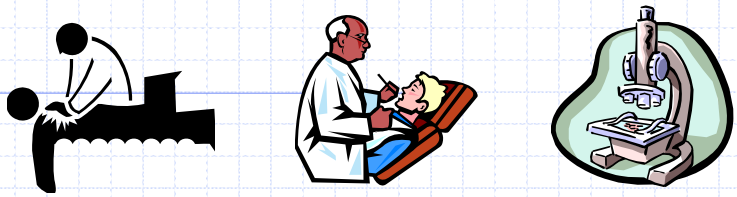
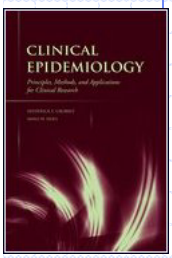
Diagnostic research: incremental value and multivariable approach



Madhukar Pai, MD, PhD
Assistant Professor of Epidemiology, McGill University
Montreal, Canada

Email: madhukar.pai@mcgill.ca

The diagnostic process is probabilistic, multivariable and sequential



1. A diagnosis starts with a patient presenting a complaint (symptom and/or sign) suggestive of a certain disease to be diagnosed.
2. The subsequent work-up is a multivariable process. It involves multiple diagnostic determinants (tests) that are applied in a logical order: from age, gender, medical history, and signs and symptoms, to more complicated, invasive, and costly tests.
3. Setting or ruling out a diagnosis is a probabilistic action in which the probability of the presence or absence of the disease is central. This probability is continuously updated based on subsequent diagnostic test results.
4. The true diagnostic value of a test is determined by the extent to which it provides diagnostic information beyond earlier tests, that is, materially changes the probability estimation of disease presence based on previous test results.
5. The goal of the diagnostic process is to eventually rule in or out the disease with enough confidence to take clinical decisions. This requires precise estimates of the probability of the presence of the target disease(s).

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

Redundancy of Single Diagnostic Test Evaluation

Karel G.M. Moons,[1,2,3] Gerri-Anne van Es,[4] Bowine C. Michel,[5] Harry R. Büller,[6] J. Dik F. Habbema,[3] and Diederick E. Grobbee[1]

Moons et al. Epidemiology 1999

Diagnostic research

Diagnostic studies as multivariable, prediction research

K G M Moons, D E Grobbee

Patient outcomes in diagnostic research

Moons et al. JECH 2002

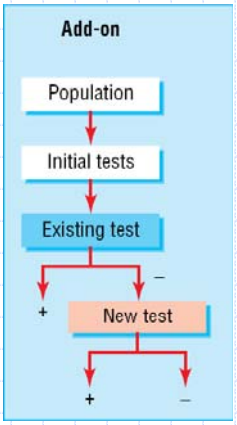
Opinion

Test Research versus Diagnostic Research

Moons et al. Clin Chem 2004

Incremental value (added value)

- What does the new test add to the diagnostic process, over and above already existing information?
- Best answered using a multivariable approach
- Sensitivity/specificity: each test is treated in isolation (which is not reflective of normal practice)



Multivariable process

- *Relate disease probability to test results*
- *Outcome = occurrence of disease (yes/no)*
- *Determinants = diagnostic tests --> dichotomous, continuous, ordinal, nominal*
- *Diagnostic function: P (D+) = f (X_1, $X_{2....}$ X_n)*
  - Where X1, X2, etc are various tests

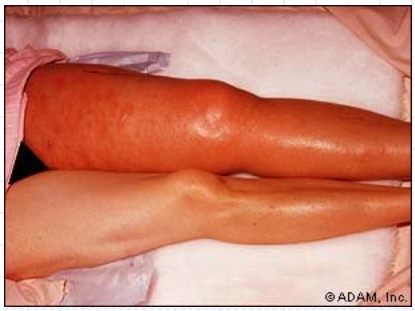
Multivariable process

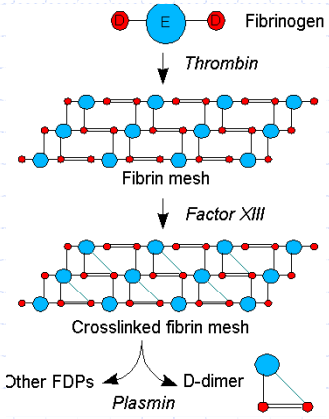
- Logistic regression model:

$$\ln\frac{P(D+|X)}{1-P(D+|X)} = b0+b1.X1+b2.X2+...+bn.Xn$$

$$P(D+|X) = \frac{1}{1+e^{-(b0+b1.X1+...+bn.Xn)}}$$

Multivariable example: does D-dimer add value to ruling out DVT?





Multivariable approach (example)

New Technologies and Diagnostic Tools

Ruling out deep venous thrombosis in primary care

A simple diagnostic algorithm including D-dimer testing

Ruud Oudega, Karel G. M. Moons, Arno W. Hoes

Julius Center for Health Sciences and Primary care, University Medical Center Utrecht, Utrecht, The Netherlands

Summary

In primary care, the physician has to decide which patients have to be referred for further diagnostic work-up. At present, only in 20% to 30% of the referred patients the diagnosis DVT is confirmed. This puts a burden on both patients and health care budgets. The question arises whether the diagnostic work-up and referral of patients suspected of DVT in primary care could be more efficient. A simple diagnostic decision rule developed in primary care is required to safely exclude the presence of DVT in patients suspected of DVT, without the need for referral. In a cross-sectional study, we investigated the data of 1295 consecutive patients consulting their primary care physician with symptoms suggestive of DVT, to develop and validate a simple diagnostic decision rule to safely exclude the presence of DVT. Independent diagnostic indicators of the presence of DVT were male gender, oral contraceptive use, presence of malignancy, recent surgery, absence of leg trauma, vein distension, calf difference and D-dimer test result. Application of this rule could reduce the number of referrals by at least 23% while only 0·7% of the patients with a DVT would not be referred. We conclude that by using eight simple diagnostic indicators from patient history, physical examination and the result of D-dimer testing, it is possible to safely rule out DVT in a large number of patients in primary care, reducing unnecessary patient burden and health care costs.

Oudega et al. Thromb Haemost 2005

Methods

- In a large cross sectional study we identified 1295 consecutive adult patients (over 18 years) who visited one of the primary care physicians adherent to three non-academic hospitals in The Netherlands, and in whom DVT was suspected by the physician on clinical grounds.
- In accordance with earlier studies, the suspicion of DVT was based on the presence of at least one of the following symptoms or signs of the lower extremities: swelling, redness, and/or pain in the legs

Oudega et al. Thromb Haemost 2005

History and physical

- After informed consent, the primary care physician systematically documented information on the patient's history and physical examination.
- Following history findings were recorded as potential diagnostic determinants: presence of previous DVT, family history of DVT, history of any malignancy (active cancer in the last 6 months), immobilization for more than 3 days, recent surgery (within past 4 weeks), leg trauma (within past 4 weeks), pain when walking, and the presence of duration of the three main symptoms (i.e. a painful, red or swollen leg).
- Physical examination items included the presence of tenderness along the deep vein system in calf or thigh, distension of collateral veins in the symptomatic leg, pitting edema in the symptomatic leg of the calf and thigh, and ≥ 3 cm difference in circumference of the calves.
- For women two additional predictors were documented, i.e. the use of oral hormonal contraception and of estrogen replacement therapy.

Oudega et al. Thromb Haemost 2005

Lab tests and reference standard

- After the standardized history taking and physical examination, all patients were referred to the hospital to undergo D-dimer testing.
- After venous blood was drawn, each patient directly underwent real time B-mode compression ultrasonography (CUS) of the lower extremities [Reference standard]

Oudega et al. Thromb Haemost 2005

Data analysis

- After univariate analysis, we first quantified which of the 16 history and physical findings independently contributed to the presence or absence of proximal DVT using multivariable logistic regression analysis.
- Starting with the overall model including all history and physical findings, model reduction (stepwise backwards) was performed by excluding variables from the model with a p-value > 0.10 based on the log likelihood ratio test.

Oudega et al. Thromb Haemost 2005

Data analysis

- Subsequently, we added the D-dimer test to this reduced model to quantify its added value, which resulted in the final model.
- The ability of a model to discriminate between patients with and without DVT was estimated using the area under the ROC curve.
- The reliability or calibration of each model was evaluated by comparing the predicted and observed probabilities for deciles of calculated patient risks and tested using the Hosmer-Lemeshow test.

Oudega et al. Thromb Haemost 2005

Results: bivariate analyses

N = 1295 patients

22% had DVT

Diagnostic variables	Total n=1295 %	DVT present n=289 %	DVT absent n=1006 %	OR (95% CI)
Patient history:				
age (years)	60.0 (17.6)[1]	62.0 (16.8)[1]	59.4 (17.8)[1]	1.01 (1.00 – 2.02)[2]
gender + OC use				
males	36	47	33	1.95 (1.47 – 2.57)
females using OC	10	10	10	1.37 (0.87 – 2.17)
females not using OC	54	43	57	-
gender + HRT use				
males	36	47	33	1.86 (1.42 – 2.43)
females using HRT	2	2	2	1.32 (0.48 – 3.63)
females not using HRT	62	51	66	-
previous DVT	24	21	25	0.82 (0.60 – 1.12)
family history of DVT	23	20	24	0.79 (0.57 – 1.09)
presence of malignancy	6	12	5	2.72 (1.71 – 4.32)
immobilization	14	13	14	0.90 (0.61 – 1.33)
recent surgery	14	19	13	1.59 (1.12 – 2.26)
absence of leg trauma	85	89	84	1.58 (1.05 – 2.36)
pain when walking	81	84	80	1.30 (0.92 – 1.84)
days of symptoms	7.9 (7.6)[1]	6.9 (6.7)[1]	8.2 (7.8)[1]	0.98 (0.96 – 0.99)[3]
Physical examination:				
vein distension	20	28	17	1.88 (1.39 – 2.55)
deep vein system tenderness	71	72	71	1.04 (0.78 – 1.39)
swelling whole leg	45	57	42	1.84 (1.41 – 2.39)
calf difference ≥ 3cm	43	67	36	3.63 (2.75 – 4.79)
D-dimer abnormal VIDAS n= 918 Tinaquant n= 377 Combined assays	 78 65 74	 99 98 99	 72 54 66	 38.2 (9.40 – 155.3) 37.3 (9.00 – 154.8) 35.7 (13.3 - 100.0)

DVT = deep vein thrombosis, n = number of patients, OR = Odds Ratio, 95%CI = 95% Confidence Interval; OC=oral contraceptive, HRT=hormonal replacement therapy; -=reference category; D-dimer abnormal for VIDAS ≥ 500 ng/ml and Tinaquant ≥ 400 ng/l; [1]Mean (standard deviation), [2] OR is estimated per year increase or decrease), [3] OR is estimated per day increase or decrease.

Results: multivariable analyses

Table 2: Independent diagnostic indicators of DVT. The final multivariate model, the figures are estimated after model validation and adjustment for over-fitting.

Diagnostic variables	Odds ratio	Regression coefficient*	p-value	Points for the rule
Male gender	1.80 (1.36 – 2.16)	0.59	<0.001	1
Oral contraceptive use	2.12 (1.32 – 3.35)	0.75	0.002	1
Presence of malignancy	1.52 (1.05 – 2.44)	0.42	0.082	1
Recent surgery	1.46 (1.02 – 2.09)	0.38	0.044	1
Absence of leg trauma	1.82 (1.25 – 2.66)	0.60	0.002	1
Vein distension	1.62 (1.19 – 2.20)	0.48	0.002	1
Calf difference ≥ 3 cm	3.10 (2.36 – 4.06)	1.13	<0.001	2
D-dimer abnormal	20.3 (8.25 – 49.9)	3.01	<0.001	6
Constant		–5.47		

DVT= deep vein thrombosis; *=natural logarithm of the odds ratio; D-dimer abnormal for VIDAS ≥ 500 ng/ml and Tinaquant ≥ 400 ng/ml. Probability of DVT as estimated by the final model =1/(1+exp-(-5·47 + 0·59*male gender + 0·75*OC use + 0·42*presence of malignancy + 0·38*recent surgery + 0·60*absence of leg trauma + 0·48*vein distension + 1·13*calf difference ≥ 3cm + 3·01*abnormal D-dimer)).

Oudega et al. Thromb Haemost 2005

Multivariable approach

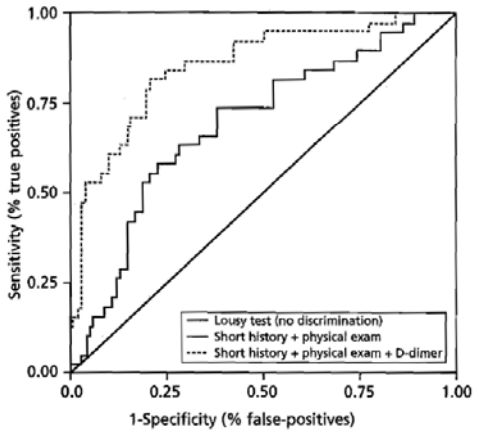


FIGURE 3.3 Example of an ROC curve of the reduced multivariable logistic regression model, including the same six determinants as in Figure 3.2. The ROC area of the "reduced history + physical model" (red) was 0.70 (95% confidence interval [CI], 0.66–0.74) and of the same model added with the D-dimer assay (green) 0.84 (95% CI, 0.80–0.88).

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

Results: scoring system

*1*male gender + 1*OC use + 1*presence of malignancy + 1*recent surgery + 1*absence of trauma + 1*vein distension + 2*calf difference ≥ 3cm +6*abnormal D-dimer test.*

Table 4: Prevalence of DVT across four score (risk) categories.

Probability or risk Category	number of patients n (%)[1]	DVT present n (%)[2]	DVT absent n (%)[3]
Very low (0–3)	293 (23)	2 (0.7)	291 (99.3)
Low (4–5)	66 (5)	3 (4.5)	63 (95.5)
Moderate (7–9)	663 (51)	144 (21.7)	519 (78.3)
High (10–13)	273 (21)	140 (51.3)	133 (48.7)

1=proportion of all (1295) patients; 2=proportion of presence of DVT within risk category; 3=proportion of absence of DVT within risk category.

Oudega et al. Thromb Haemost 2005

Another example

Acta Pædiatr 90: 611–617. 2001

Prediction of bacterial meningitis in children with meningeal signs: reduction of lumbar punctures

R Oostenbrink[1,2], KGM Moons[1,2], ART Donders[2,3], DE Grobbee[2] and HA Moll[1]

Outpatient Department of Paediatrics[1], Sophia Children's Hospital, Rotterdam; Julius Centre for Patient Oriented Research[2], University Medical Centre, Utrecht; Centre for Biostatistics[3], University of Utrecht, The Netherlands

Oostenbrink R, Moons KGM, Donders ART, Grobbee DE, Moll HA. Prediction of bacterial meningitis in children with meningeal signs: reduction of lumbar punctures. Acta Pædiatr 2001; 90: 611–617. Stockholm. ISSN 0803-5253

Physicians often have to perform a lumbar puncture to ascertain the diagnosis in patients with meningeal signs, because of the serious consequences of missing bacterial meningitis. The aim of this study was to derive and validate a clinical rule to predict bacterial meningitis in children with meningeal signs, to guide decisions on the performance of lumbar punctures. Information was collected from records of patients (aged 1 mo to 15 y) consulting the emergency department of the Sophia Children's Hospital between 1988 and 1998 with meningeal signs. Bacterial meningitis was defined as cerebrospinal fluid (CSF) leucocyte count >5 cells μl^{-1} with a positive bacterial culture of CSF or blood. The diagnostic value of predictors was judged using multivariate logistic modelling and area under the receiver operating characteristic curves (ROC area). In the derivation set (286 patients, years 1988–1995) the duration of the main complaint, vomiting, meningeal irritation, cyanosis, petechiae and disturbed consciousness were independent clinical predictors of bacterial meningitis. The ROC area of this model was 0.92. The only independent predictor from subsequent laboratory tests was the serum C-reactive protein concentration, increasing the ROC area to 0.95. Without missing a single case, this final model identified 99 patients (35%) without bacterial meningitis. Validation on 74 consecutive patients in 3 subsequent years (1996–1998) yielded similar results.

Conclusion: This prediction rule identifies about 35% of the patients with meningeal signs in whom a lumbar puncture can be withheld without missing a single case of bacterial meningitis. For the individual patient this prediction rule is valuable in deciding whether or not to perform a lumbar puncture.

Table 3. Independent predictors for bacterial meningitis

Variable	Clinical evaluation model OR (95% CI)	Clinical evaluation + laboratory model OR (95% CI)	Risk score
Patient history			
Duration of the main complaint (per day)[a]	1.5 (1.2–1.9)	1.5 (1.2–1.9)	1
Vomiting	2.4 (1.0–5.4)	2.3 (0.9–5.5)	2
Physical examination			
Meningeal irritation	25.0 (3.2–197.5)	21.1 (2.6–172.4)	7.5
Cyanosis	24.0 (2.0–289.4)	13.0 (1.1–151.3)	6.5
Petechiae or ecchymoses	7.5 (2.2–25.6)	4.9 (1.4–17.9)	4
Disturbed consciousness	22.2 (9.4–52.4)	21.8 (8.6–55.2)	8
Laboratory tests			
Serum CRP (per 10 mg l^{-1})[b]		1.1 (1.0–1.1)	0.1
ROC area (95% CI) in derivation set	0.92 (0.89–0.95)	0.95 (0.92–0.97)	0.94 (0.91–0.97)
ROC area (95% CI) in validation set	0.92 (0.86–0.98)	0.92 (0.86–0.98)	0.92 (0.86–0.98)

[a] Duration of the main complaint rounded off to half days, with a maximum of 7 points.
[b] Points assigned to serum CRP: 0.1 point per 10 mg l^{-1} increase, thus 0–9 mg l^{-1}: 0 points; 10–19 mg l^{-1}: 0.1 points; etc., with a maximum of 2 points.
OR: odds ratio; CI: confidence interval; CRP: C-reactive protein; ROC: receiver operating characteristic.

$$\text{Total score} = 1 \times \text{duration main complaint (d)} + 2 \times \text{vomiting} + 7.5 \times \text{meningeal irritation} + 6.5 \times \text{cyanosis} + 4 \times \text{petechiae} + 8 \times \text{disturbed consciousness} + 0.1 \times \text{serum CRP (per 10 mg } l^{-1}\text{)}$$

Table 4. Frequency of bacterial meningitis (BM) related to the risk score

	Derivation set ($n = 286$)		Validation set ($n = 74$)	
Risk score (points)	BM present	BM absent	BM present	BM absent
0–4.9	0	64 (100%)	0	20 (100%)
5.0–9.4	0	35 (100%)	0	14 (100%)
9.5–14.9	17 (16%)	88 (84%)	3 (15%)	17 (85%)
15.0–19.9	24 (63%)	14 (37%)	4 (44%)	5 (56%)
≥20.0	43 (98%)	1 (2%)	8 (73%)	3 (27%)

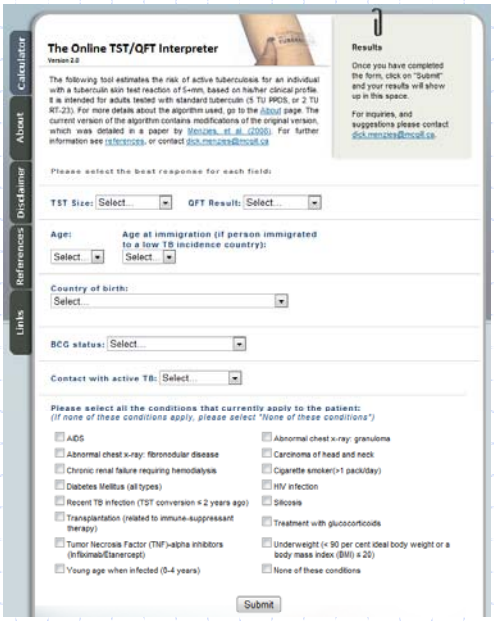
Bacterial meningitis was absent in all patients with a score <9.5 and present in almost all patients with a score >=20. The threshold value <9.5 identified 99 patients without bacterial meningitis (35%; 95% CI 29–40%), without missing a single case of bacterial meningitis. In patients with meningeal signs, a lumbar puncture can be withheld in 35% of cases without missing a single case of bacterial meningitis.

TB EXAMPLES

Composite risk prediction models for latent TB

http://www.tstin3d.com

Composite risk prediction models that incorporate biomarker and risk factors



22

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

AJRCCM 2010

TABLE 2. COEFFICIENTS AND SUMMARY STATISTICS FOR PREDICTION MODELS

	Baseline Clinical Prediction Model	Baseline Prediction Model with IFN-γ Results	Baseline Prediction Model with Clinician Suspicion	Baseline Prediction Model with Clinician Suspicion and IFN-γ Results
CXR, active disease*	2.92	3.66	0.92	1.18
Night sweats or weight loss	1.60	2.22	1.12	1.45
Previous active disease	0.29	0.27	0.24	0.0 23
US birth†	1.80	2.85	2.01	2.95
Foreign birth, ≤2 yr in US†	1.41	1.58	2.33	2.45
Foreign born, 3–12 yr in US†	2.71	3.37	2.09	2.65
Contact to active case	2.43	2.11	3.69	3.09
High clinical suspicion‡			19.43	19.31
Intermediate clinical suspicion‡			5.53	4.83
Quantitative IFN-γ result (effect size per each doubling, IU/ml)		1.07		1.07
AIC	400	374	346	323
AUC	0.71 (0.64–0.77)	0.78§ (0.73–0.84)	0.82 (0.77–0.88)	0.86‖ (0.81–0.91)

Definition of abbreviations: AIC = Akaike information criterion, a measure of the goodness of fit of a statistical model with lower values indicating better fit; AUC = Area under the receiver operating curve, the probability that a randomly selected case will have a higher test value than a randomly selected noncase; a perfect test has an area under the curve of 1.0, while a worthless test has an area of 0.5; CXR = chest radiograph.

* Reference category: inactive disease or normal CXR.

† Reference category: foreign born, >12 years in US.

‡ Reference category: low clinical suspicion.

§ Significant difference ($P < 0.001$) between this model and previous model without quantitative IFN-γ results.

‖ Significant difference ($P = 0.02$) between this model and previous model without quantitative IFN-γ results.

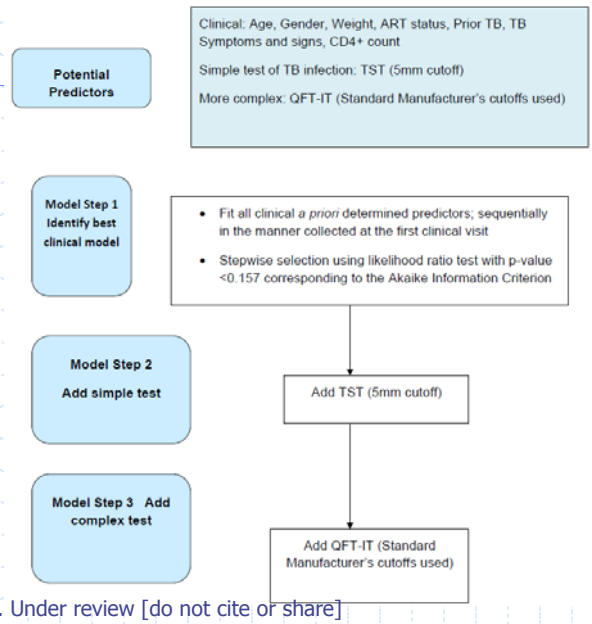
AJRCCM 2010

Discriminatory value of Mantoux and interferon release added to clinical algorithms for smear negative HIV-tuberculosis

Molebogeng X Rangaka[1,2,3], Hannah P Gideon[1], Katalin A Wilkinson[4,1,5], Madhukar Pai[6], Judith Mwansa[1], Gary Maartens[7,5], Andrew Boulle[2], Gilles van Cutsem[2], Judith Glynn[3], Katherine Fielding[3], Rene Goliath[1], Raylene Titus[1], Shahied Mathee[8], Robert J Wilkinson[9,1,4,5]

Rangaka M et al. Under review [do not cite or share]

Figure 1 Flow-chart detailing steps in the development of multivariate models



Rangaka M et al. Under review [do not cite or share]

Table 4 Multivariate logistic regression estimates for culture positive disease in clinical model without and with tests of TB infection

	A. Clinical model	B. With TST (5mm)	C. With QFT	D. With TST (5mm) and QFT
Multivariate Predictors	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**	**OR (95% CI)**
Clinical				
Weight less than 60kg	2.3 (1.3-4.2)	2.6 (1.4-4.8)	2.6 (1.4-4.8)	2.7 (1.5-5.1)
No prior TB	2.8 (1.3-6.0)	2.6 (1.2-5.7)	2.6 (1.2-5.5)	2.5 (1.1-5.4)
Anyone positive TB symptom/sign	3.1 (1.5-6.2)	3.0 (1.5-6.1)	3.1 (1.5-6.4)	3.0 (1.5-6.2)
CD4+ count less than 250 cells/mm^3	1.7 (0.8-3.5)	2.0 (1.0-4.4)	1.8 (0.9-3.9)	2.1 (1.0-4.6)
Tests of TB infection				
TST positive at 5mm		3.5 (1.8-6.6)		2.7 (1.4-5.4)
QFT (manufacturer's cutoffs)				
Positive			3.0 (1.5-5.7)	2.1 (1.0-4.1)
Indeterminate			1.5 (0.4-5.6)	1.5 (0.4-5.6)
Negative			1	1
Incremental value performance measures				
*H-L Goodness of Fit p-value	0.639	0.817	0.658	0.793
**Akaike Information Criterion (AIC)	349	335	341	334
***AUC (95% CI)	0.72 (0.65-0.79)	0.78 (0.72-0.84)	0.74 (0.64-0.82)	0.79 (0.72-0.86)
*** AUC comparison p-value	-	0.03	0.41	0.01
***LRT p-value	-	<0.001	0.003	<0.001

* Hosmer-Lemeshow (H-L) goodness of fit test, H_0= There is no difference between observed and model-predicted probabilities. ** A small AIC infers minimum prediction error. ***Models B-D compared to A. LRT: Likelihood Ratio Test, for the reduced model nested in the full model. P<0.05 indicate that the added predictor should not be dropped from the full model.

Rangaka M et al. Under review [do not cite or share]

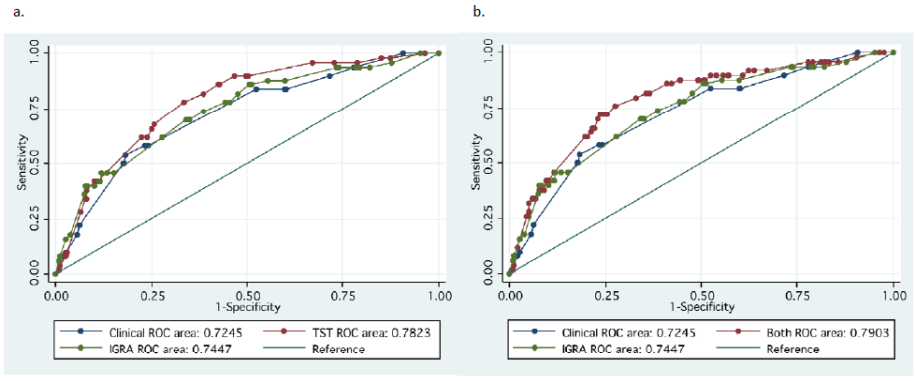


Figure 3a & b. Figure 3a & b. Receiver operator characteristic area under the curve comparisons of the simple clinical model against the model extended with TST and/or QFT

Rangaka M et al. Under review [do not cite or share]

STATISTICS IN MEDICINE
Statist. Med. 2008; **27**:157–172
Published online 13 June 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2929

Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond

Michael J. Pencina[1,*,†], Ralph B. D'Agostino Sr[1], Ralph B. D'Agostino Jr[2] and Ramachandran S. Vasan[3]

[1]*Department of Mathematics and Statistics, Framingham Heart Study, Boston University, 111 Cummington St., Boston, MA 02215, U.S.A.*
[2]*Department of Biostatistical Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, U.S.A.*
[3]*Framingham Heart Study, Boston University School of Medicine, 73 Mount Wayte Avenue, Suite 2, Framingham, MA 01702-5803, U.S.A.*

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

AJRCCM 2010

Based on prespecified risk thresholds, the NRI reflects the net proportion of patients with culture-positive tuberculosis reclassified into a higher-risk category, plus the net proportion of patients without culture positive tuberculosis reclassified into a lower-risk category.

TABLE 4. RISK RECLASSIFICATION FOLLOWING INCORPORATION OF IFN-γ RESULTS. COMPARISON TO EXPANDED 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	9	0	16	56
5–20% risk	3	6	7	16	25
>20% risk	1	0	32	33	−3
Total No.	11	15	39	65	
In 595 patients who ruled out for active tuberculosis					
≤5% risk	334	121	0	455	−27
5–20% risk	20	34	14	68	9
>20% risk	9	18	45	72	38
Total No.	363	173	59	595	

Net reclassification improvement = 3.7% ($P = 0.31$). Reclassification among patients who developed culture-positive disease = 18.5% ($P < 0.01$); reclassification among patients who ruled out for active tuberculosis = −14.8% ($P = 1$).

Metcalfe JZ et al. AJRCCM 2010