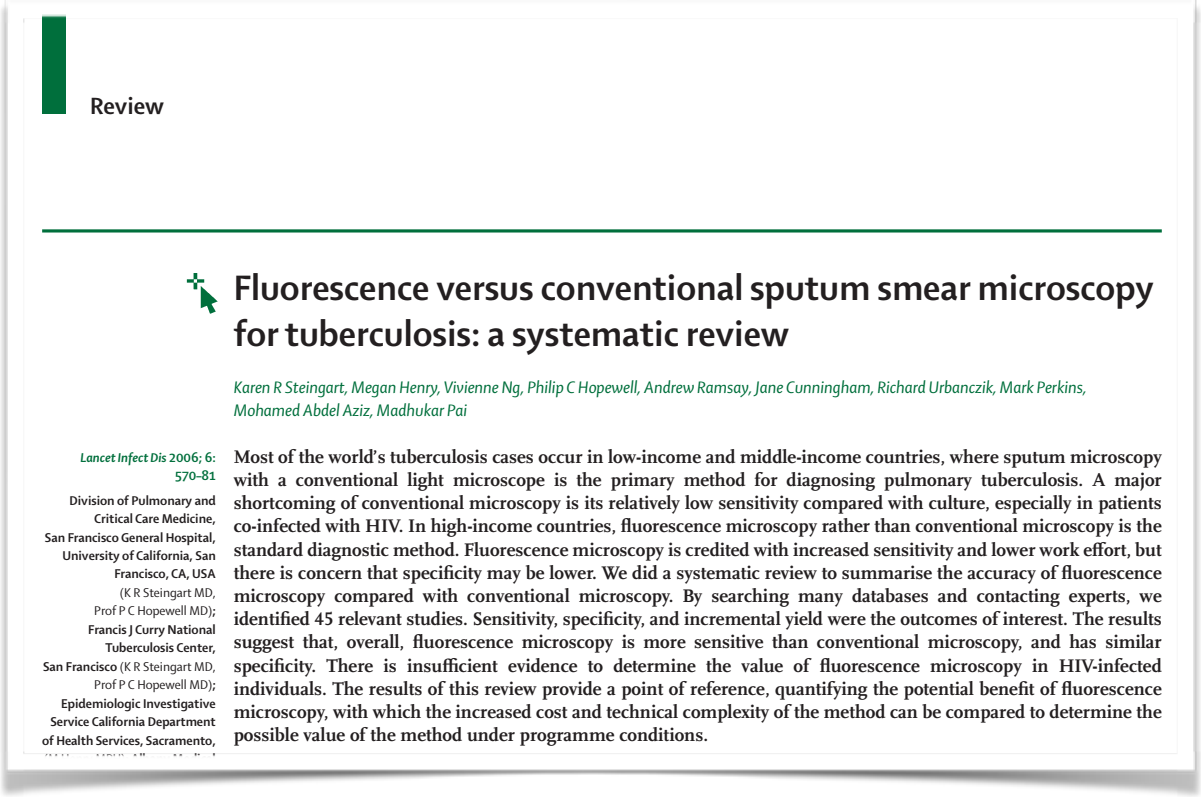
Methods

- **ref. stand.:** culture
- **baseline:** ZN 1 + ZN 2
- **Index test:** ZN 3
- **Stats:** ($Positives_{ZN\text{-}3}$ - $Positives_{ZN\text{-}1,2}$) / $Positives_{Culture}$

Results

- IV of third smear: 2-5%

FM versus ZN



Review

Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review

Karen R Steingart, Megan Henry, Vivienne Ng, Philip C Hopewell, Andrew Ramsay, Jane Cunningham, Richard Urbanczik, Mark Perkins, Mohamed Abdel Aziz, Madhukar Pai

Lancet Infect Dis 2006; 6: 570–81

Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, University of California, San Francisco, CA, USA (K R Steingart MD, Prof P C Hopewell MD); Francis J Curry National Tuberculosis Center, San Francisco (K R Steingart MD, Prof P C Hopewell MD); Epidemiologic Investigative Service California Department of Health Services, Sacramento,

Most of the world's tuberculosis cases occur in low-income and middle-income countries, where sputum microscopy with a conventional light microscope is the primary method for diagnosing pulmonary tuberculosis. A major shortcoming of conventional microscopy is its relatively low sensitivity compared with culture, especially in patients co-infected with HIV. In high-income countries, fluorescence microscopy rather than conventional microscopy is the standard diagnostic method. Fluorescence microscopy is credited with increased sensitivity and lower work effort, but there is concern that specificity may be lower. We did a systematic review to summarise the accuracy of fluorescence microscopy compared with conventional microscopy. By searching many databases and contacting experts, we identified 45 relevant studies. Sensitivity, specificity, and incremental yield were the outcomes of interest. The results suggest that, overall, fluorescence microscopy is more sensitive than conventional microscopy, and has similar specificity. There is insufficient evidence to determine the value of fluorescence microscopy in HIV-infected individuals. The results of this review provide a point of reference, quantifying the potential benefit of fluorescence microscopy, with which the increased cost and technical complexity of the method can be compared to determine the possible value of the method under programme conditions.

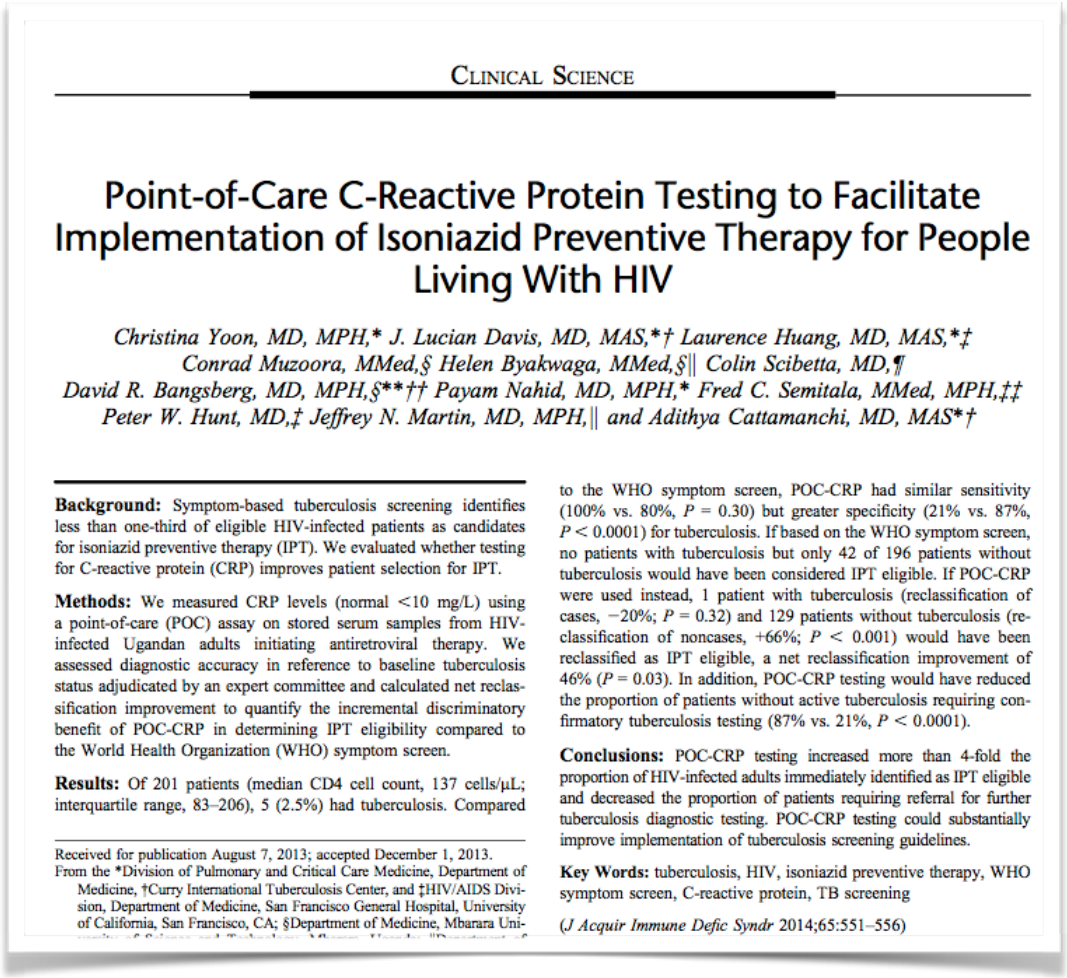
Methods

- **ref. stand.:** culture
- **baseline:** ZN
- **Index test:** FM
- **Stats:** Δ Sensitivity

Results

- IV of FM over ZN: 10%

CRP vs WHO symptom screen to determine IPT eligibility



CLINICAL SCIENCE

Point-of-Care C-Reactive Protein Testing to Facilitate Implementation of Isoniazid Preventive Therapy for People Living With HIV

Christina Yoon, MD, MPH, J. Lucian Davis, MD, MAS,*† Laurence Huang, MD, MAS,*‡ Conrad Muzoora, MMed,§ Helen Byakwaga, MMed,§‖ Colin Scibetta, MD,¶ David R. Bangsberg, MD, MPH,$**†† Payam Nahid, MD, MPH,* Fred C. Semitala, MMed, MPH,‡‡ Peter W. Hunt, MD,‡ Jeffrey N. Martin, MD, MPH,‖ and Adithya Cattamanchi, MD, MAS*†*

Background: Symptom-based tuberculosis screening identifies less than one-third of eligible HIV-infected patients as candidates for isoniazid preventive therapy (IPT). We evaluated whether testing for C-reactive protein (CRP) improves patient selection for IPT.

Methods: We measured CRP levels (normal <10 mg/L) using a point-of-care (POC) assay on stored serum samples from HIV-infected Ugandan adults initiating antiretroviral therapy. We assessed diagnostic accuracy in reference to baseline tuberculosis status adjudicated by an expert committee and calculated net reclassification improvement to quantify the incremental discriminatory benefit of POC-CRP in determining IPT eligibility compared to the World Health Organization (WHO) symptom screen.

Results: Of 201 patients (median CD4 cell count, 137 cells/μL; interquartile range, 83–206), 5 (2.5%) had tuberculosis. Compared to the WHO symptom screen, POC-CRP had similar sensitivity (100% vs. 80%, $P = 0.30$) but greater specificity (21% vs. 87%, $P < 0.0001$) for tuberculosis. If based on the WHO symptom screen, no patients with tuberculosis but only 42 of 196 patients without tuberculosis would have been considered IPT eligible. If POC-CRP were used instead, 1 patient with tuberculosis (reclassification of cases, −20%; $P = 0.32$) and 129 patients without tuberculosis (reclassification of noncases, +66%; $P < 0.001$) would have been reclassified as IPT eligible, a net reclassification improvement of 46% ($P = 0.03$). In addition, POC-CRP testing would have reduced the proportion of patients without active tuberculosis requiring confirmatory tuberculosis testing (87% vs. 21%, $P < 0.0001$).

Conclusions: POC-CRP testing increased more than 4-fold the proportion of HIV-infected adults immediately identified as IPT eligible and decreased the proportion of patients requiring referral for further tuberculosis diagnostic testing. POC-CRP testing could substantially improve implementation of tuberculosis screening guidelines.

Key Words: tuberculosis, HIV, isoniazid preventive therapy, WHO symptom screen, C-reactive protein, TB screening

(*J Acquir Immune Defic Syndr* 2014;65:551–556)

Received for publication August 7, 2013; accepted December 1, 2013.

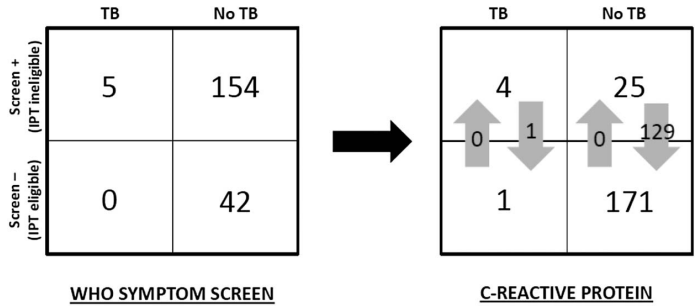
From the *Division of Pulmonary and Critical Care Medicine, Department of Medicine, †Curry International Tuberculosis Center, and ‡HIV/AIDS Division, Department of Medicine, San Francisco General Hospital, University of California, San Francisco, CA; §Department of Medicine, Mbarara Uni-

Methods

- **ref. stand.:** expert panel
- **baseline:** WHO symp. screen
- **Index test:** CRP
- **Stats:** Δ Sensitivity, Δ Specificity

Results

- Δ Sensitivity = -20%
- Δ Specificity = +66%
- NRI = 46%



Limitations of this Methodology

1. Works for simple situations where
   - all tests are **binary**
   - baseline test is **one** single test

2. definition of incremental value *"technical"*
   - **"*practical*"** incremental value may differ
   - other methods needed to assess "*practical*" incremental value

What if…

…“the baseline test” is not a single binary test

What if…

…“the baseline test” is not a single binary test

- routine for chronic diseases

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common
- in TB

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common
- in TB
  - extra-pulmonary TB

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common
- in TB
  - extra-pulmonary TB
  - smear-negative TB

What if…

…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common
- in TB
  - extra-pulmonary TB
  - smear-negative TB
  - paediatric TB

What if…

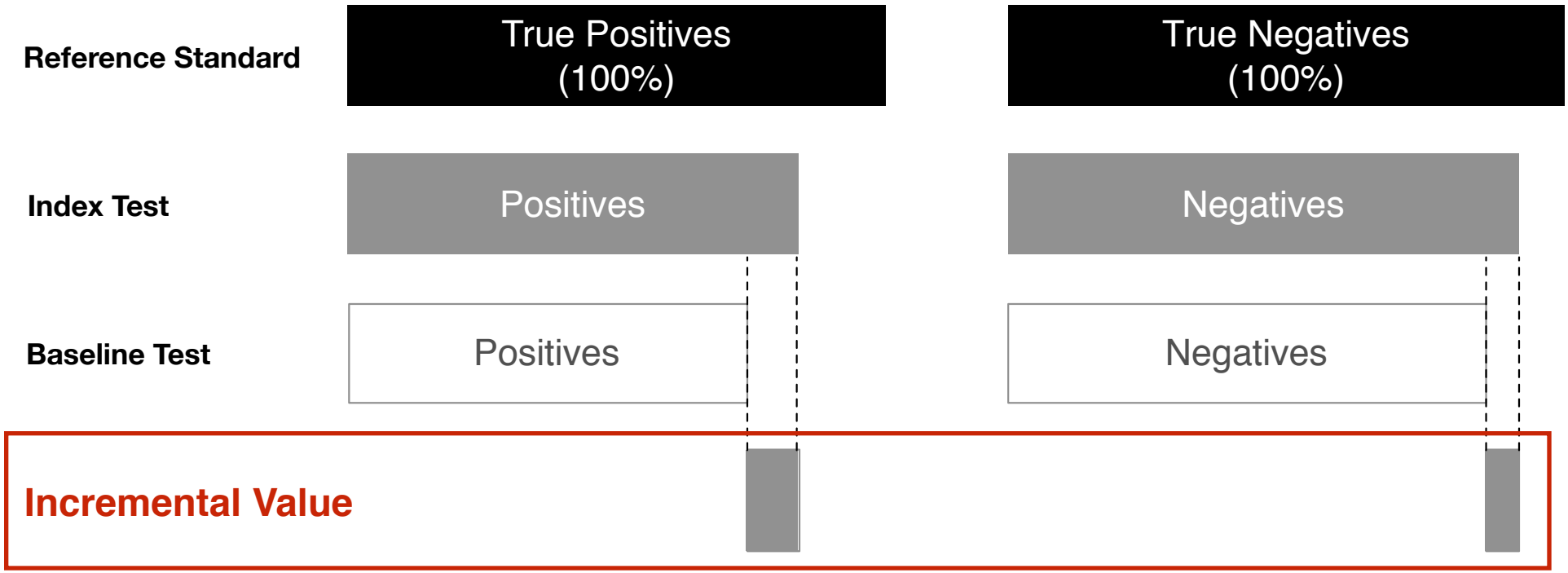
…"the baseline test" is not a single binary test

- routine for chronic diseases
  - multiple tests & characteristics (multivariable)
  - prognosis rather than diagnosis (probabilistic)
  - prediction models or decision rules more common
- in TB
  - extra-pulmonary TB
  - smear-negative TB
  - paediatric TB
  - latent TB infection

"*Multivariable model case*": **Dx prediction models**

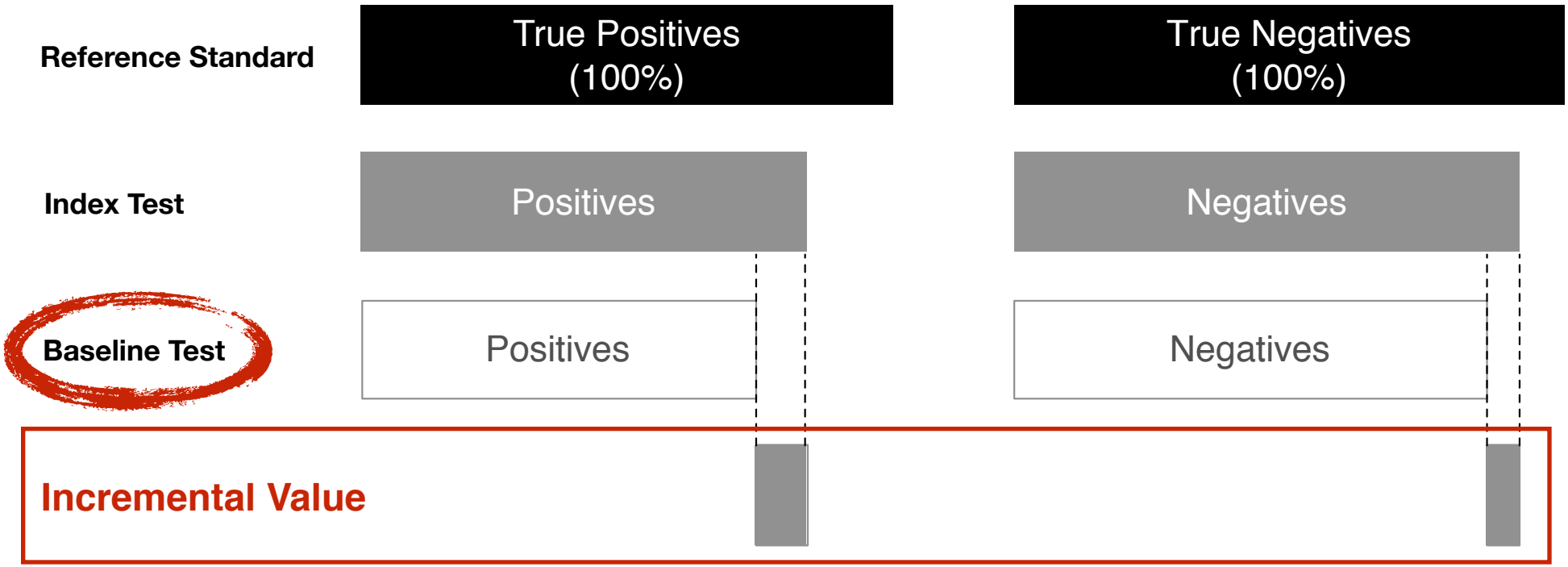
Incremental value

"Multivariable model case"



Incremental value

"Multivariable model case"



Differences to “binary test case”

Differences to "binary test case"

1. information of baseline tests needs to be combined
   - diagnostic prediction model

Differences to "binary test case"

1. information of baseline tests needs to be combined
   - diagnostic prediction model

2. incremental value is the change you observe when adding the index test to the baseline model
   - assessing this *change* needs different metrics

Differences to "binary test case"

1. information of baseline tests needs to be combined
   - diagnostic prediction model

2. incremental value is the change you observe when adding the index test to the baseline model
   - assessing this *change* needs different metrics

3. output of models is not binary but a probability
   - need to think about cut-offs & decision-making

Differences to "binary test case"

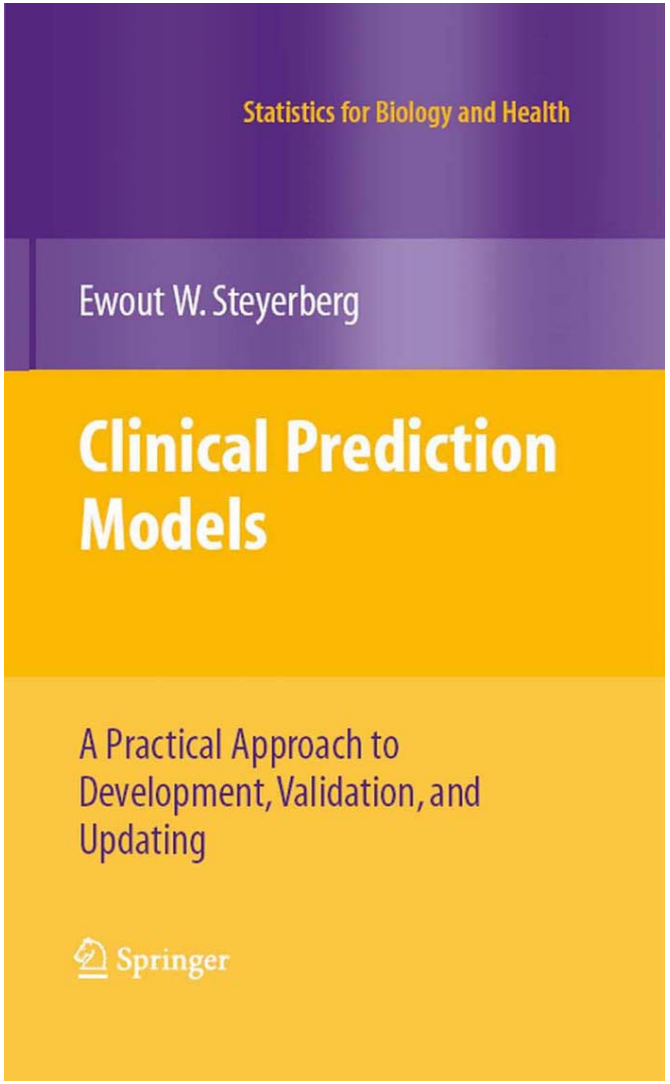
1. information of baseline tests needs to be combined
   - diagnostic prediction model

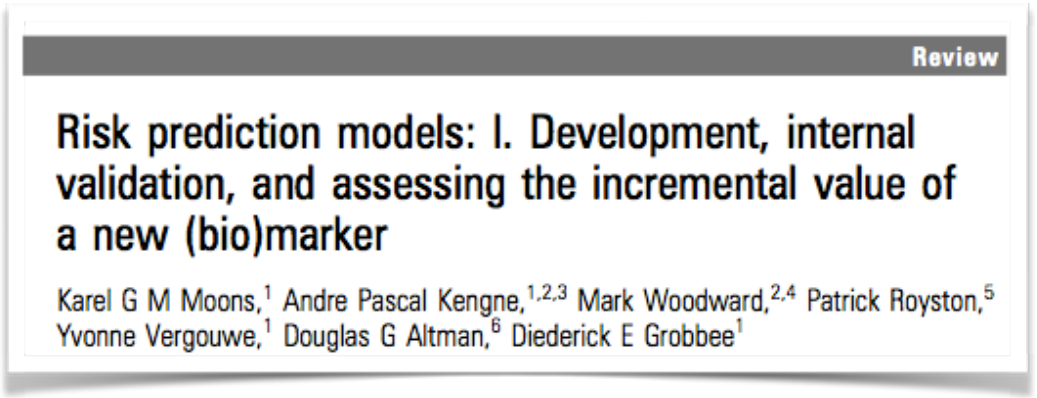
2. incremental value is the change you observe when adding the index test to
   - assessing this

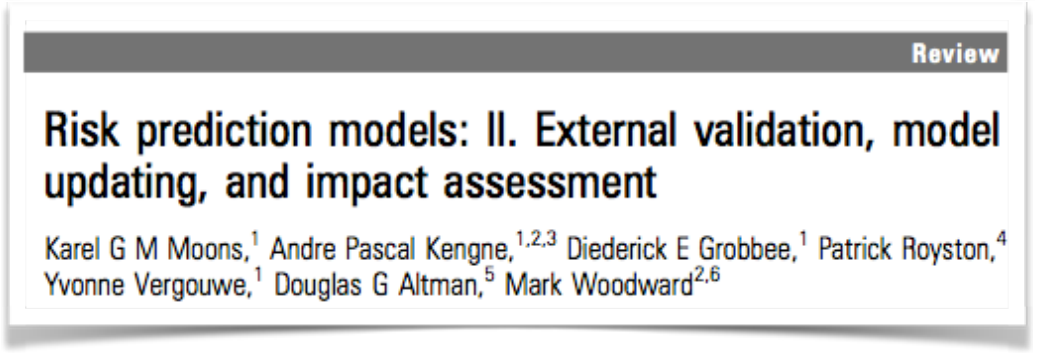
3. output of models is not binary but a probability
   - need to think about cut-offs & decision-making

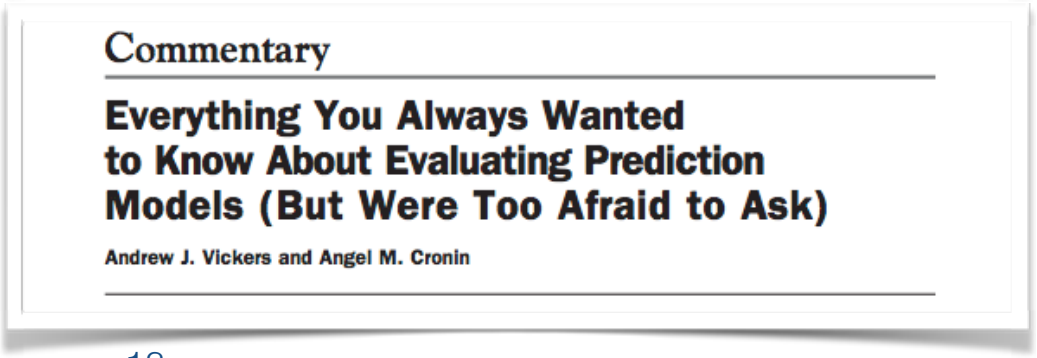
Prediction Models

Further readings









Prediction Models

1. **Model development**
   - logistic regression model / machine learning
   - modelling continuous predictors
   - predictor selection

2. **Model assessment**
   - discrimination (e.g. AUC)
   - calibration (e.g. calibration plot)

3. **Model validation**
   - internal validation
   - external validation

Differences to "binary test case"

1. information of baseline tests needs to be combined
   - diagnostic prediction model

2. incremental value is the change you observe when adding the index test to the baseline model
   - assessing this *change* needs different metrics

3. output of models is not binary but a probability
   - need to think about cut-offs & decision-making

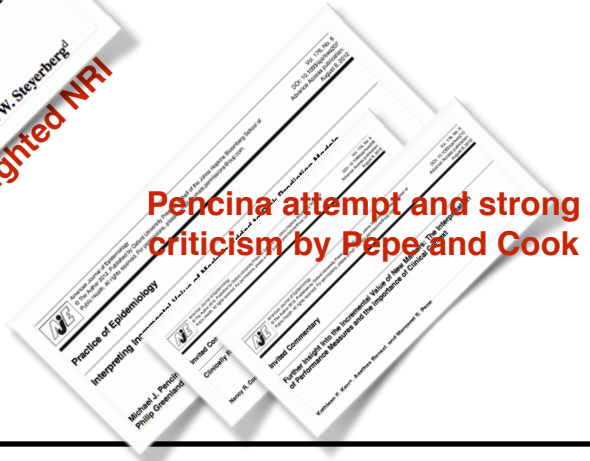
risk stratification (reclassification) table

NRI, IDI, case-stratified reclassification table

Greenland criticism

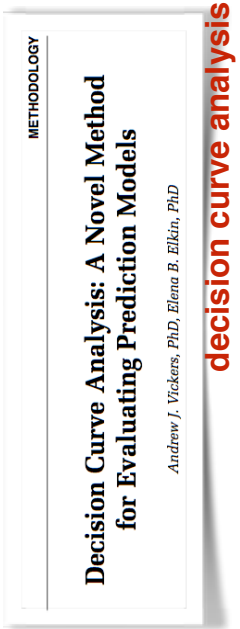
category-free and weighted NRI

very good reviews

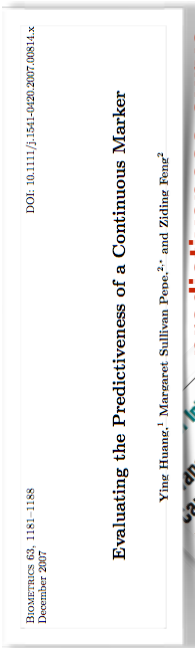


Pencina attempt and strong criticism by Pepe and Cook

2006 2007 2008 2009 2010 2011 2012 2013 2014



decision curve analysis



predictiveness curve

more analysis around risk stratification table

Pepe criticism of NRI and certain analyses and testing around reclassification table

very critical editorial of NRI

very critical review of NRI

Incremental Value

Further readings

DOI: 10.1111/j.1365-2362.2011.02562.x

METHODS

Assessing the incremental value of diagnostic and prognostic markers: a review and illustration

Ewout W. Steyerberg[*], Michael J. Pencina[†], Hester F. Lingsma[*], Michael W. Kattan[‡], Andrew J. Vickers[§] and Ben Van Calster[*,¶]

Clinical Chemistry 58:10
1408–1417 (2012)

Review

Quantifying the Added Value of a Diagnostic Test or Marker

Karel G.M. Moons,[1*†] Joris A.H. de Groot,[1†] Kristian Linnet,[2] Johannes B. Reitsma,[1] and Patrick M.M. Bossuyt[3]

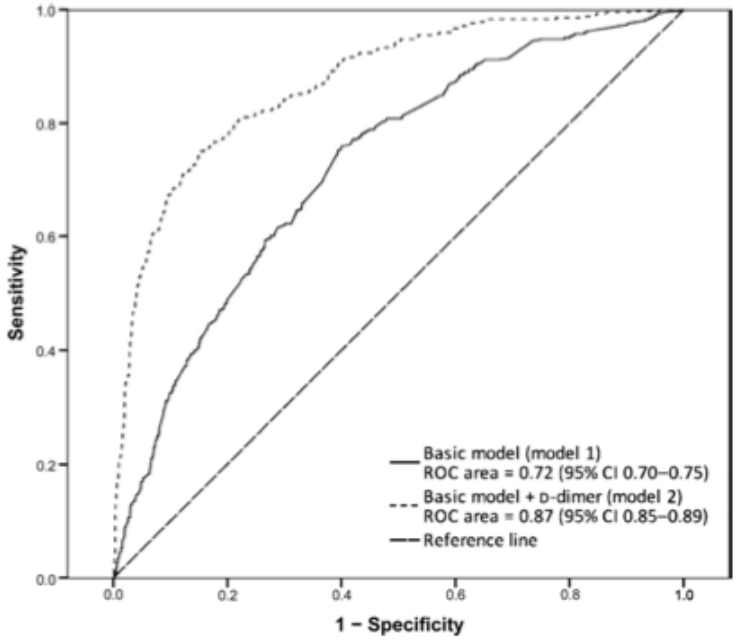
Incremental Value

Limitations of Selected Metrics

Incremental Value

Limitations of Selected Metrics

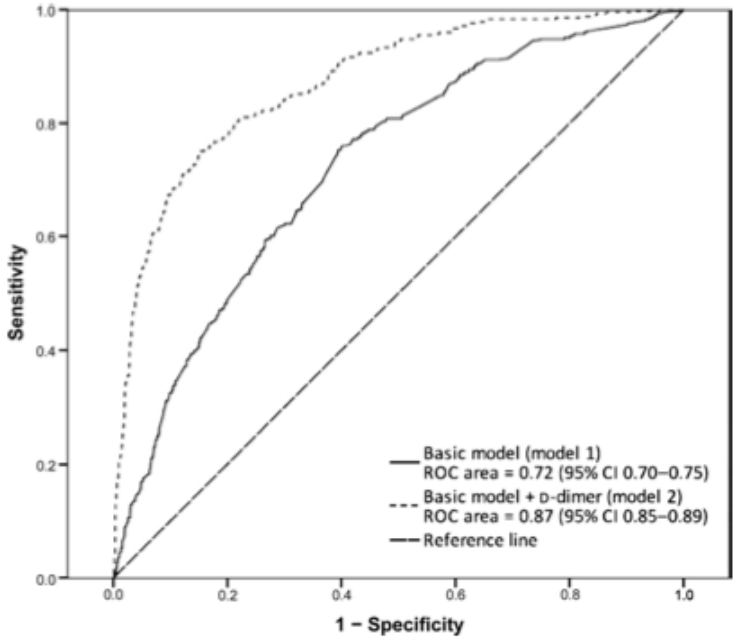
Area Under The ROC Curve & Integr. Discrimin. Improv.



Incremental Value

Limitations of Selected Metrics

Area Under The ROC Curve & Integr. Discrimin. Improv.

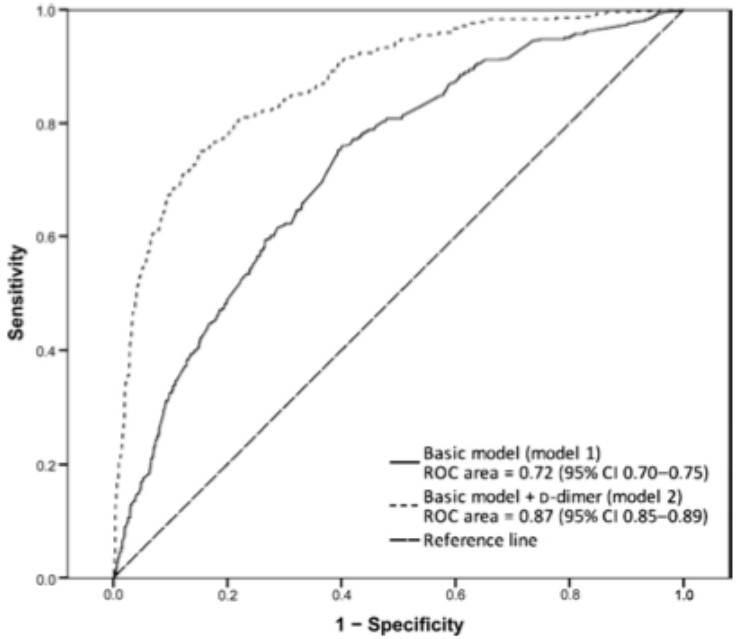


- clinical interpretation
- incorporate irrelevant info.
- does not incorporate info. on consequences

Incremental Value

Limitations of Selected Metrics

Area Under The ROC Curve & Integr. Discrimin. Improv.



Reclassification table & Net Reclass. Index

DVT yes (n = 416)			
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	92	123	215
>25	26	175	201
Total	118	298	416

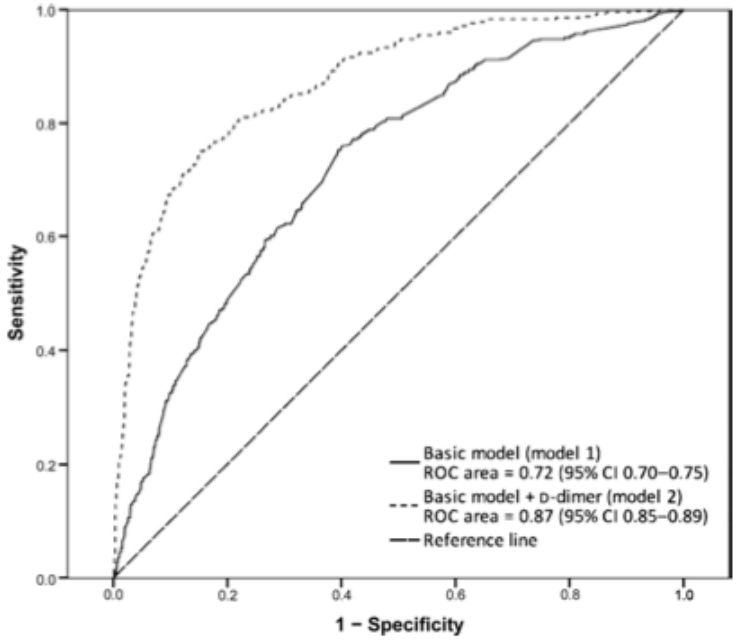
DVT no (n = 1670)			
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	1223	116	1339
>25	227	104	331
Total	1450	220	1670

- clinical interpretation
- incorporate irrelevant info.
- does not incorporate info. on consequences

Incremental Value

Limitations of Selected Metrics

Area Under The ROC Curve & Integr. Discrimin. Improv.



- clinical interpretation
- incorporate irrelevant info.
- does not incorporate info. on consequences

Reclassification table & Net Reclass. Index

DVT yes (n = 416)			
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	92	123	215
>25	26	175	201
Total	118	298	416

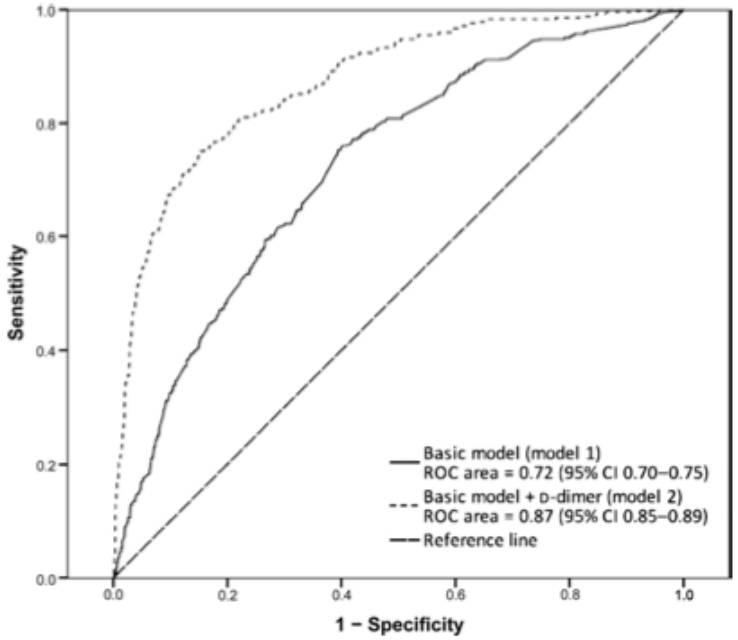
DVT no (n = 1670)			
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	1223	116	1339
>25	227	104	331
Total	1450	220	1670

- choice of cut-offs
- weighting / “nature” of reclassification
- does not incorporate info. on consequences

Incremental Value

Limitations of Selected Metrics

Area Under The ROC Curve & Integr. Discrimin. Improv.



- clinical interpretation
- incorporate irrelevant info.
- does not incorporate info. on consequences

Reclassification table & Net Reclass. Index

DVT yes (n = 416)

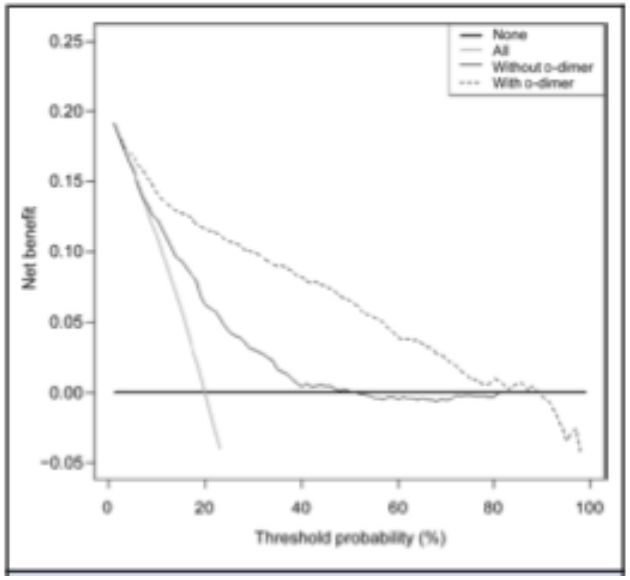
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	92	123	215
>25	26	175	201
Total	118	298	416

DVT no (n = 1670)

	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	1223	116	1339
>25	227	104	331
Total	1450	220	1670

- choice of cut-offs
- weighting / “nature” of reclassification
- does not incorporate info. on consequences

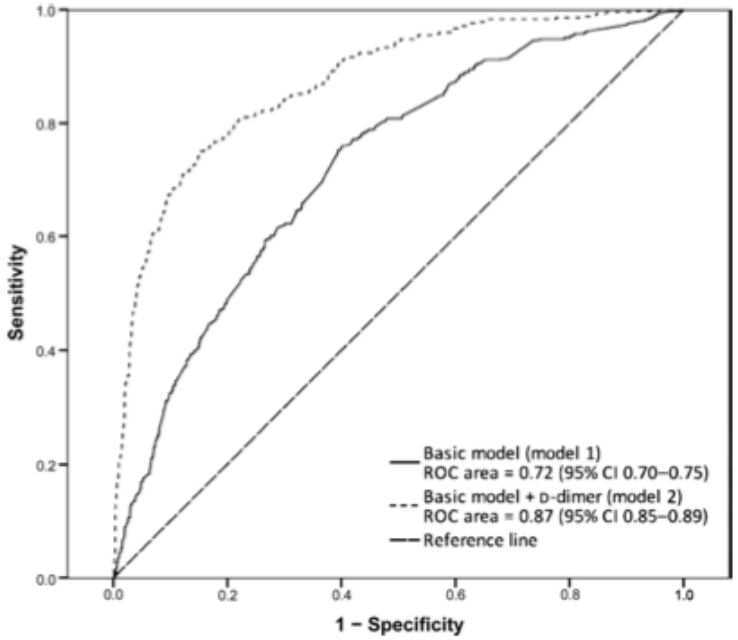
Decision Curve Analysis & Net Benefit



Incremental Value

Limitations of Selected Metrics

Area Under The ROC Curve & Integr. Discrimin. Improv.



- clinical interpretation
- incorporate irrelevant info.
- does not incorporate info. on consequences

Reclassification table & Net Reclass. Index

DVT yes (n = 416)

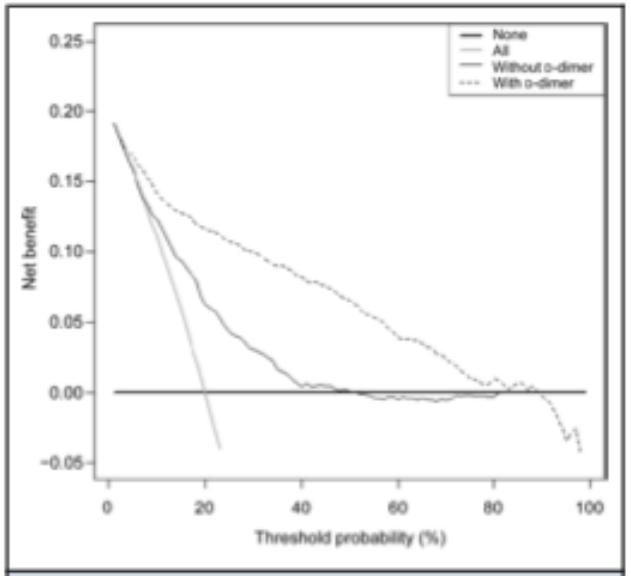
	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	92	123	215
>25	26	175	201
Total	118	298	416

DVT no (n = 1670)

	Model 2 with D-dimer		
	≤25	>25	Total
Model 1 without D-dimer			
≤25	1223	116	1339
>25	227	104	331
Total	1450	220	1670

- choice of cut-offs
- weighting / “nature” of reclassification
- does not incorporate info. on consequences

Decision Curve Analysis & Net Benefit



- requires thinking about cut-offs (decision-thresholds)
- perhaps less easily interpreted

Differences to "binary test case"

1. information of baseline tests needs to be combined
   - diagnostic prediction model

2. incremental value assessed by adding index test to
   - assessing incremental value needs different metrics

3. output of models is not binary but a probability
   - need to think about cut-offs & decision-making

LAM for TB Meningitis

OPEN ACCESS Freely available online

PLoS one

Comparison of a Clinical Prediction Rule and a LAM Antigen-Detection Assay for the Rapid Diagnosis of TBM in a High HIV Prevalence Setting

Vinod B. Patel[1], Ravesh Singh[2], Cathy Connolly[3], Victoria Kasprowicz[2], Allimudin Zumla[4], Thumbi Ndungu[2], Keertan Dheda[4,5,6]*

1 Department of Neurology, University of KwaZulu Natal, Berea, South Africa, **2** HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R. Mandela School of Medicine, University of KwaZulu Natal, Berea, South Africa, **3** Biostatistics Unit, Medical Research Council, Durban, South Africa, **4** Department of Infection, Centre for Infectious Diseases and International Health, University College London, London, United Kingdom, **5** Lung Infection and Immunity Unit, Division of Pulmonology and Department of Medicine, University of Cape Town Lung Institute, University of Cape Town, Cape Town, South Africa, **6** Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

Abstract

Background/Objective: The diagnosis of tuberculous meningitis (TBM) in resource poor TB endemic environments is challenging. The accuracy of current tools for the rapid diagnosis of TBM is suboptimal. We sought to develop a clinical-prediction rule for the diagnosis of TBM in a high HIV prevalence setting, and to compare performance outcomes to conventional diagnostic modalities and a novel lipoarabinomannan (LAM) antigen detection test (Clearview-TB®) using cerebrospinal fluid (CSF).

Methods: Patients with suspected TBM were classified as definite-TBM (CSF culture or PCR positive), probable-TBM and non-TBM.

Results: Of the 150 patients, 84% were HIV-infected (median [IQR] CD4 count = 132 [54; 241] cells/μl). There were 39, 55 and 54 patients in the definite, probable and non-TBM groups, respectively. The LAM sensitivity and specificity (95%CI) was 31% (17;48) and 94% (85;99), respectively (cut-point ≥0.18). By contrast, smear-microscopy was 100% specific but detected none of the definite-TBM cases. LAM positivity was associated with HIV co-infection and low CD4 T cell count (CD4<200 vs. >200 cells/μl; p = 0.03). The sensitivity and specificity in those with a CD4<100 cells/μl was 50% (27;73) and 95% (74;99), respectively. A clinical-prediction rule ≥6 derived from multivariate analysis had a sensitivity and specificity (95%CI) of 47% (31;64) and 98% (90;100), respectively. When LAM was combined with the clinical-prediction-rule, the sensitivity increased significantly (p<0.001) to 63% (47;68) and specificity remained high at 93% (82;98).

Conclusions: Despite its modest sensitivity the LAM ELISA is an accurate rapid rule-in test for TBM that has incremental value over smear-microscopy. The rule-in value of LAM can be further increased by combination with a clinical-prediction rule, thus enhancing the rapid diagnosis of TBM in HIV-infected persons with advanced immunosuppression.

Method

- **ref. stand.**: culture or PCR
- **baseline:** logistic regression model (clinical predictors and lab tests)
- **Index test:** LAM ELISA

Result

- Δ Sensitivity = +16%
- Δ Specificity = -5%

LAM for TB Meningitis

Methods

LAM for TB Meningitis

Methods

1 logistic regression model

Table 4. Univariable and multivariable analysis for the prediction of definite TB meningitis.

Characteristic	OR	95%CI	p value	β coefficient	Score
Univariate analysis					
Lymphocyte count >200 (cells/μl)	6.5	(2–22)	0.003		
Neutrophil count ≥36 (cells/μl)	5.0	(2–12)	<0.001		
Protein Level ≥2.5 g/l	3.6	(1–10)	0.02		
CSF glucose ≤1 mmol/l	8.4	(3–24)	<0.001		
Ratio of CSF/serum glucose ≤0.2	9.3	(3–28)	<0.001		
CD4 count (<200 cells/μl)	2.9	(1–7)	0.03		
CLAT test (NEG)	8.7	(3–28)	<0.001		
Previous TB (no)	3.1	(1.2–8.0)	0.02		
Multivariate analysis					
Ratio of CSF/serum glucose ≤0.2	7.1	(1.8–29)	0.006	2	2
Lymphocyte count >200 (cells/μl)	7.6	(1.5–40)	0.017	2	2
CD4 count (<200cells/μl)	6.8	(1.9–24)	0.003	1.9	2
CLAT test (NEG)	12.9	(3–52)	<0.001	2.6	3

LAM for TB Meningitis

Methods

1 logistic regression model

2 chose cut-off

Table 4. Univariable and multivariable analysis for the prediction of definite TB meningitis.

Characteristic	OR	95%CI	p value	β coefficient	Score
Univariate analysis					
Lymphocyte count >200 (cells/μl)	6.5	(2–22)	0.003		
Neutrophil count ≥36 (cells/μl)	5.0	(2–12)	<0.001		
Protein Level ≥2.5 g/l	3.6	(1–10)	0.02		
CSF glucose ≤1 mmol/l	8.4	(3–24)	<0.001		
Ratio of CSF/serum glucose ≤0.2	9.3	(3–28)	<0.001		
CD4 count (<200 cells/μl)	2.9	(1–7)	0.03		
CLAT test (NEG)	8.7	(3–28)	<0.001		
Previous TB (no)	3.1	(1.2–8.0)	0.02		
Multivariate analysis					
Ratio of CSF/serum glucose ≤0.2	7.1	(1.8–29)	0.006	2	2
Lymphocyte count >200 (cells/μl)	7.6	(1.5–40)	0.017	2	2
CD4 count (<200cells/μl)	6.8	(1.9–24)	0.003	1.9	2
CLAT test (NEG)	12.9	(3–52)	<0.001	2.6	3

LAM for TB Meningitis

Methods

1 logistic regression model

2 chose cut-off

3 compare sens/spec

Table 4. Univariable and multivariable analysis for the prediction of definite TB meningitis.

Characteristic	OR	95%CI	p value	β coefficient	Score
Univariate analysis					
Lymphocyte count >200 (cells/μl)	6.5	(2–22)	0.003		
Neutrophil count ≥36 (cells/μl)	5.0	(2–12)	<0.001		
Protein Level ≥2.5 g/l	3.6	(1–10)	0.02		
CSF glucose ≤1 mmol/l	8.4	(3–24)	<0.001		
Ratio of CSF/serum glucose ≤0.2	9.3	(3–28)	<0.001		
CD4 count (<200 cells/μl)	2.9	(1–7)	0.03		
CLAT test (NEG)	8.7	(3–28)	<0.001		
Previous TB (no)	3.1	(1.2–8.0)	0.02		
Multivariate analysis					
Ratio of CSF/serum glucose ≤0.2	7.1	(1.8–29)	0.006	2	2
Lymphocyte count >200 (cells/μl)	7.6	(1.5–40)	0.017	2	2
CD4 count (<200cells/μl)	6.8	(1.9–24)	0.003	1.9	2
CLAT test (NEG)	12.9	(3–52)	<0.001	2.6	3

Definite TBM (n = 38)* compared to unselected Non TBM (n = 5

Cut Point	Sens (CI)	Spec (CI)
CPR[†] ≥4	87%	70%
(excluding LAM)	(72;96)	(56;82)
CPR (≥6)	47%[‡]	98%
(excluding LAM)	(31;64)	(90;100)
LAM (OD) ≥0.18	31%[#]	94%
	(17;48)	(85;99)
CPR (≥4) + LAM	89%	65%
	(75;97)	(51;77)
CPR (≥6) + LAM	63%[‡, #]	93%
	(46–78)	(82;98)

IGRA & TST for pre-IPT assessment in HIV+

Eur Respir J 2012; 39: 163–171
DOI: 10.1183/09031936.00058911

Interferon release does not add discriminatory value to smear-negative HIV–tuberculosis algorithms

M.X. Rangaka*,#,¶, H.P. Gideon#, K.A. Wilkinson#,+,§, M. Pai/, J. Mwansa-Kambafwile#, G. Maartens§,, J.R. Glynn¶, A. Boulle*, K. Fielding¶, R. Goliath#, R. Titus#, S. Mathee## and R.J. Wilkinson#,+,§,¶¶**

ABSTRACT: Clinical algorithms for evaluating HIV-infected individuals for tuberculosis (TB) prior to isoniazid preventive therapy (IPT) perform poorly, and interferon-γ release assays (IGRAs) have moderate accuracy for active TB. It is unclear whether, when used as adjunct tests, IGRAs add any clinical discriminatory value for active TB diagnosis in the pre-IPT assessment.

779 sputum smear-negative HIV-infected persons, established on or about to commence combined antiretroviral therapy (ART), were screened for TB prior to IPT. Stepwise multivariable logistic regression was used to develop clinical prediction models. The discriminatory ability was assessed by receiver operator characteristic area under the curve (AUC). QuantiFERON®-TB Gold in-tube (QFT-GIT) was evaluated.

The prevalence of smear-negative TB by culture was 6.4% (95% CI 4.9–8.4%). Used alone, QFT-GIT and the tuberculin skin test (TST) had comparable performance; the post-test probability of disease based on single negative tests was 3–4%. In a multivariable model, the QFT-GIT test did not improve the ability of a clinical algorithm, which included not taking ART, weight <60 kg, no prior history of TB, any one positive TB symptom/sign (cough ⩾2 weeks) and CD4+ count <250 cells per mm^3, to discriminate smear-negative culture-positive and -negative TB (72% to 74%; AUC comparison p=0.33). The TST marginally improved the discriminatory ability of the clinical model (to 77%, AUC comparison p=0.04).

QFT-GIT does not improve the discriminatory ability of current TB screening clinical algorithms used to evaluate HIV-infected individuals for TB ahead of preventive therapy. Evaluation of new TB diagnostics for clinical relevance should follow a multivariable process that goes beyond test accuracy.

AFFILIATIONS
*Centre for Infectious Disease and Epidemiology Research (CIDER), #Clinical Infectious Diseases Research Initiative, Institute of Infectious Diseases and Molecular Medicine, §Dept of Medicine, **Division of Clinical Pharmacology, University of Cape Town, and ##Provincial Administration of the Western Cape, Cape Town, South Africa. ¶London School of Hygiene and Tropical Medicine, +MRC National Institute for Medical Research, and ¶¶Division of Medicine, Imperial College London, London, UK. /McGill University and Montreal Chest Institute, Montreal, QC, Canada.

CORRESPONDENCE
M.X. Rangaka
Centre for Infectious Disease and

Method

- **ref. stand.:** culture
- **baseline:** logistic regression model (demographics, clinical, CD4)
- **Index tests:** IGRA, TST

Results

- Δ AUC(IGRA) =+2%
- Δ AUC(TST) = +5%

IGRA & TST for pre-IPT assessment in HIV+

Methods

IGRA & TST for pre-IPT assessment in HIV+

Methods

1 logistic regression model

IGRA & TST for pre-IPT assessment in HIV+

Methods

1 logistic regression model

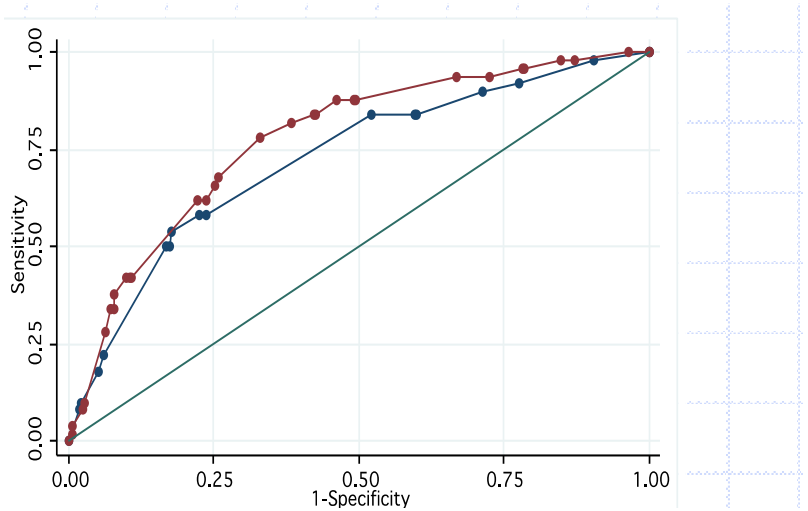
2 compare AUC

IGRA & TST for pre-IPT assessment in HIV+

Methods

1 logistic regression model

2 compare AUC



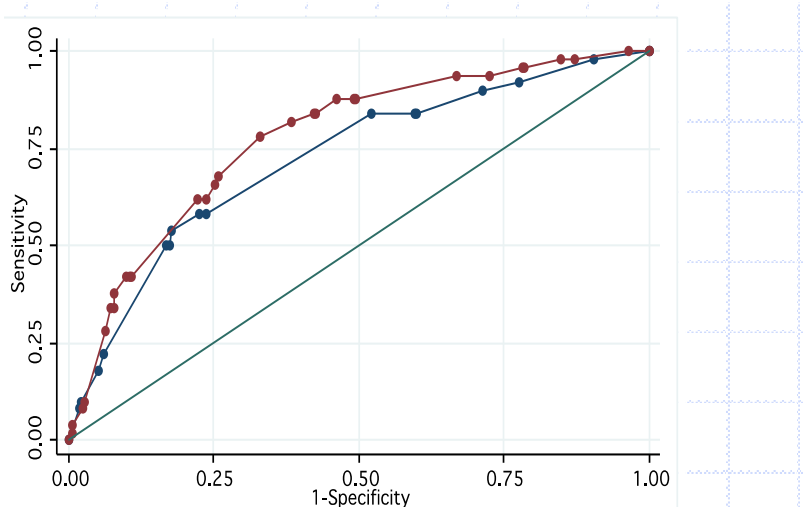
Clinical **(blue, AUC=72%)** AND TST at 5mm **(red, AUC=77%)**
Comparison p-value=0.03

IGRA & TST for pre-IPT assessment in HIV+

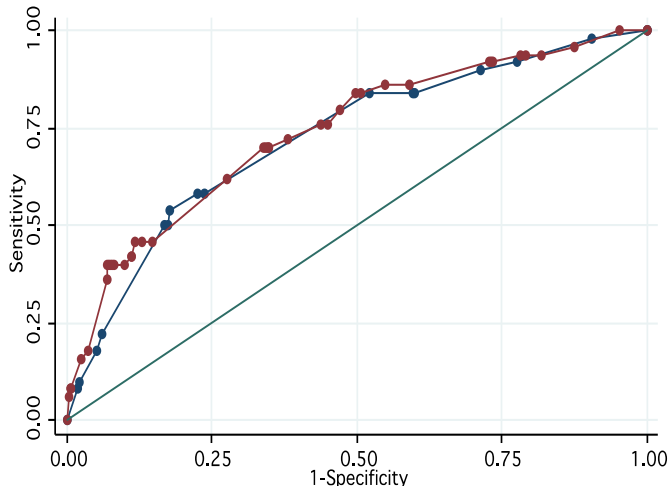
Methods

1 logistic regression model

2 compare AUC

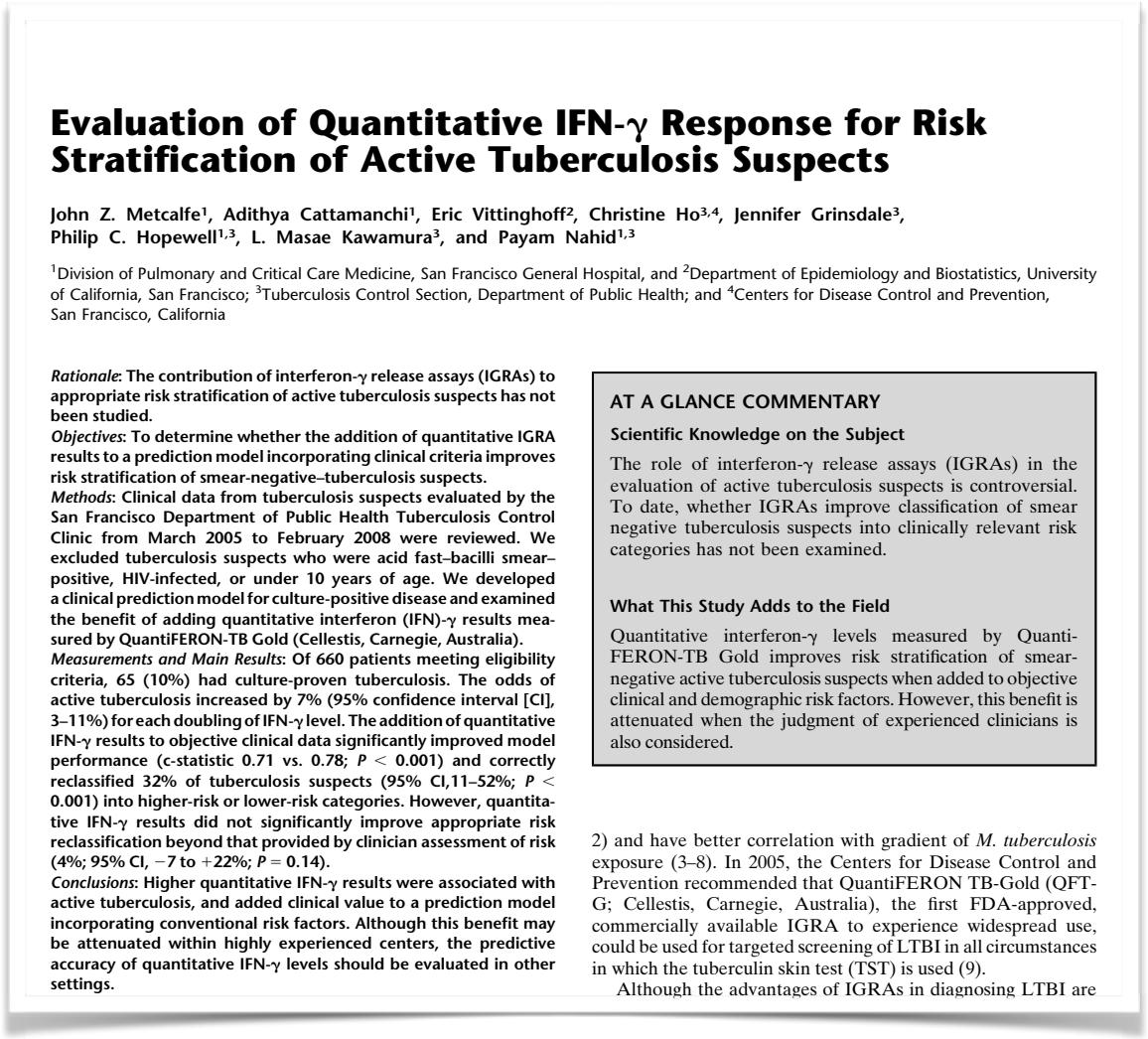


Clinical (**blue, AUC=72%**) AND TST at 5mm (**red, AUC=77%**)
Comparison p-value=0.03



Clinical (**blue, AUC=72%**) AND QFT (**red, AUC=74%**)
Comparison p-value=0.41

IGRA for Risk Stratification of Active Tuberculosis Suspects



Evaluation of Quantitative IFN-γ Response for Risk Stratification of Active Tuberculosis Suspects

John Z. Metcalfe[1], Adithya Cattamanchi[1], Eric Vittinghoff[2], Christine Ho[3,4], Jennifer Grinsdale[3], Philip C. Hopewell[1,3], L. Masae Kawamura[3], and Payam Nahid[1,3]

[1]Division of Pulmonary and Critical Care Medicine, San Francisco General Hospital, and [2]Department of Epidemiology and Biostatistics, University of California, San Francisco; [3]Tuberculosis Control Section, Department of Public Health; and [4]Centers for Disease Control and Prevention, San Francisco, California

Rationale: The contribution of interferon-γ release assays (IGRAs) to appropriate risk stratification of active tuberculosis suspects has not been studied.
Objectives: To determine whether the addition of quantitative IGRA results to a prediction model incorporating clinical criteria improves risk stratification of smear-negative–tuberculosis suspects.
Methods: Clinical data from tuberculosis suspects evaluated by the San Francisco Department of Public Health Tuberculosis Control Clinic from March 2005 to February 2008 were reviewed. We excluded tuberculosis suspects who were acid fast–bacilli smear–positive, HIV-infected, or under 10 years of age. We developed a clinical prediction model for culture-positive disease and examined the benefit of adding quantitative interferon (IFN)-γ results measured by QuantiFERON-TB Gold (Cellestis, Carnegie, Australia).
Measurements and Main Results: Of 660 patients meeting eligibility criteria, 65 (10%) had culture-proven tuberculosis. The odds of active tuberculosis increased by 7% (95% confidence interval [CI], 3–11%) for each doubling of IFN-γ level. The addition of quantitative IFN-γ results to objective clinical data significantly improved model performance (c-statistic 0.71 vs. 0.78; $P < 0.001$) and correctly reclassified 32% of tuberculosis suspects (95% CI,11–52%; $P < 0.001$) into higher-risk or lower-risk categories. However, quantitative IFN-γ results did not significantly improve appropriate risk reclassification beyond that provided by clinician assessment of risk (4%; 95% CI, −7 to +22%; $P = 0.14$).
Conclusions: Higher quantitative IFN-γ results were associated with active tuberculosis, and added clinical value to a prediction model incorporating conventional risk factors. Although this benefit may be attenuated within highly experienced centers, the predictive accuracy of quantitative IFN-γ levels should be evaluated in other settings.

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

The role of interferon-γ release assays (IGRAs) in the evaluation of active tuberculosis suspects is controversial. To date, whether IGRAs improve classification of smear negative tuberculosis suspects into clinically relevant risk categories has not been examined.

What This Study Adds to the Field

Quantitative interferon-γ levels measured by QuantiFERON-TB Gold improves risk stratification of smear-negative active tuberculosis suspects when added to objective clinical and demographic risk factors. However, this benefit is attenuated when the judgment of experienced clinicians is also considered.

2) and have better correlation with gradient of *M. tuberculosis* exposure (3–8). In 2005, the Centers for Disease Control and Prevention recommended that QuantiFERON TB-Gold (QFT-G; Cellestis, Carnegie, Australia), the first FDA-approved, commercially available IGRA to experience widespread use, could be used for targeted screening of LTBI in all circumstances in which the tuberculin skin test (TST) is used (9).

Although the advantages of IGRAs in diagnosing LTBI are

Method

- **ref. stand.**: culture
- **baseline:** logistic regression model (demographics, clinical, CXR)
- **Index test:** IGRA (quant.)

Results

- Δ AUC = +7%
- NRI = +32%
- no improvement beyond clinician assessment

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

2 assess reclassification

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

2 assess reclassification

TABLE 3. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN-γ RESULTS. COMPARISON TO BASELINE CLINICAL PREDICTION MODEL

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results				Percent Appropriately Reclassified
	≤5% risk	5–20% risk	>20% risk	Total No.	
In 65 patients who developed culture-positive disease					
≤5% risk	7	3	0	10	30
5–20% risk	1	27	12	40	28
>20% risk	1	0	14	15	7
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis					
≤5% risk	158	18	0	176	−10
5–20% risk	89	241	27	357	17
>20% risk	17	10	35	62	44
Total No.	264	269	62	595	

Net reclassification improvement = 31.9% ($P < 0.001$). Reclassification among patients who developed culture-positive disease = 20% ($P < 0.01$); reclassification among patients who ruled out for active tuberculosis = 11.9% ($P < 0.001$).

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

2 assess reclassification

TABLE 3. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN-γ RESULTS. COMPARISON TO BASELINE CLINICAL PREDICTION MODEL

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results				Percent Appropriately Reclassified
	≤5% risk	5–20% risk	>20% risk	Total No.	
In 65 patients who developed culture-positive disease					
≤5% risk	7	3	0	10	30
5–20% risk	1	27	12	40	28
>20% risk	1	0	14	15	7
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis					
≤5% risk	158	18	0	176	−10
5–20% risk	89	241	27	357	17
>20% risk	17	10	35	62	44
Total No.	264	269	62	595	

Net reclassification improvement = 31.9% ($P < 0.001$). Reclassification among patients who developed culture-positive disease = 20% ($P < 0.01$); reclassification among patients who ruled out for active tuberculosis = 11.9% ($P < 0.001$).

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

2 assess reclassification

TABLE 3. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN-γ RESULTS. COMPARISON TO BASELINE CLINICAL PREDICTION MODEL

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results			Total No.	Percent Appropriately Reclassified
	≤5% risk	5–20% risk	>20% risk		
In 65 patients who developed culture-positive disease					
≤5% risk	7	3	0	10	30
5–20% risk	1	27	12	40	28
>20% risk	1	0	14	15	7
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis					
≤5% risk	158	18	0	176	−10
5–20% risk	89	241	27	357	17
>20% risk	17	10	35	62	44
Total No.	264	269	62	595	

Net reclassification improvement = 31.9% ($P < 0.001$). Reclassification among patients who developed culture-positive disease = 20% ($P < 0.01$); reclassification among patients who ruled out for active tuberculosis = 11.9% ($P < 0.001$).

IGRA for Risk Stratification of Active Tuberculosis Suspects

Methods

1 logistic regression model

2 assess reclassification

TABLE 3. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN-γ RESULTS. COMPARISON TO BASELINE CLINICAL PREDICTION MODEL

Model with Clinical Predictors Alone	Model with Clinical Predictors and Quantitative IFN-γ Results			Total No.	Percent Appropriately Reclassified
	≤5% risk	5–20% risk	>20% risk		
In 65 patients who developed culture-positive disease					
≤5% risk	7	3	0	10	30
5–20% risk	1	27	12	40	28
>20% risk	1	0	14	15	7
Total No.	9	30	26	65	
In 595 patients who ruled out for active tuberculosis					
≤5% risk	158	18	0	176	−10
5–20% risk	89	241	27	357	17
>20% risk	17	10	35	62	44
Total No.	264	269	62	595	

Net reclassification improvement = 31.9% ($P < 0.001$). Reclassification among patients who developed culture-positive disease = 20% ($P < 0.01$); reclassification among patients who ruled out for active tuberculosis = 11.9% ($P < 0.001$).

Net Benefit & Decision Curve Analysis

Decision Thresholds

Net Benefit & Decision Curve Analysis

Decision Thresholds

- Taking clinical consequences of testing into account explicitly

Net Benefit & Decision Curve Analysis

Decision Thresholds

- Taking clinical consequences of testing into account explicitly
- Decision threshold (p_t): probability where expected benefit of treatment is equal to the expected benefit of avoiding treatment

Net Benefit & Decision Curve Analysis

Decision Thresholds

- Taking clinical consequences of testing into account explicitly
- Decision threshold (p_t): probability where expected benefit of treatment is equal to the expected benefit of avoiding treatment
  - e.g. p_t = 20% = willing to over-treat 80 to correctly treat 20; or under-treat 20 to correctly withhold treatment in 80

Net Benefit & Decision Curve Analysis

The Basic Idea

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

- key idea: by committing to a particular p_t, you implicitly state your view of what the trade-off should be between avoiding FPs and FNs

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

- key idea: by committing to a particular p_t, you implicitly state your view of what the trade-off should be between avoiding FPs and FNs
  - this can then be used to calculate the NB of a test/model at that threshold

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

- key idea: by committing to a particular p_t, you implicitly state your view of what the trade-off should be between avoiding FPs and FNs
  - this can then be used to calculate the NB of a test/model at that threshold

- NB indicates how many more TP classifications can be made with a model for the same number of FP classifications, compared to not using a model

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

- key idea: by committing to a particular p_t, you implicitly state your view of what the trade-off should be between avoiding FPs and FNs
  - this can then be used to calculate the NB of a test/model at that threshold

- NB indicates how many more TP classifications can be made with a model for the same number of FP classifications, compared to not using a model
  - e.g. NB of 0.05: use of test/model leads to 5 extra TPs per 100 tested without an increase in FPs, compared to assuming all are test-negative

Net Benefit & Decision Curve Analysis

The Basic Idea

$$\text{Net benefit} = \frac{\text{true-positive count}}{n} - \frac{\text{false-positive count}}{n}\left(\frac{p_t}{1-p_t}\right).$$

- key idea: by committing to a particular p_t, you implicitly state your view of what the trade-off should be between avoiding FPs and FNs
  - this can then be used to calculate the NB of a test/model at that threshold
- NB indicates how many more TP classifications can be made with a model for the same number of FP classifications, compared to not using a model
  - e.g. NB of 0.05: use of test/model leads to 5 extra TPs per 100 tested without an increase in FPs, compared to assuming all are test-negative
- DCA: NB over a range of decision thresholds

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions
  1. assume all will recur / all radical (perfect sens.)

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions
  1. assume all will recur / all radical (perfect sens.)
  2. assume none will recur / none radical (perfect spec.)

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions
  1. assume all will recur / all radical (perfect sens.)
  2. assume none will recur / none radical (perfect spec.)
  3. decision based on cancer grade (AUC 0.56)

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions
  1. assume all will recur / all radical (perfect sens.)
  2. assume none will recur / none radical (perfect spec.)
  3. decision based on cancer grade (AUC 0.56)
  4. decision based on cancer grade & stage (AUC 0.58)

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

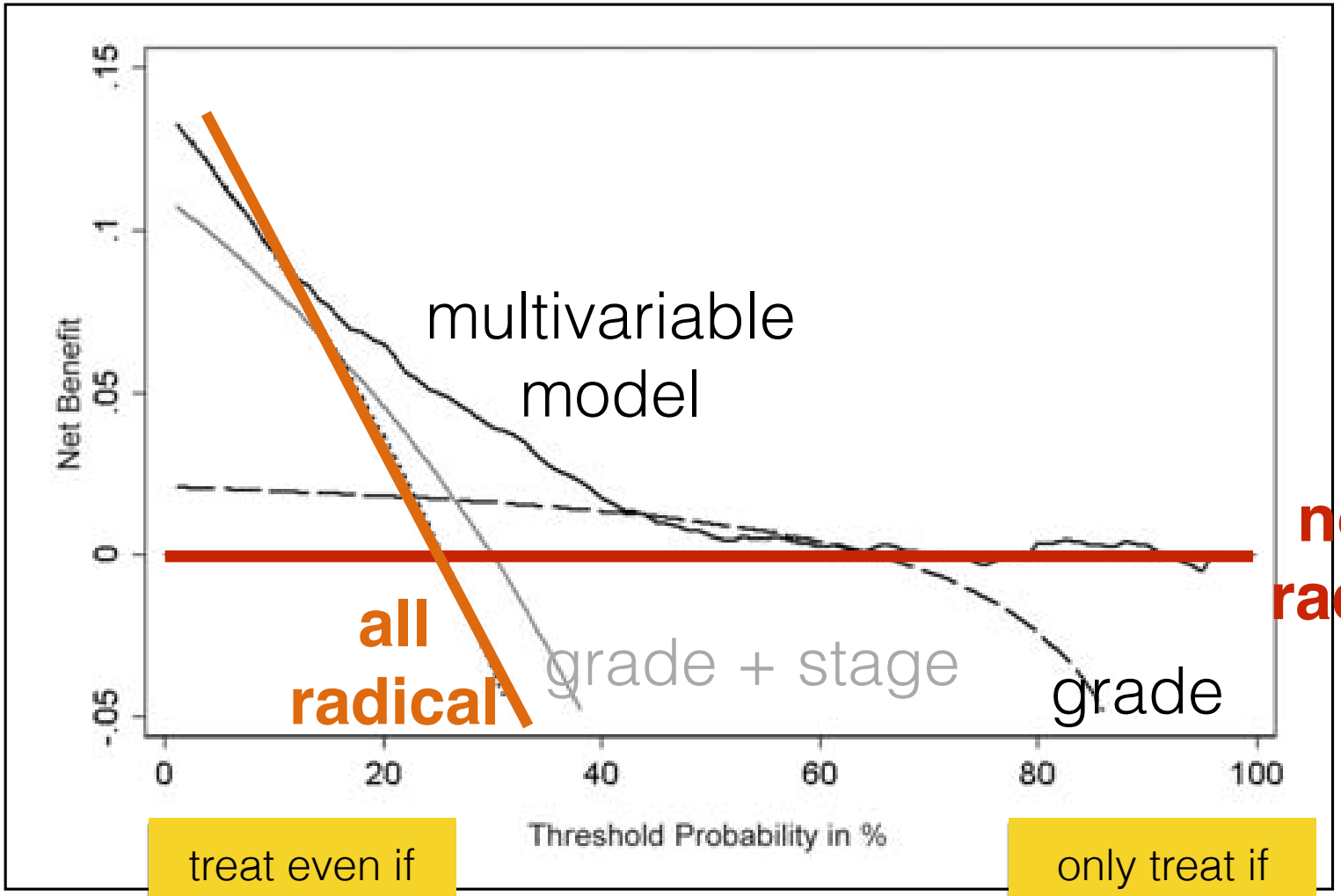
Example of Decision Curve Analysis

- clinical decision: more radical vs. less radical surgery?
  a) less radical: avoid impotence/incontinence, risk recurrence
  b) more radical: risk impotence/incontinence, avoid recurrence

- possible decisions
  1. assume all will recur / all radical (perfect sens.)
  2. assume none will recur / none radical (perfect spec.)
  3. decision based on cancer grade (AUC 0.56)
  4. decision based on cancer grade & stage (AUC 0.58)
  5. decision based on multivariable model (AUC 0.73)

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

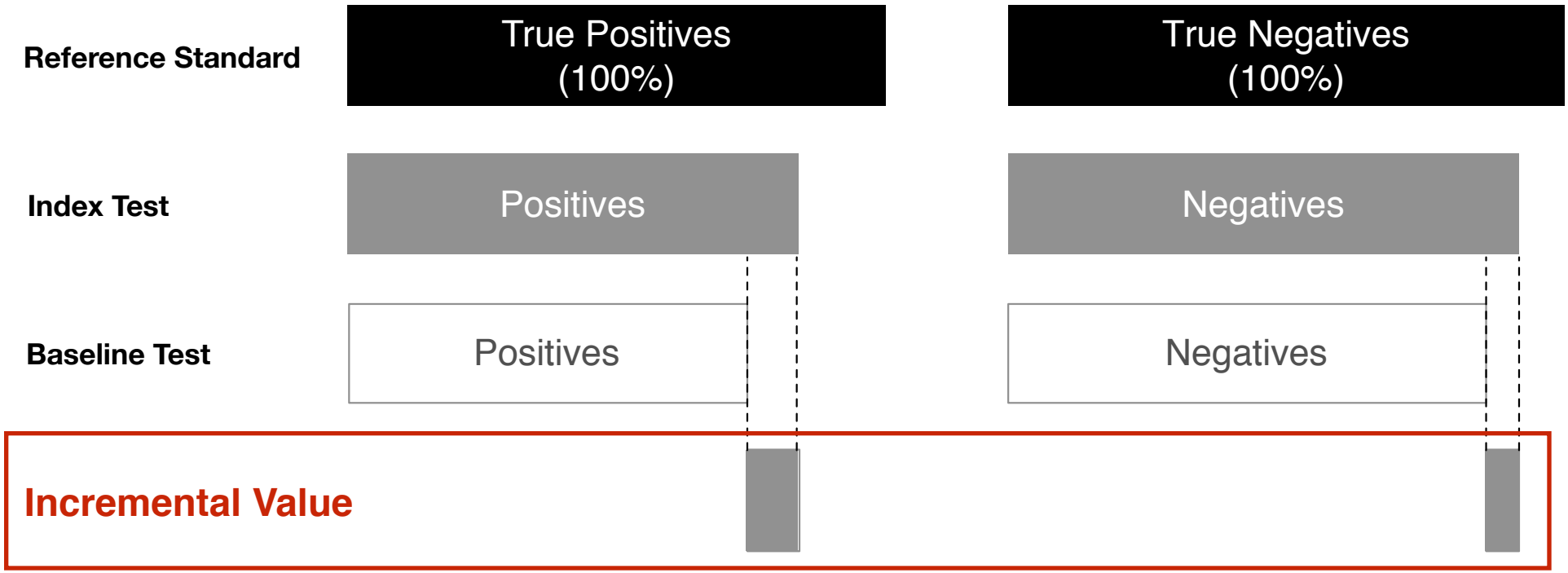


treat even if P(D) is low

only treat if P(D) very high

Imperfect Reference Standard

Incremental value: 3 tests



Xpert for Paediatric TB

Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study

Mark P Nicol, Lesley Workman, Washiefa Isaacs, Jacinta Munro, Faye Black, Brian Eley, Catharina C Boehme, Widaad Zemanay, Heather J Zar

Summary

Background WHO recommends that Xpert MTB/RIF replaces smear microscopy for initial diagnosis of suspected HIV-associated tuberculosis or multidrug-resistant pulmonary tuberculosis, but no data exist for its use in children. We aimed to assess the accuracy of the test for the diagnosis of pulmonary tuberculosis in children in an area with high tuberculosis and HIV prevalences.

Methods In this prospective, descriptive study, we enrolled children aged 15 years or younger who had been admitted to one of two hospitals in Cape Town, South Africa, with suspected pulmonary tuberculosis between Feb 19, 2009, and Nov 30, 2010. We compared the diagnostic accuracy of MTB/RIF and concentrated, fluorescent acid-fast smear with a reference standard of liquid culture from two sequential induced sputum specimens (primary analysis).

Results 452 children (median age 19·4 months, IQR 11·1–46·2) had at least one induced sputum specimen; 108 children (24%) had HIV infection. 27 children (6%) had a positive smear result, 70 (16%) had a positive culture result, and 58 (13%) had a positive MTB/RIF test result. With mycobacterial culture as the reference standard, MTB/RIF tests when done on two induced sputum samples detected twice as many cases (75·9%, 95% CI 64·5–87·2) as did smear microscopy (37·9%, 25·1–50·8), detecting all of 22 smear-positive cases and 22 of 36 (61·1%, 44·4–77·8) smear-negative cases. For smear-negative cases, the incremental increase in sensitivity from testing a second specimen was 27·8% for MTB/RIF, compared with 13·8% for culture. The specificity of MTB/RIF was 98·8% (97·6–99·9). MTB/RIF results were available in median 1 day (IQR 0–4) compared with median 12 days (9–17) for culture (p<0·0001).

Interpretation MTB/RIF testing of two induced sputum specimens is warranted as the first-line diagnostic test for children with suspected pulmonary tuberculosis.

Funding National Institutes of Health, the National Health Laboratory Service Research Trust, the Medical Research Council of South Africa, and Wellcome Trust.

Published Online July 18, 2011 DOI:10.1016/S1473-3099(11)70167-0

See Online/Comment DOI:10.1016/S1473-3099(11)70187-6

Division of Medical Microbiology and Institute for Infectious Diseases and Molecular Medicine (Prof M P Nicol PhD, W Zemanay PhD), **Division of Clinical Pharmacology, Department of Medicine** (L Workman MSc), **and Department of Paediatrics and Child Health** (W Isaacs BSc, J Munro, F Black MBBCh, B Eley FCPaed, Prof H J Zar PhD), **University of Cape Town, Cape Town, South Africa; National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa** (Prof M P Nicol, W Zemanay); **Red Cross War Memorial Children's Hospital, Cape Town, South Africa** (W Isaacs, J Munro, F Black, B Eley, Prof H J Zar); and

Rapid diagnosis of pulmonary tuberculosis in African children in a primary care setting by use of Xpert MTB/RIF on respiratory specimens: a prospective study

Heather J Zar, Lesley Workman, Washiefa Isaacs, Keertan Dheda, Widaad Zemanay, Mark P Nicol

Summary

Background In children admitted to hospital, rapid, accurate diagnosis of pulmonary tuberculosis with the Xpert MTB/RIF assay is possible, but no paediatric studies have been done in the primary care setting, where most children are given care, and where microbiological diagnosis is rarely available. We assessed the diagnostic accuracy of Xpert MTB/RIF in children in primary care.

Methods For this prospective study, we obtained repeat induced sputum and nasopharyngeal aspirate specimens from children (<15 years) with suspected pulmonary tuberculosis at a clinic in Khayelitsha, Cape Town, South Africa. We compared the diagnostic accuracy of Xpert MTB/RIF with a reference standard of culture and smear microscopy on induced sputum specimens. For the main analysis, specificity of Xpert MTB/RIF versus liquid culture, we included only children with two interpretable Xpert MTB/RIF and induced sputum culture results.

Findings Between Aug 1, 2010, and July 30, 2012, we enrolled 384 children (median age 38·3 months, IQR 21·2–56·5) who had one paired induced sputum and nasopharyngeal specimen, 309 (81%) of whom had two paired specimens. Five children (1%) tested positive for tuberculosis by smear microscopy, 26 (7%) tested positive by Xpert MTB/RIF, and 30 (8%) tested positive by culture. Xpert MTB/RIF on two induced sputum specimens detected 16 of 28 culture-confirmed cases (sensitivity of 57·1%, 95% CI 39·1–73·5) and on two nasopharyngeal aspirates detected 11 of 28 culture-confirmed cases (sensitivity of 39·3, 23·6–57·6; p=0·18). The specificity of Xpert MTB/RIF on induced sputum was 98·9% (95% CI 96·9–99·6) and on nasopharyngeal aspirates was 99·3% (97·4–99·8).

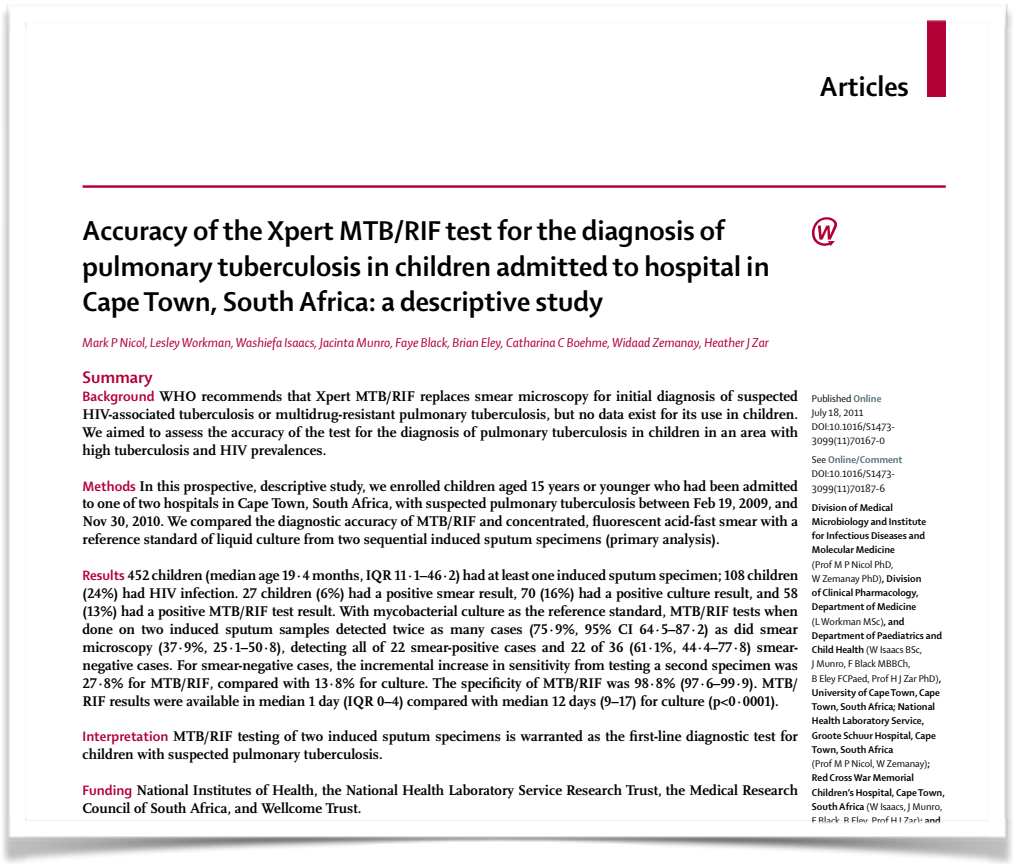
Interpretation Our findings suggest that Xpert MTB/RIF on respiratory secretions is a useful test for rapid diagnosis of paediatric pulmonary tuberculosis in primary care.

Lancet Glob Health 2013; 1: e97–104

Department of Paediatrics and Child Health (Prof H J Zar PhD, L Workman MPH, W Isaacs MSc), **Department of Medicine** (Prof K Dheda PhD), **and Division of Medical Microbiology, Institute for Infectious Diseases, and Molecular Medicine** (W Zemanay PhD, Prof M P Nicol PhD), **University of Cape Town, Cape Town, South Africa; Red Cross War Memorial Children's Hospital, Cape Town, South Africa** Prof H J Zar, L Workman, W Isaacs); **and the National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa** (W Zemanay, Prof M P Nicol)

Correspondence to: Prof Heather J Zar, Department

Xpert for Paediatric TB



Articles

Accuracy of the Xpert MTB/RIF test for the diagnosis of pulmonary tuberculosis in children admitted to hospital in Cape Town, South Africa: a descriptive study

Mark P Nicol, Lesley Workman, Washiefa Isaacs, Jacinta Munro, Faye Black, Brian Eley, Catharina C Boehme, Widaad Zemanay, Heather J Zar

Summary

Background WHO recommends that Xpert MTB/RIF replaces smear microscopy for initial diagnosis of suspected HIV-associated tuberculosis or multidrug-resistant pulmonary tuberculosis, but no data exist for its use in children. We aimed to assess the accuracy of the test for the diagnosis of pulmonary tuberculosis in children in an area with high tuberculosis and HIV prevalences.

Methods In this prospective, descriptive study, we enrolled children aged 15 years or younger who had been admitted to one of two hospitals in Cape Town, South Africa, with suspected pulmonary tuberculosis between Feb 19, 2009, and Nov 30, 2010. We compared the diagnostic accuracy of MTB/RIF and concentrated, fluorescent acid-fast smear with a reference standard of liquid culture from two sequential induced sputum specimens (primary analysis).

Results 452 children (median age 19·4 months, IQR 11·1–46·2) had at least one induced sputum specimen; 108 children (24%) had HIV infection. 27 children (6%) had a positive smear result, 70 (16%) had a positive culture result, and 58 (13%) had a positive MTB/RIF test result. With mycobacterial culture as the reference standard, MTB/RIF tests when done on two induced sputum samples detected twice as many cases (75·9%, 95% CI 64·5–87·2) as did smear microscopy (37·9%, 25·1–50·8), detecting all of 22 smear-positive cases and 22 of 36 (61·1%, 44·4–77·8) smear-negative cases. For smear-negative cases, the incremental increase in sensitivity from testing a second specimen was 27·8% for MTB/RIF, compared with 13·8% for culture. The specificity of MTB/RIF was 98·8% (97·6–99·9). MTB/RIF results were available in median 1 day (IQR 0–4) compared with median 12 days (9–17) for culture (p<0·0001).

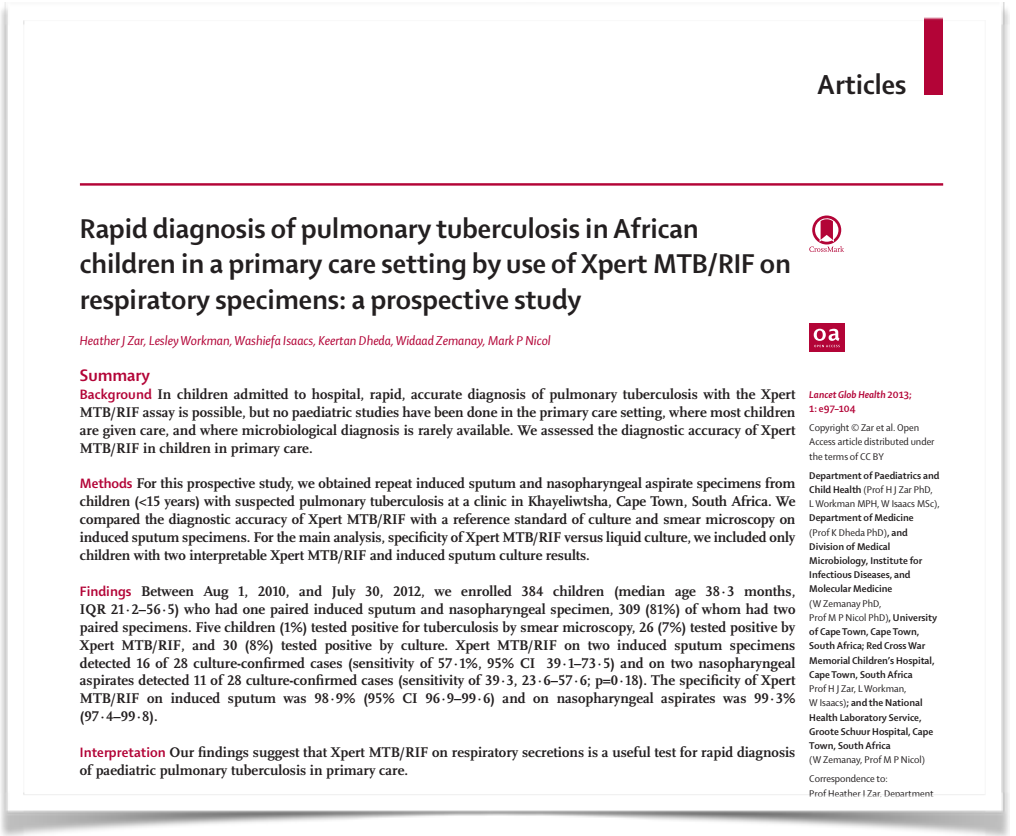
Interpretation MTB/RIF testing of two induced sputum specimens is warranted as the first-line diagnostic test for children with suspected pulmonary tuberculosis.

Funding National Institutes of Health, the National Health Laboratory Service Research Trust, the Medical Research Council of South Africa, and Wellcome Trust.

Published Online July 18, 2011 DOI:10.1016/S1473-3099(11)70167-0

See Online/Comment DOI:10.1016/S1473-3099(11)70187-6

Division of Medical Microbiology and Institute for Infectious Diseases and Molecular Medicine (Prof M P Nicol PhD, W Zemanay PhD), **Division of Clinical Pharmacology, Department of Medicine** (L Workman MSc), **and Department of Paediatrics and Child Health** (W Isaacs BSc, J Munro, F Black MBBCh, B Eley FCPaed, Prof H J Zar PhD), **University of Cape Town, Cape Town, South Africa; National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa** (Prof M P Nicol, W Zemanay); **Red Cross War Memorial Children's Hospital, Cape Town, South Africa** (W Isaacs, J Munro, F Black, B Eley, Prof H J Zar); and



Articles

Rapid diagnosis of pulmonary tuberculosis in African children in a primary care setting by use of Xpert MTB/RIF on respiratory specimens: a prospective study

Heather J Zar, Lesley Workman, Washiefa Isaacs, Keertan Dheda, Widaad Zemanay, Mark P Nicol

Summary

Background In children admitted to hospital, rapid, accurate diagnosis of pulmonary tuberculosis with the Xpert MTB/RIF assay is possible, but no paediatric studies have been done in the primary care setting, where most children are given care, and where microbiological diagnosis is rarely available. We assessed the diagnostic accuracy of Xpert MTB/RIF in children in primary care.

Methods For this prospective study, we obtained repeat induced sputum and nasopharyngeal aspirate specimens from children (<15 years) with suspected pulmonary tuberculosis at a clinic in Khayeliwtsha, Cape Town, South Africa. We compared the diagnostic accuracy of Xpert MTB/RIF with a reference standard of culture and smear microscopy on induced sputum specimens. For the main analysis, specificity of Xpert MTB/RIF versus liquid culture, we included only children with two interpretable Xpert MTB/RIF and induced sputum culture results.

Findings Between Aug 1, 2010, and July 30, 2012, we enrolled 384 children (median age 38·3 months, IQR 21·2–56·5) who had one paired induced sputum and nasopharyngeal specimen, 309 (81%) of whom had two paired specimens. Five children (1%) tested positive for tuberculosis by smear microscopy, 26 (7%) tested positive by Xpert MTB/RIF, and 30 (8%) tested positive by culture. Xpert MTB/RIF on two induced sputum specimens detected 16 of 28 culture-confirmed cases (sensitivity of 57·1%, 95% CI 39·1–73·5) and on two nasopharyngeal aspirates detected 11 of 28 culture-confirmed cases (sensitivity of 39·3, 23·6–57·6; p=0·18). The specificity of Xpert MTB/RIF on induced sputum was 98·9% (95% CI 96·9–99·6) and on nasopharyngeal aspirates was 99·3% (97·4–99·8).

Interpretation Our findings suggest that Xpert MTB/RIF on respiratory secretions is a useful test for rapid diagnosis of paediatric pulmonary tuberculosis in primary care.

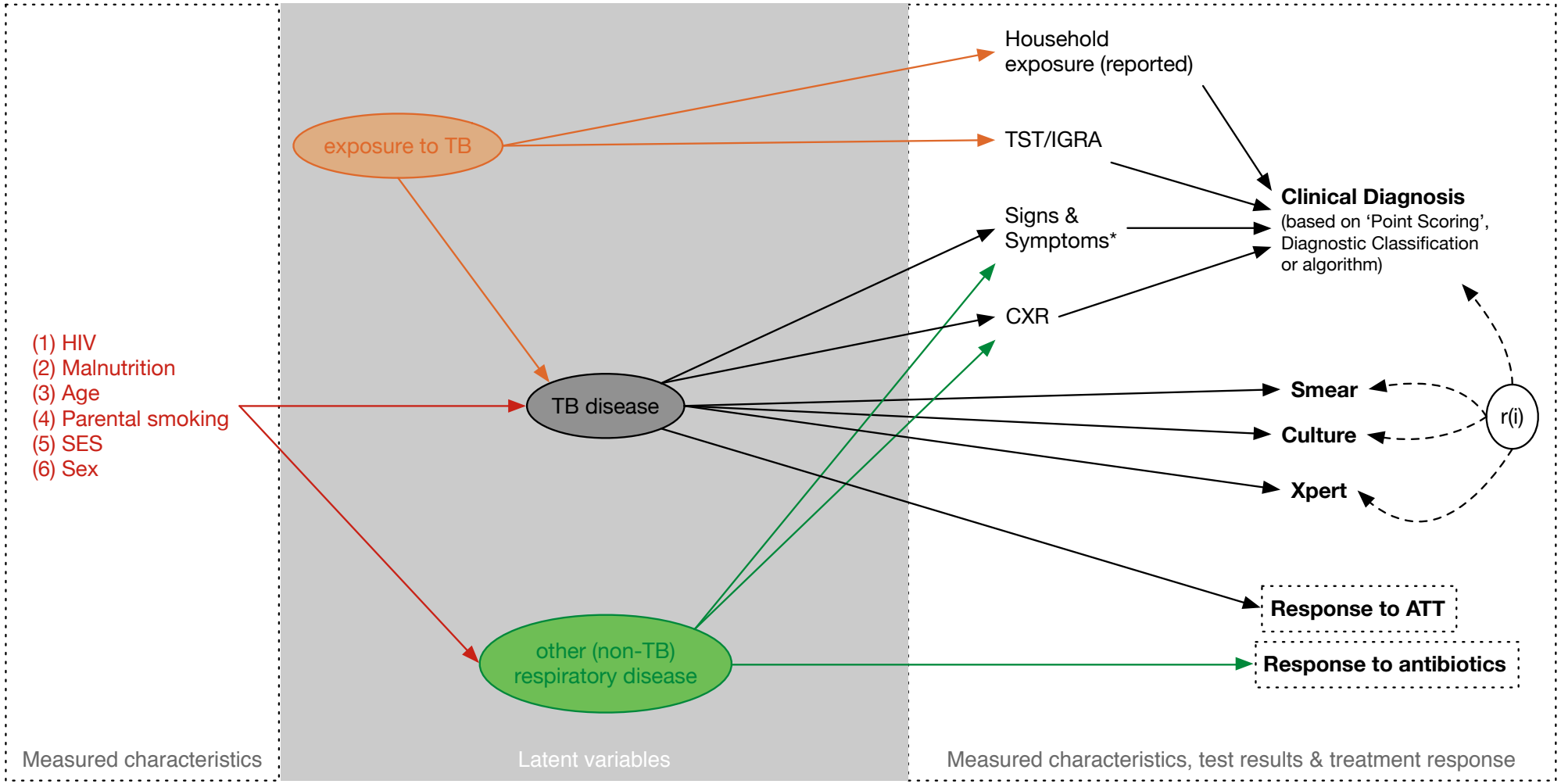
Lancet Glob Health 2013; 1: e97–104

Department of Paediatrics and Child Health (Prof H J Zar PhD, L Workman MPH, W Isaacs MSc), **Department of Medicine** (Prof K Dheda PhD), **and Division of Medical Microbiology, Institute for Infectious Diseases, and Molecular Medicine** (W Zemanay PhD, Prof M P Nicol PhD), **University of Cape Town, Cape Town, South Africa; Red Cross War Memorial Children's Hospital, Cape Town, South Africa** Prof H J Zar, L Workman, W Isaacs); **and the National Health Laboratory Service, Groote Schuur Hospital, Cape Town, South Africa** (W Zemanay, Prof M P Nicol)

Correspondence to: Prof Heather J Zar, Department

~10-20% culture+ but ~50% treated for TB

Latent Class Analysis



Other options

- improve upon simple reference standard
  - composite reference standard
  - panel diagnosis
- look downstream of “diagnosis”
  - look at treatment decisions
  - look at final outcomes
- decision-analytic modelling
  - based on evidence/assumptions about accuracy of index test and baseline test

Conclusions

Conclusions

- incremental value: first step towards impact
- “binary test case” is simple: should always be done
- “multivariable test case” based on diagnostic prediction models: consider DCA/NB
- imperfect reference standard

ADDITIONAL SLIDES

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis

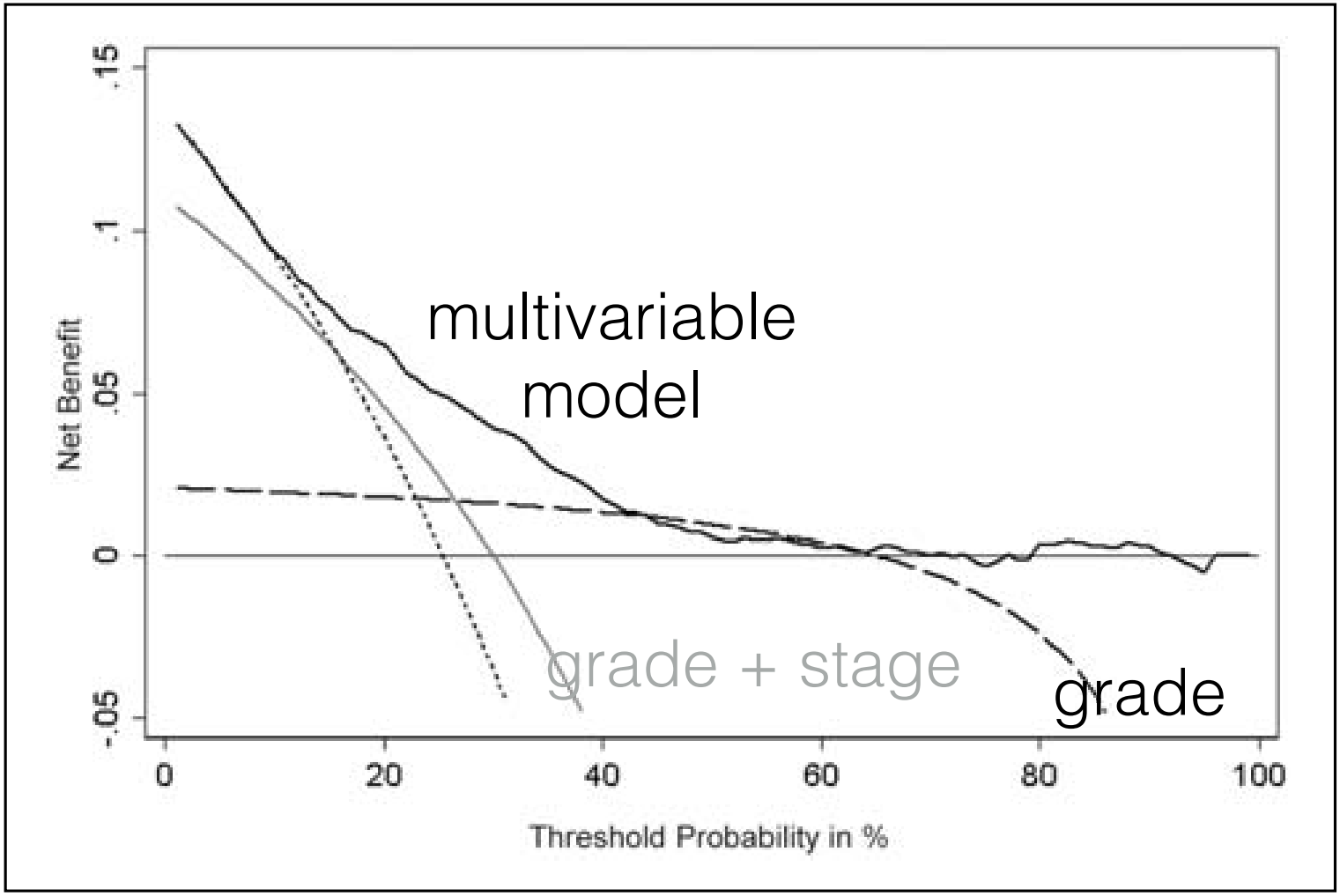
multivariable model

grade + stage

grade

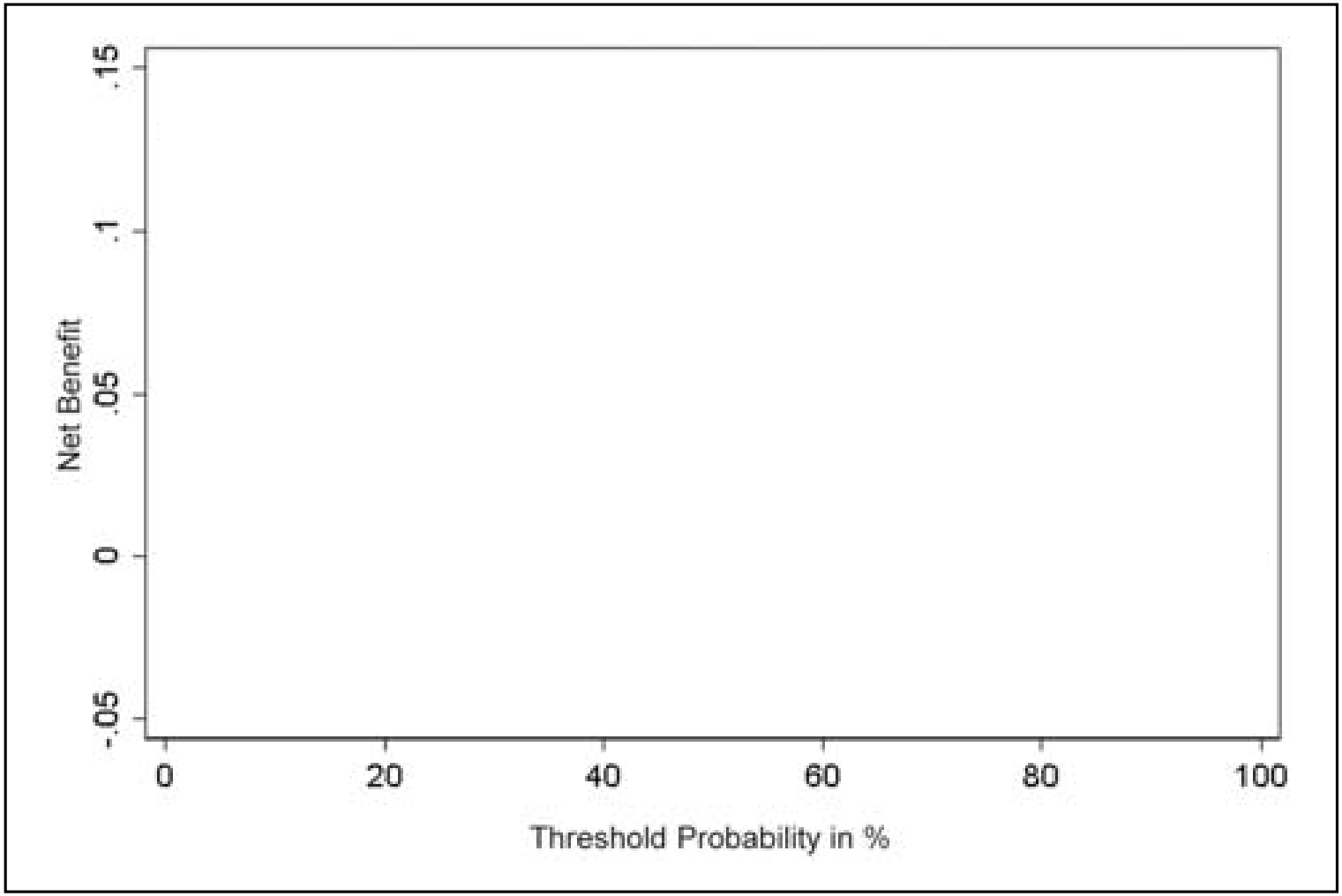
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



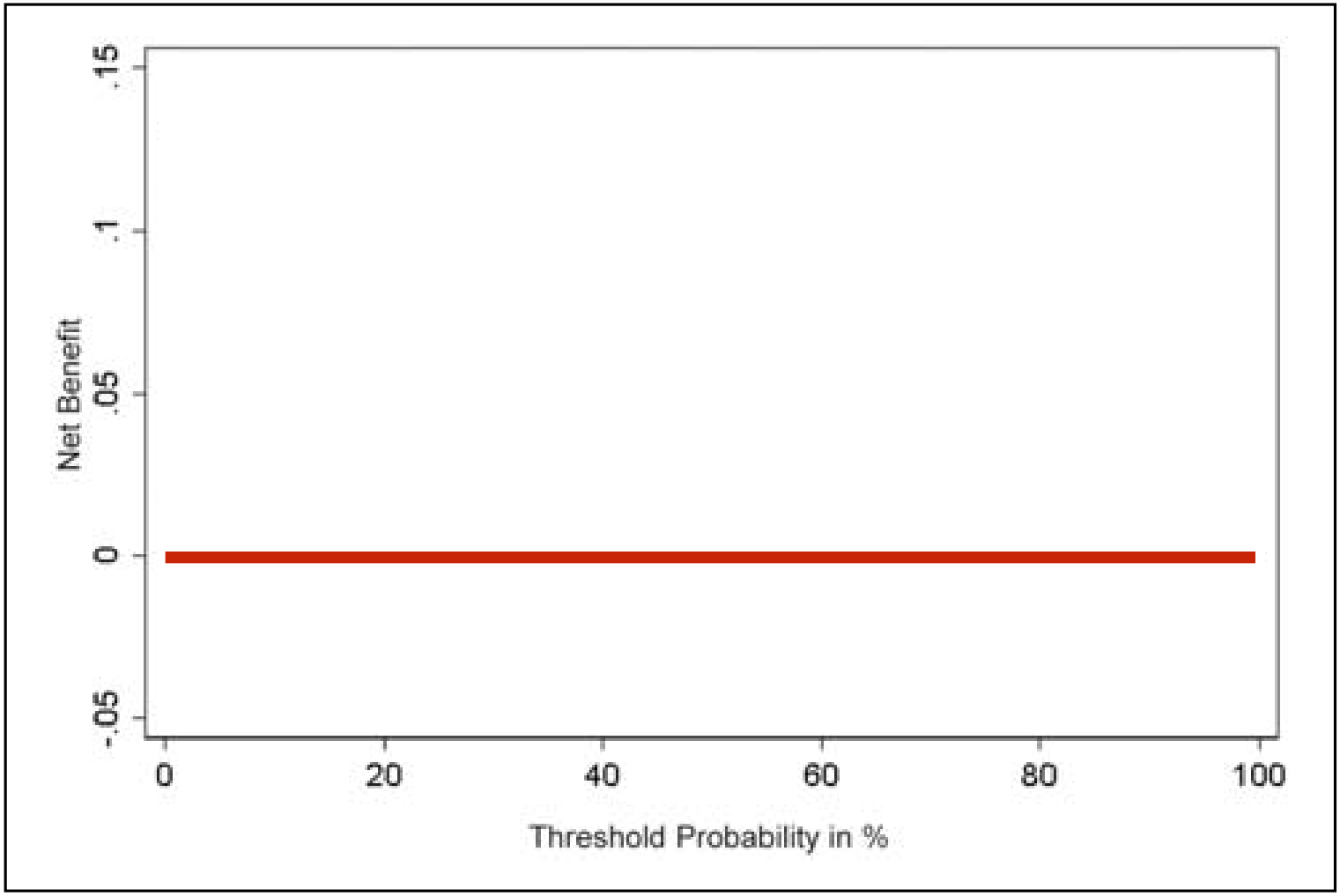
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



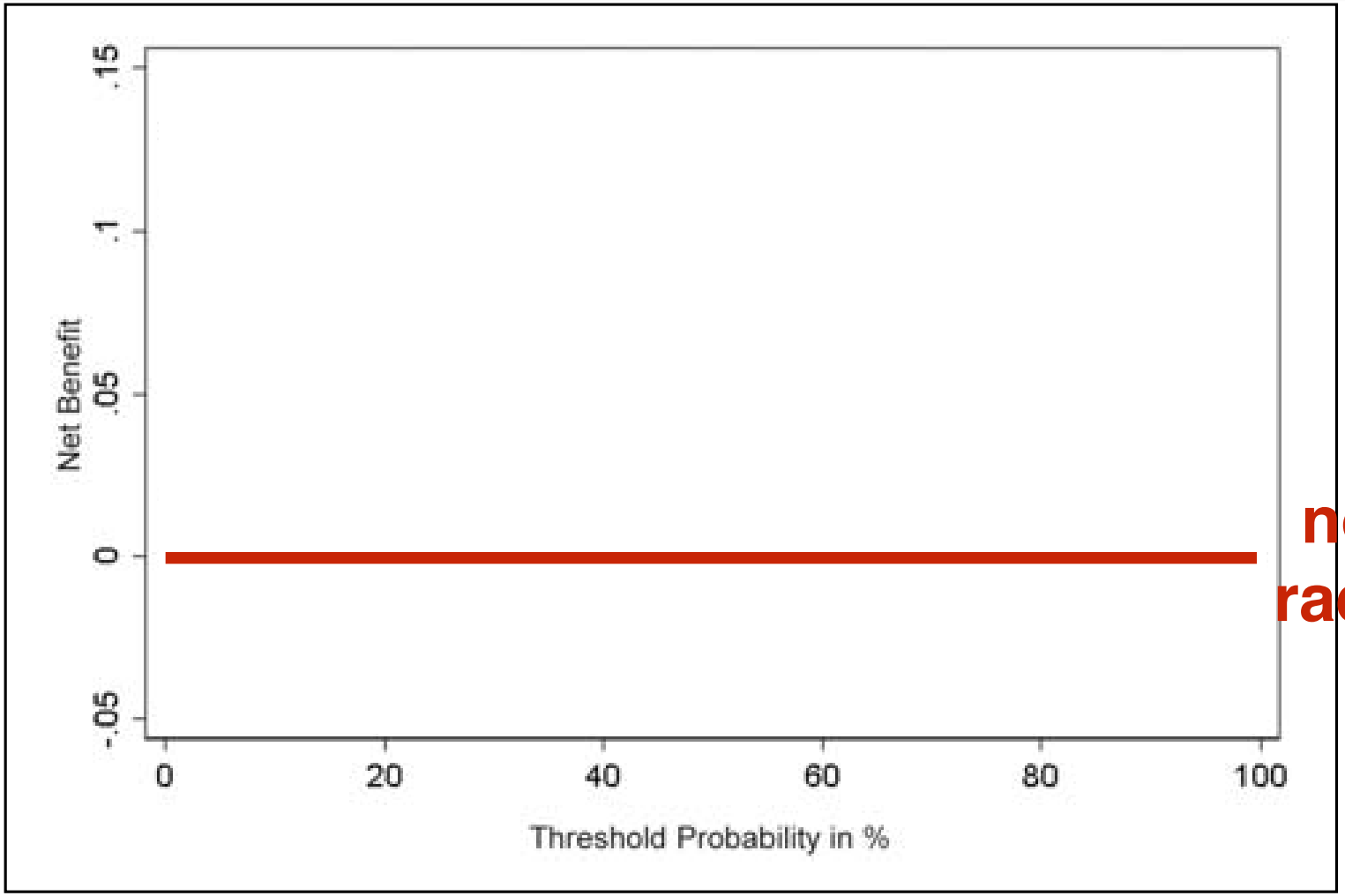
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



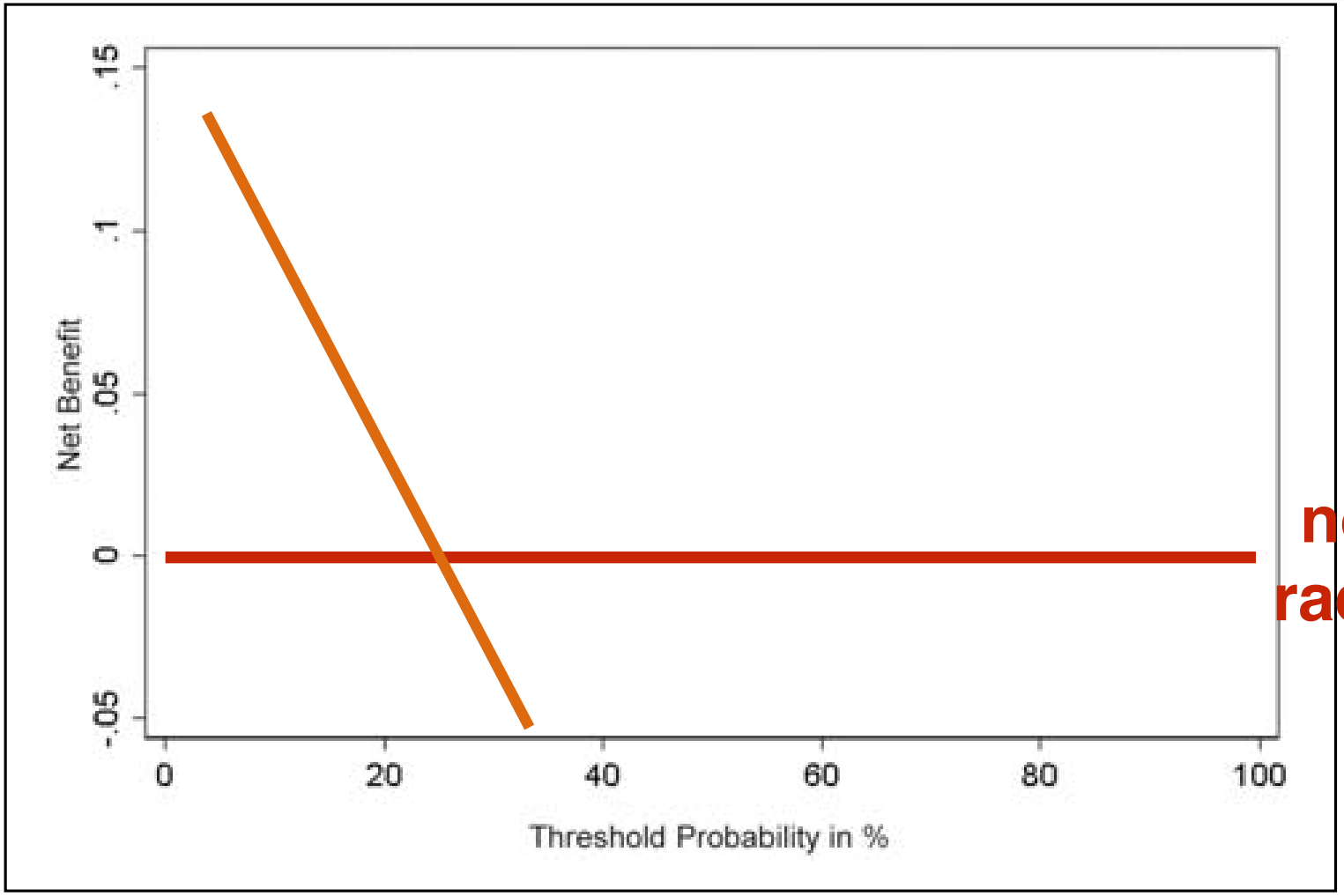
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



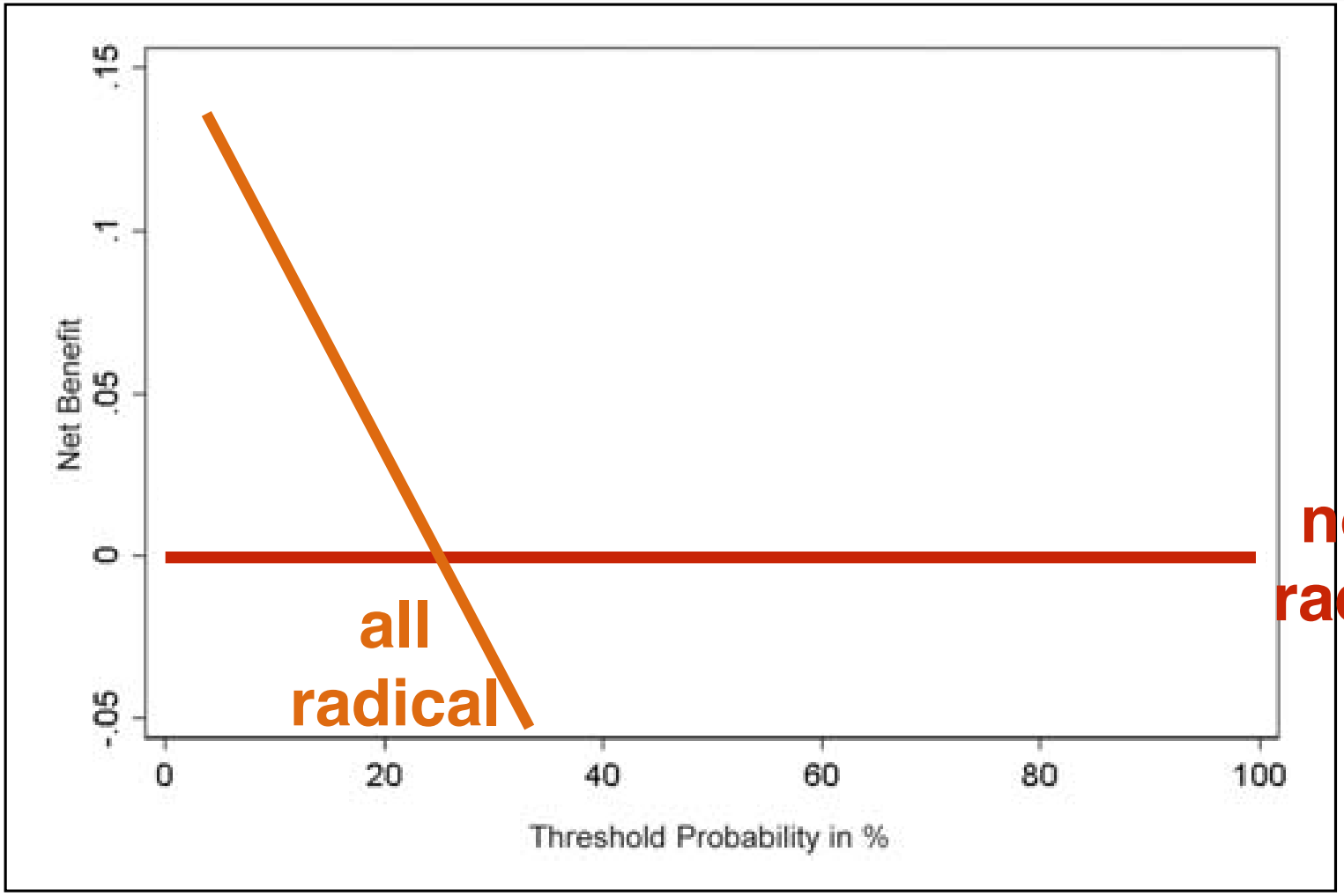
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



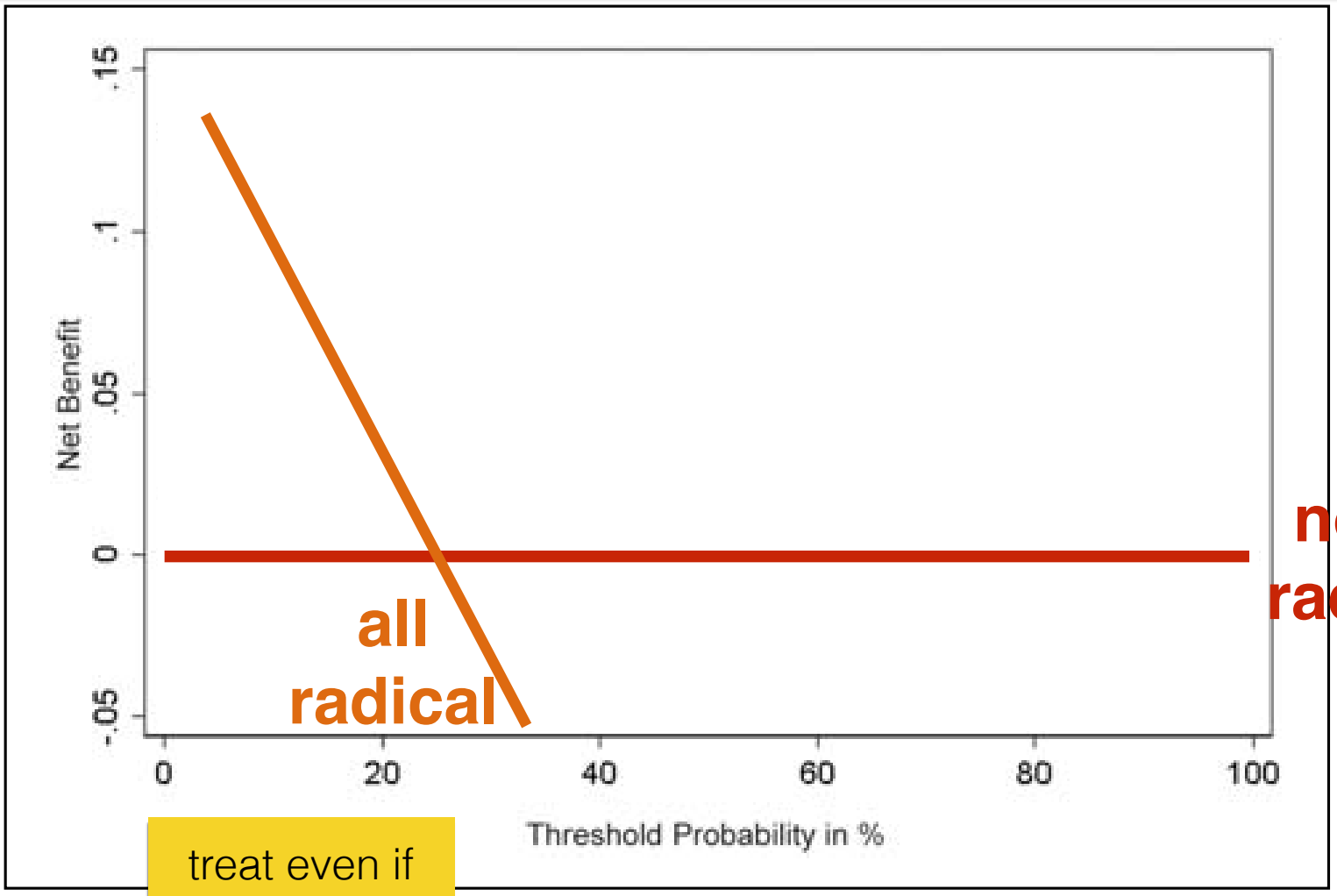
Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

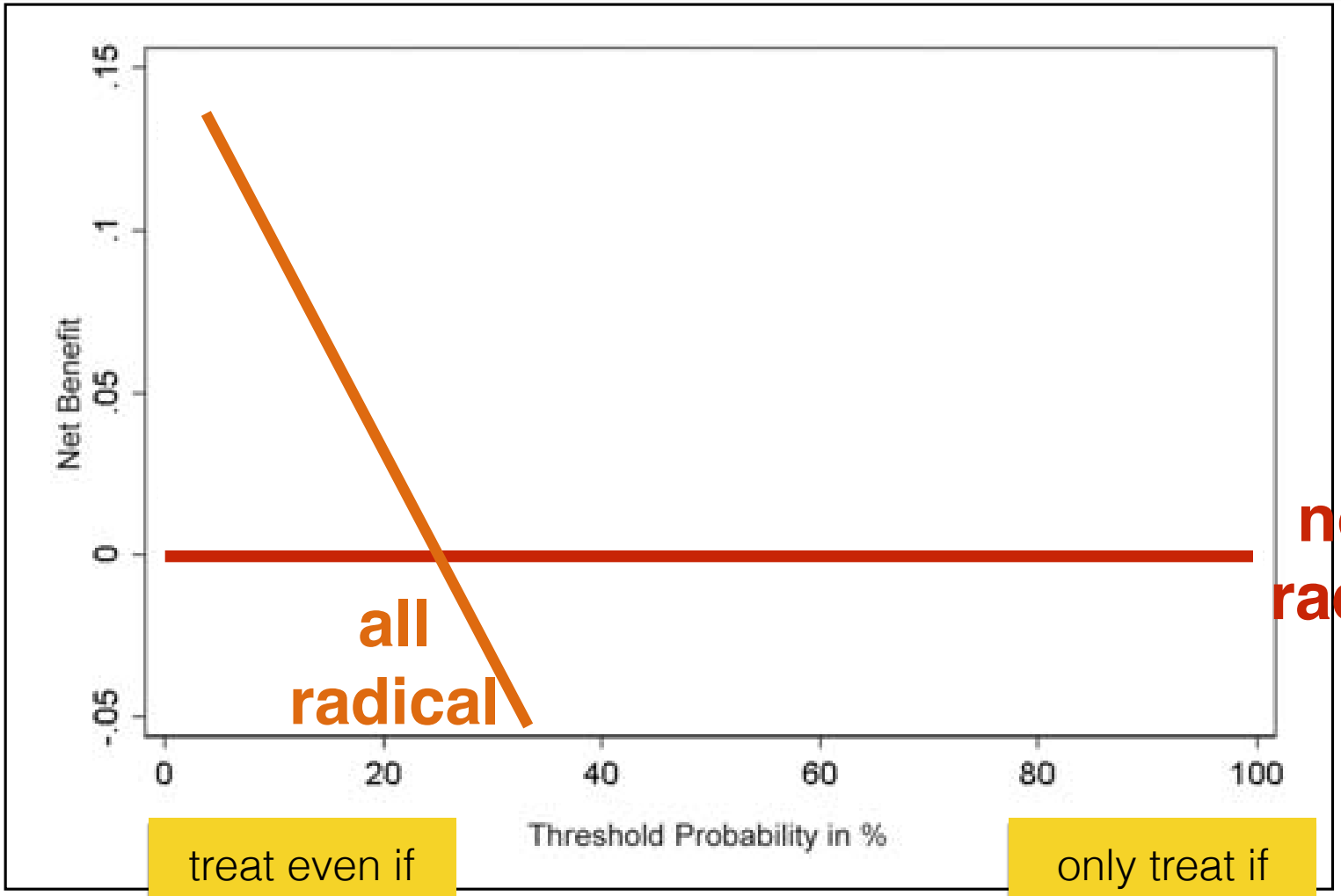
Example of Decision Curve Analysis



treat even if P(D) is low

Prediction of Seminal Vesicle Invasion at Radical Prostatectomy

Example of Decision Curve Analysis



Two paradigms for IV

	ID	NCD
Goal of testing	diagnosis	(diagnosis) & prognosis
Outcome	high and acute risk of mortality if untreated	risk of poor outcome usually not acute and sometimes not high
Treatment	very good risk/benefit profile	moderate risk/benefit profile
Reference standard	usually some kind of culture (or serology / molecular)	usually long-term follow-up or invasive testing
Management decision	based on single test	based on many parameters

Where the classic ID paradigm does not fit

	Infectious diseaes	Latent TB
Goal of testing	diagnosis	prognosis
Outcome	acute and high risk of mortality if untreated	risk of poor outcome usually not acute and sometimes not high
Treatment	very good risk/benefit profile	moderate risk/benefit profile
Reference standard	usually some kind of culture (or serology / molecular)	long-term follow-up
Management decision	based on single test	based on many parameters

PREDICTION RESEARCH

Where the classic ID paradigm does not fit

	Infectious diseaes	Childhood TB
Goal of testing	diagnosis	diagnosis/prognosis
Outcome	acute and high risk of mortality if untreated	acute and high risk of mortality if untreated
Treatment	very good risk/benefit profile	very good risk/benefit profile
Reference standard	usually some kind of culture (or serology / molecular)	?
Management decision	based on single test	based on many parameters

PREDICTION RESEARCH

Where the classic ID paradigm does not fit

	Infectious diseaes	Extrapulmonary TB
Goal of testing	diagnosis	diagnosis/prognosis
Outcome	acute and high risk of mortality if untreated	acute and high risk of mortality if untreated
Treatment	very good risk/benefit profile	very good risk/benefit profile
Reference standard	usually some kind of culture (or serology / molecular)	?
Management decision	based on single test	based on many parameters

PREDICTION RESEARCH

Where the classic ID paradigm does not fit

	Infectious diseaes	HIV-associated TB
Goal of testing	diagnosis	diagnosis/prognosis
Outcome	acute and high risk of mortality if untreated	acute and high risk of mortality if untreated
Treatment	very good risk/benefit profile	very good risk/benefit profile
Reference standard	usually some kind of culture (or serology / molecular)	?
Management decision	based on single test	based on many parameters

PREDICTION RESEARCH

Prediction Models & Incremental Value Metrics

1. Model development
2. Model assessment
   - discrimination (e.g. AUC)
   - calibration (e.g. calibration plot)
3. Internal validation
4. Assess incremental value
   - Δ AUC & IDI
   - reclassification table & NRI
   - DCA & NB
5. Next steps
   - external validation
   - model updating
   - impact assessment

How to integrate multiple parameters into treatment decision

- ideally: personalized management by highly-trained specialist based on any available data on patient, perhaps supported by a validated prediction model
- reality: based on algorithms (essentially a decision-rule)
  - based on expert opinion rather than empiric evidence?
  - either simple algorithms or simple scoring rule
- Opportunities
  - prediction models may offer improvement in correct classification
  - may enable lower-level HC workers to diagnose and treat
- Challenges
  - for diseases where multiple measures are used there is usually no good reference standard but long-term f/u is not practical in low-resource settings
  - potentially challenging to implement

Where the classic ID paradigm does not fit

- Latent TB infection
  - no detection of agent: prognosis more relevant than "diagnosis"
  - no gold standard test available
- Childhood TB
  - no detection of agent for majority: predicting prognosis rather than making diagnosis is relevant
  - no gold standard test available
- HIV-associated TB

Table 16.6 Developing and evaluating clinical prediction models and decision rules (based on Reilly and Evans[344])

Level of evidence	Definitions and standards of evaluation	Clinical implications
Level 1		
Derivation of prediction model	Identification of predictors for multivariable model; blinded assessment of outcomes	Needs validation and further evaluation before using in actual patient care
Level 2		
Narrow validation of prediction model	Assessment of predictive ability when tested prospectively in one setting; blinded assessment of outcomes	Needs validation in varied settings; may use predictions cautiously in patients similar to sample studied
Level 3		
Broad validation of prediction model	Assessment of predictive ability in varied settings with wide spectrum of patients and physicians	Needs impact analysis; may use predictions with confidence in their accuracy
Level 4		
Narrow impact analysis of prediction model used as decision rule	Prospective demonstration in one setting that use of decision rule improves physicians' decisions (quality or cost-effectiveness of patient care)	May use cautiously to inform decisions in settings similar to that studied
Level 5		
Broad impact analysis of prediction model used as decision rule	Prospective demonstration in varied settings that use of decision rule improves physicians' decisions for wide spectrum of patients	May use in varied settings with confidence that its use will benefit patient care quality or effectiveness

Incremental Value Metrics

AUC & IDI

- AUC: probability of correct classification for a pair of patients with and without the outcome
- IDI: sum of the average increase in predicted probability among D+ and average decrease in predicted probability among D- (related to ΔAUC); equivalent to the increase in mean sensitivity given no changes in specificity

$$\text{IDI} = [(P_{\text{extended}} \mid D = 1) - (P_{\text{basic}} \mid D = 1)] - [(P_{\text{extended}} \mid D = 0) - (P_{\text{basic}} \mid D = 0)].$$

NRI

Net fraction of reclassifications in the right direction by making decisions based on predictions with the marker compared to decisions without the marker; default weights are by prevalence of disease.

$$\begin{aligned}\text{NRI} = &[\text{P}(\text{up} \mid \text{D} = 1) - \text{P}(\text{down} \mid \text{D} = 1)] \\ &- [\text{P}(\text{up} \mid \text{D} = 0) - \text{P}(\text{down} \mid \text{D} = 0)],\end{aligned}$$

Decision Curve Analysis