

Evaluation of reproducibility and reliability of TB diagnostics

Advanced TB Diagnostic Research Course
6 March 2012



Richard van Zyl-Smit
Division of Pulmonology & UCT Lung Institute,
Department of Medicine,
University of Cape Town, South Africa



“No financial relationships to disclose”



Scenario

You (or a friend) are exposed to someone at work with TB....

The staff clinic suggests you have a 'TB test'...

A week later they phone to say it is **“positive”**

Now what....?

Options:

1. Ignore it!
2. Phone a friend... or even ask the audience...
3. Ask for another one as you are sure they did the wrong test...
4. Ask to have it repeated in 3 months as you don't believe the result...
5. Ask them what cut-point they used and what the sensitivity, specificity and positive predictive values for the test are, and what the equated risk is for you developing active TB in the next 5 years, and what the risk reduction is if taking isoniazid preventative therapy, and how this equates to the risk of drug induced hepatotoxicity or neuropathy !

Overview of the next 45 min...

The aim of this talk is to make you skeptical...

- 1) What is a diagnostic test?
- 2) What factors can affect a test
- 3) Do test results always remain positive/ negative
- 4) How does one define a “positive” result...
- 5) Interpreting serial tests...

What is a diagnostic test ?

- HIV, hepatitis, syphilis,
 - Pregnancy
 - Lipogram
 - Glucose, creatinine, haemoglobin
 - X ray, MRI scan
-
- Tuberculin skin test
 - Interferon gamma release assay
 - LAM
 - Sputum smear
 - Xpert MTB/RIF

What factors can affect a test...

- Important to understand how a test is done...
- **X ray** – “ionizing radiation beam passed through the body causing a shadows dependent on the density of the part of the body...”
- **Haemoglobin** – “finger prick” measure the light absorbance through a drop of blood and then calculate how much Hb there is...”
- **HIV test** “measure the antibody level to the HI virus...”
- **Tuberculin skin test** “inject PPD (crushed mycobacterium) into the skin and wait 3 days then measure the lump on the arm”
- **Interferon gamma release assay (IGRA)**
- **Xpert MTB RIF** : wash away all degraded bugs and PCR what remains...

What are reliability and reproducibility ...

Reliability [noun]

- The quality of being **reliable, dependable or trustworthy**.
- The quality of a measurement indicating the **degree to which the measure is consistent**, that is, **repeated measurements would give the same result**.

Reproducible [adjective]

- *(of a measurement, experiment etc)* **Capable of being reproduced at a different time** or place and by different people.

Wikipedia



Lingo continued...

- Variability
- Within-subject variability
- Inter subject variability
- Test- retest reproducibility
- Repeatability

Measures of reliability...

- Standard error (SEM) – indicative of precision
 - When measures are repeated on the same subject
- Coefficient of variation = $SD/mean$
- 95% limits of agreement
 - Calculate the mean difference between two measurements – then $2 SD = 95\%$ limits
- Cronbach's- Alpha (psychometric instruments)
- For dichotomous data
 - Inter-reader reliability - Kappa...

Calculating and evaluating Kappa

- How reliable (agreement) is a CXR for diagnosing TB?
 - Inter-reader agreement (κ as dichotomous)
 - First calculate raw/crude agreement - then
 - $\kappa = (\text{observed} - \text{chance}) / \text{chance}$

0 = 0.2 poor agreement

0.21 to 0.4 = fair agreement

0.41 to 0.6 = moderate agreement

0.61 to 0.8 = good agreement

0.81 to 1.0 = excellent agreement

		Reader 1	
		Lymphadenopathy	No
Reader 2	Yes	30	16
	No	24	305

Raw agreement: 89.3%

Kappa 0.54 (0.40-0.67)

		Reader 1	
		Pleural abnormalities	No
Reader 2	Yes	100	31
	No	18	226

Raw agreement: 89.6%

Kappa 0.71 (0.63-0.78)

		Reader 1	
		Consistent with TB	No
Reader 2	Yes	349	22
	No	1	3

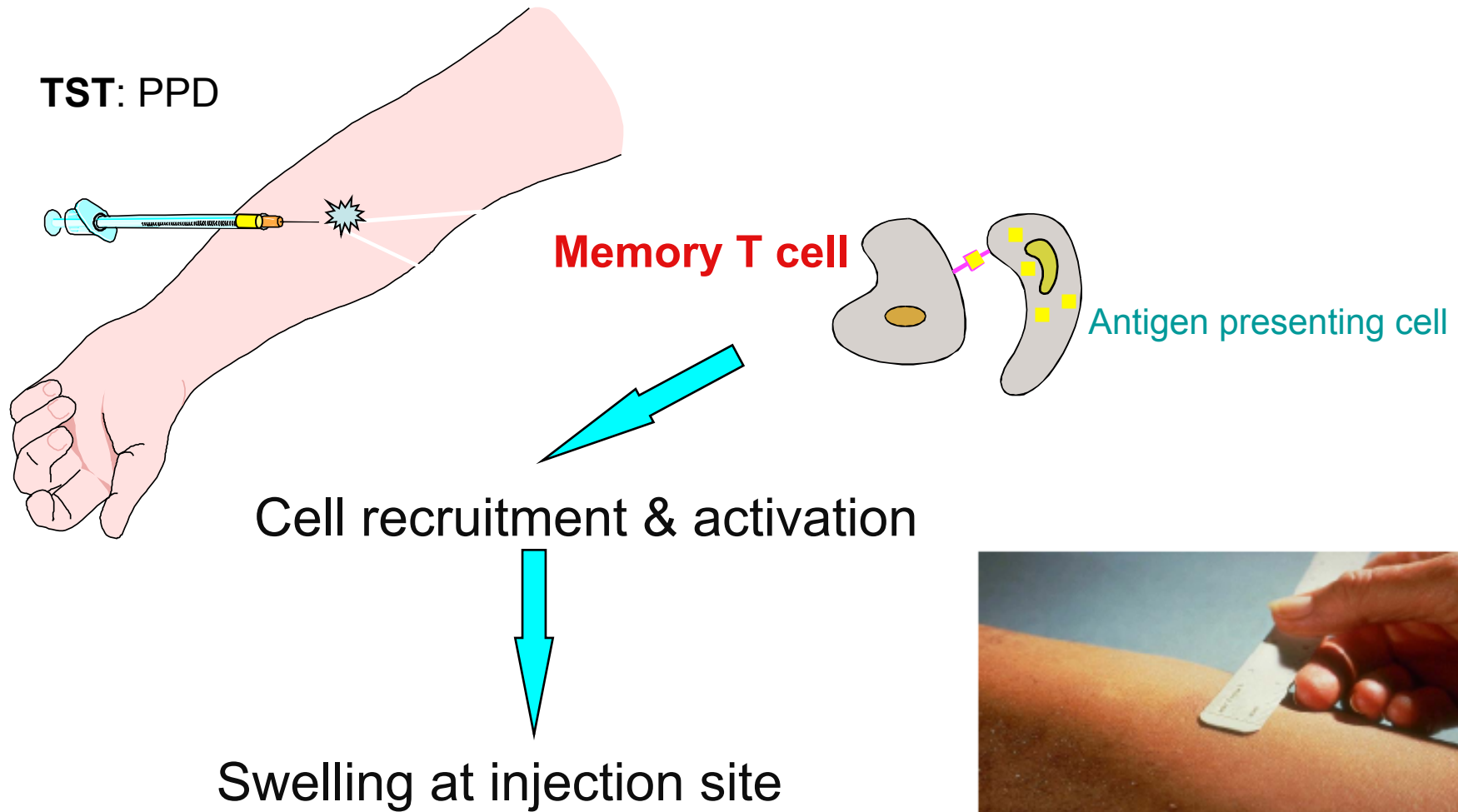
Raw agreement: 93.9%

Kappa 0.19 (-0.13 – 0.51)

Factors affecting a test....

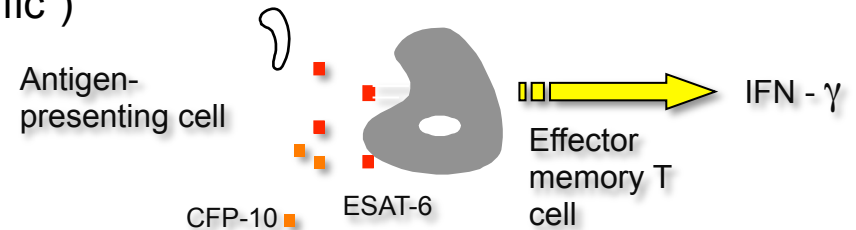
- Administration and performing the test....
 - Questionnaire (very prone to reliability issues)
 - Radiology (technical issues and interpretation)
 - Skin test (must be administered and read correctly)
 - Blood test
 - “fill the correct tube, correctly”
 - Acquisition - Time of day/ tubes/ ambient conditions
 - Processing

LTBI diagnosis: Mantoux or Tuberculin Skin Test (TST)

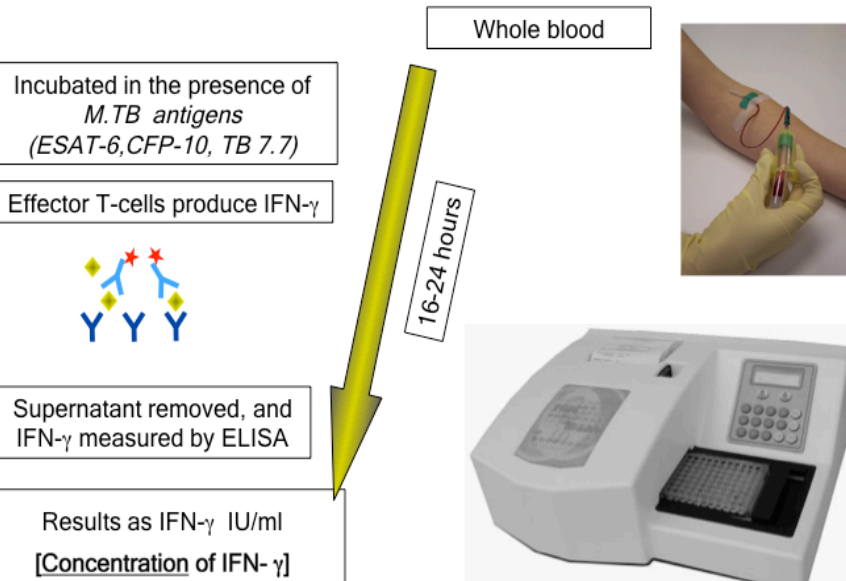


Principle underlying immuno-diagnosis ...

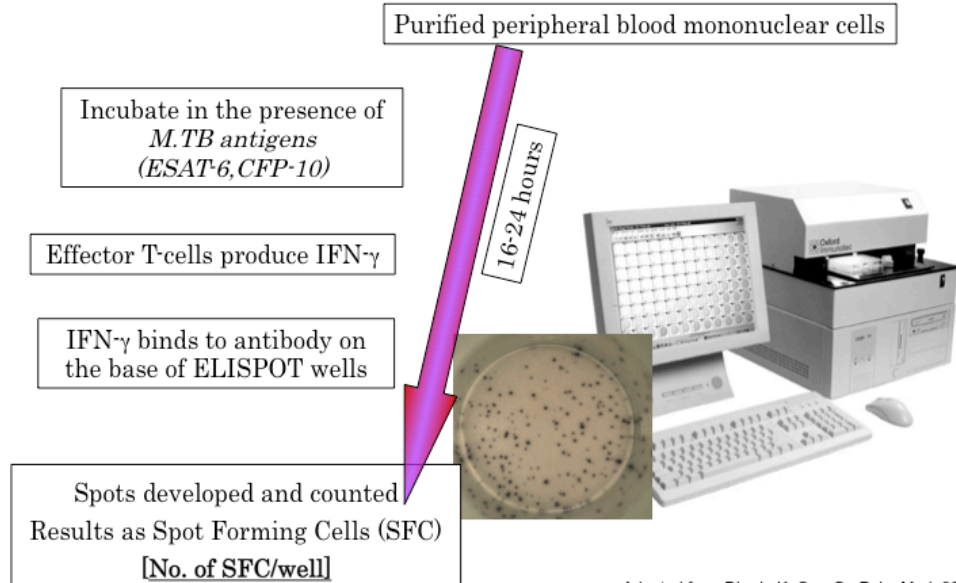
- ‘Previously sensitized’ effector memory T cells, when challenged by TB antigens, will release inflammatory cytokines including IFN- γ
- PPD non-specific
- IGRA
 - Utilizes RD-1 antigens (“Relatively TB specific”)
 - detects IFN- γ production *in vitro*



QuantiFERON®-TB Gold (In-Tube Method) (Cellestis, Australia)



T-SPOT.® TB assay (Oxford Immunotec, UK)



“Repeat testing”

- Important questions & concepts
 - What happens if I take two samples and do the test?
 - What happens if I take the sample twice on one day?
 - What happens if the lab tech processes the same sample twice?
 - What happens if we do the same test three days apart?
 - What happens if we do the same test a month apart?
 - What happens if we do the test after an “intervention”?
 - What happens when I test the same sample with two similar methods?
 - What happens when I test the same sample with two different methods?
 - What happens when I test different samples with different methods?
- Why are these questions important?
 - By example: If one wishes to determine if an intervention has any effect
 - ? Does treating for TB affect IGRA results?
 - ? Does doing a TST affect the IGRA result?

A very real life example...

From:

Sent: 09 April 2010 03:39 PM

To: rvzs@iafrica.com

Subject: QFT series

Hi Richard,

Hope this mail finds you in good shape in Cape Town, inches away from a PhD ceremony!!

Can I ask you for your opinion about the QuantiFERON test-results series of **a 44 year old medical microbiology analyst**, working in the infectious disease sector, including mycobacteria. **They are checked three times/ year.** She cannot recall a lab incident (and I assume they work under adequate BSL 3 conditions) or documented TB contact recently. A CXR was normal in January. She is referred to me by the hospital's occupational health physician and I will see her next week. Can I still interpret the latest QFT result as "wobbling around the curve" and "within-subject variability" and simply repeat the next (routine) QFT in 4 months? Thanks in advance for your attention.

Cheers

QFT series...

- QFT Series:
 - okt 2007 0,05
 - mrt 2008 0,02
 - jun 2008 0,06
 - okt 2008 0,21
 - mei 2009 0,14
 - **mrt 2010 0,47**
 - **June 2010 0.16 !**

Hi Richard,

- Thanks for your practical advice. I will repeat after 3 -4 months (and hope it will wobble back under the cut-off). If persistent conversion I will consider 3HR, since we consider conversions < 2 years as recent and in the context of possible professional exposure to (10% H-monoresistant in the NL) mycobacteria.

Real life experiments...

Within-Subject Variability and Boosting of T-Cell Interferon- γ Responses after Tuberculin Skin Testing

Richard N. van Zyl-Smit¹, Madhukar Pai², Kwaku Peprah¹, Richard Meldau¹, Jackie Kieck³, June Juritz⁴, Motasim Badri¹, Alimuddin Zumla⁵, Leonardo A. Sechi⁶, Eric D. Bateman¹, and Keertan Dheda^{1,5,7}

Am J Respir Crit Care Med Vol 180. pp 49–58, 2009

Originally Published in Press as DOI: 10.1164/rccm.200811-1704OC on April 16, 2009

Internet address: www.atsjournals.org

OPEN ACCESS Freely available online



Within-Subject Variability of Interferon-g Assay Results for Tuberculosis and Boosting Effect of Tuberculin Skin Testing: A Systematic Review

Richard N. van Zyl-Smit¹, Alice Zwerling^{2,3}, Keertan Dheda^{1,4,5}, Madhukar Pai^{2,3*}

¹ Lung Infection and Immunity Unit, Division of Pulmonology and UCT Lung Institute, Department of Medicine, University of Cape Town, Cape Town, South Africa,

² Department of Epidemiology & Biostatistics, McGill University, Montreal, Canada, ³ Montreal Chest Institute, Montreal, Canada, ⁴ Institute of Infectious Disease and

Molecular Medicine, University of Cape Town, South Africa, ⁵ Centre for Infectious Diseases and International Health, University College Medical School, London, United

Kingdom

Study Aims

1. To document 'within-subject' variability
(Short term reproducibility over a 3 week period)
2. To examine the effect of the TST on IGRA responses
(Does TST 'boost" subsequent IGRA responses)

Within-Subject Variability and Boosting of T-Cell Interferon- γ Responses after Tuberculin Skin Testing

Richard N. van Zyl-Smit¹, Madhukar Pai², Kwaku Pephrah¹, Richard Meldau¹, Jackie Kieck³, June Juritz⁴, Motasim Badri¹, Alimuddin Zumla⁵, Leonardo A. Sechi⁶, Eric D. Bateman¹, and Keertan Dheda^{1,5,7}

Am J Respir Crit Care Med Vol 180, pp 49-58, 2009

Originally Published in Press as DOI: 10.1164/rccm.200811-1704OC on April 16, 2009

Internet address: www.atsjournals.org

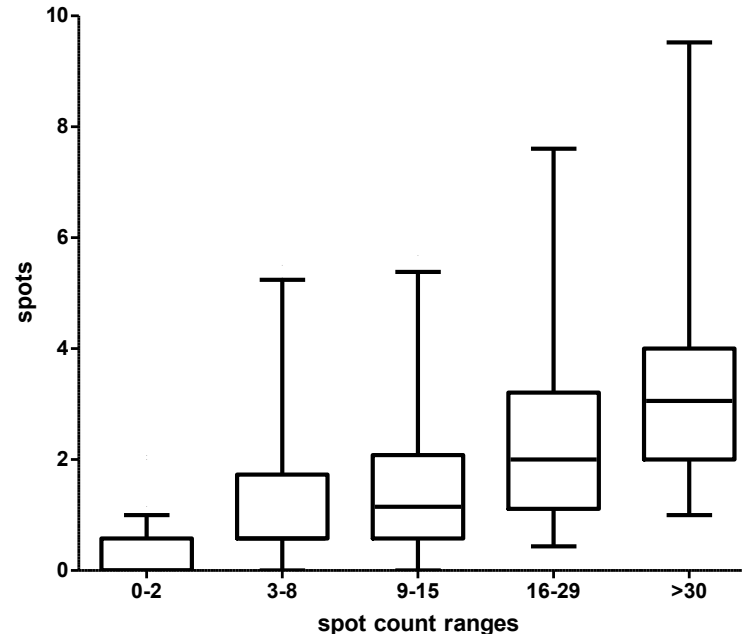
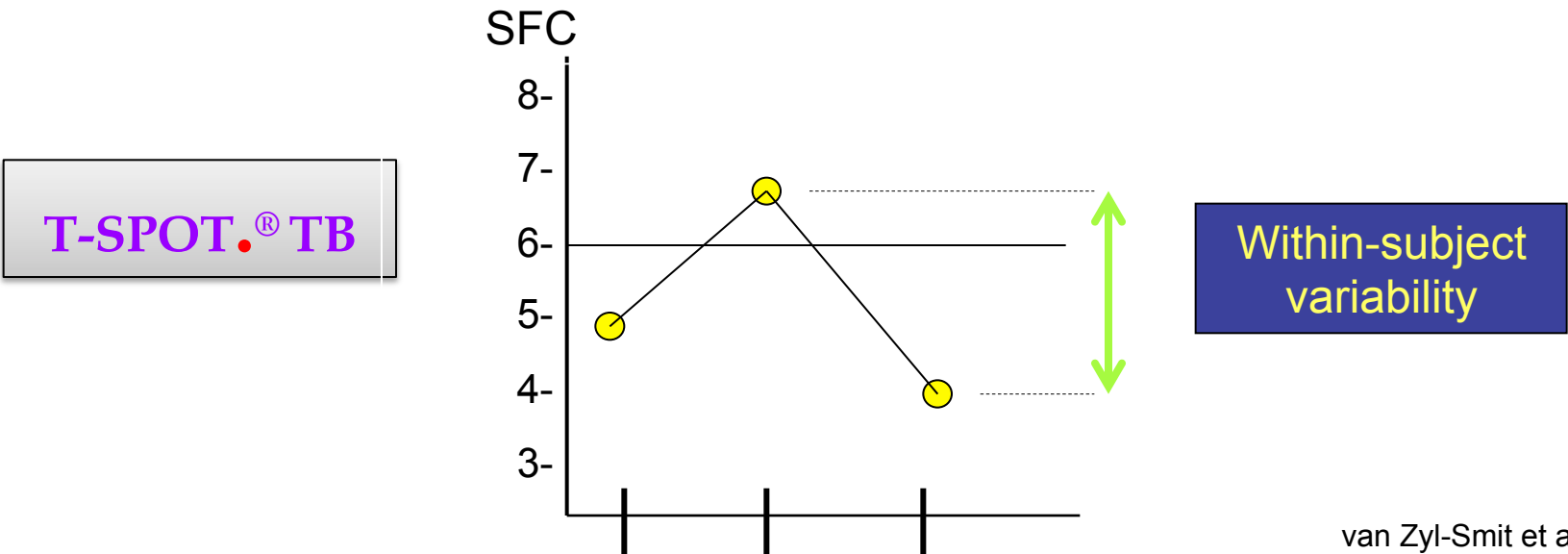
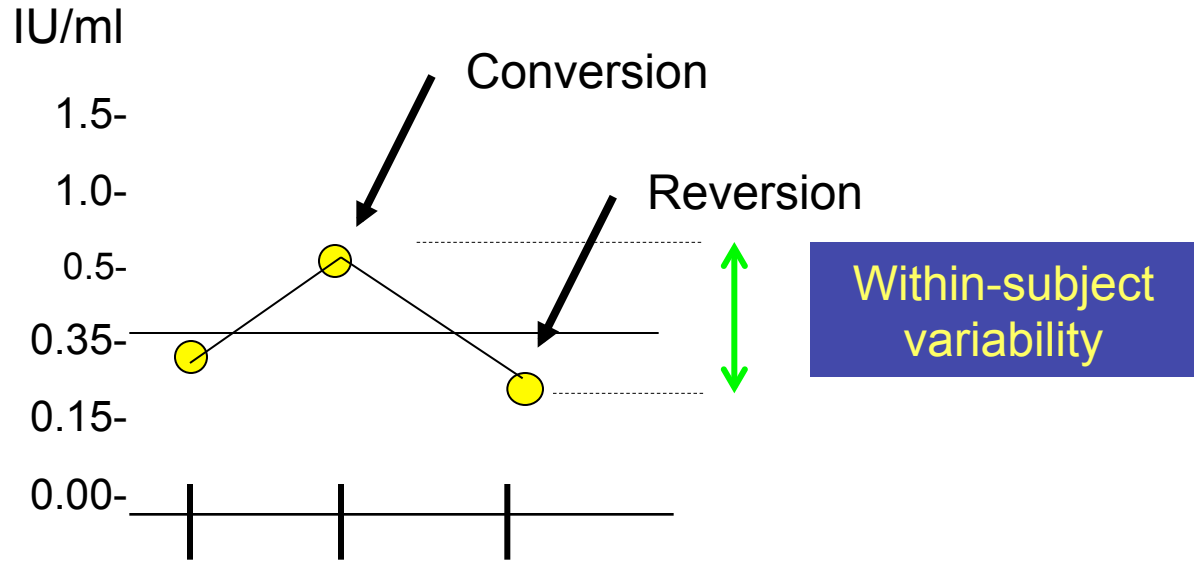


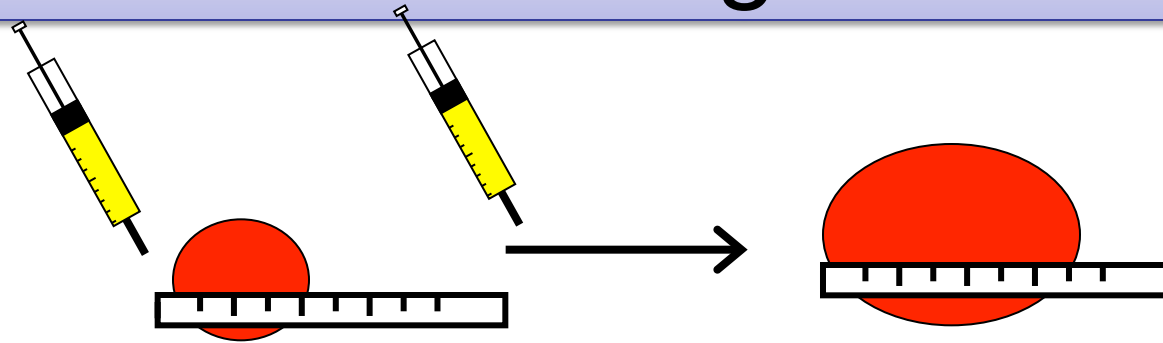
Figure E1 Repeated counting of 216 ELISPOT wells using the same AID reader on three separate days. The median is depicted by the horizontal line within box and whiskers depict the 95% confidence intervals.

Within-subject variability and “conversion & reversions”

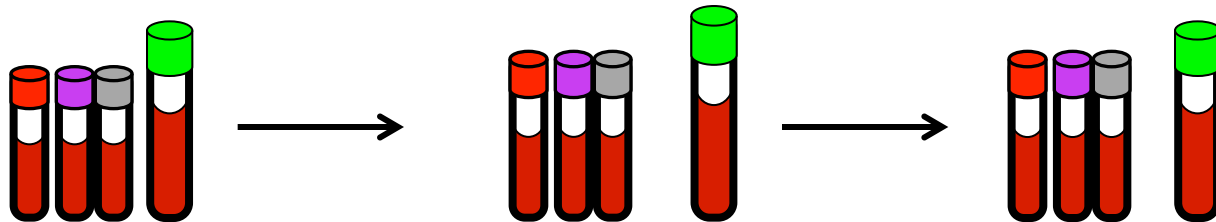


Boosting

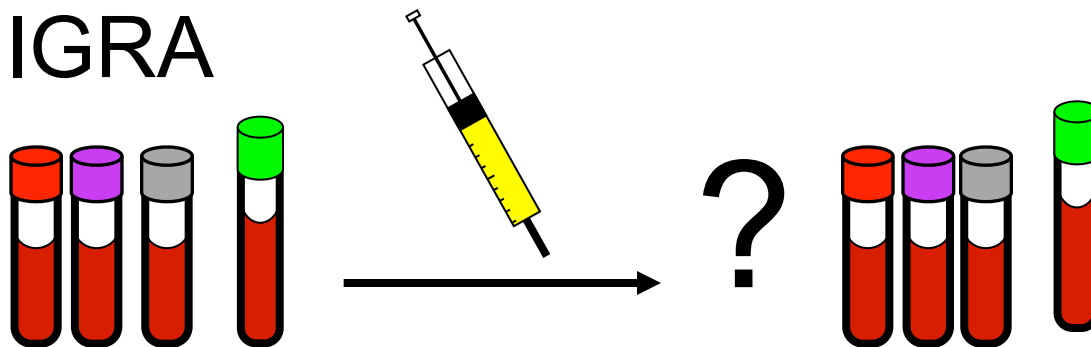
- TST



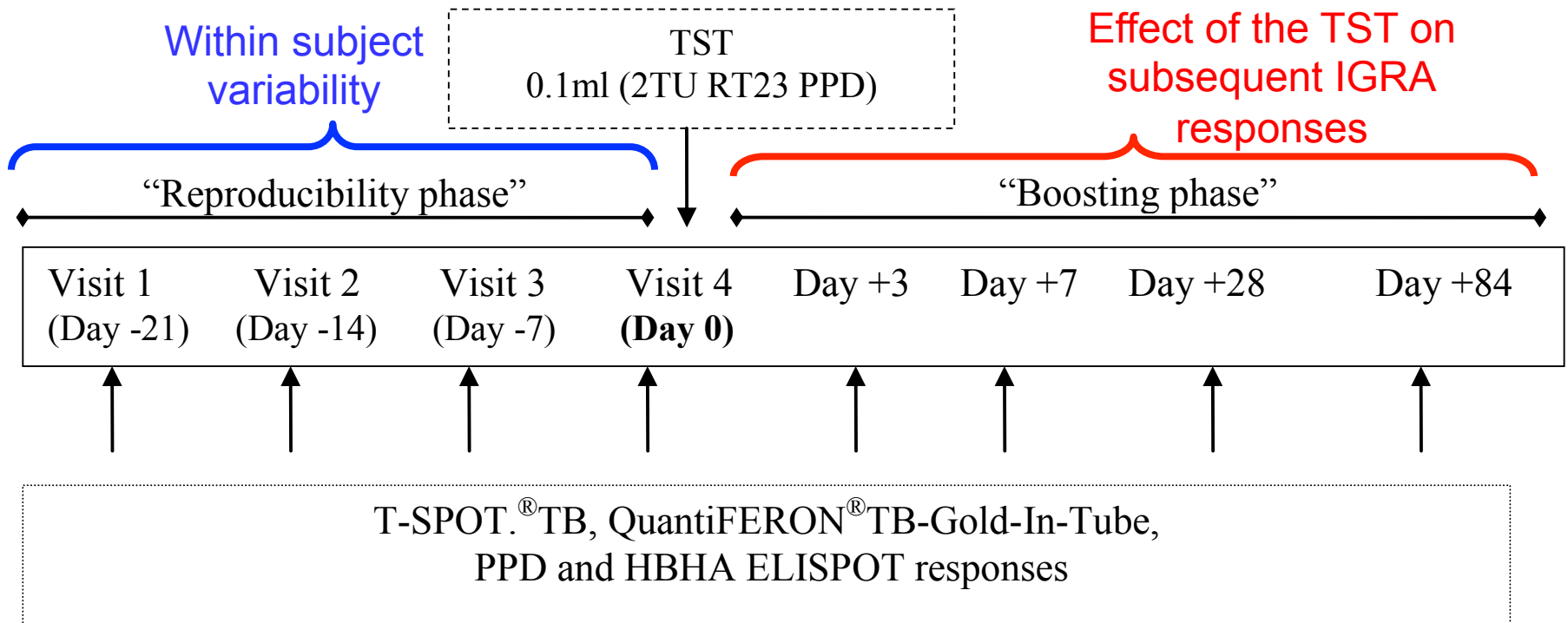
- IGRA



- TST - IGRA

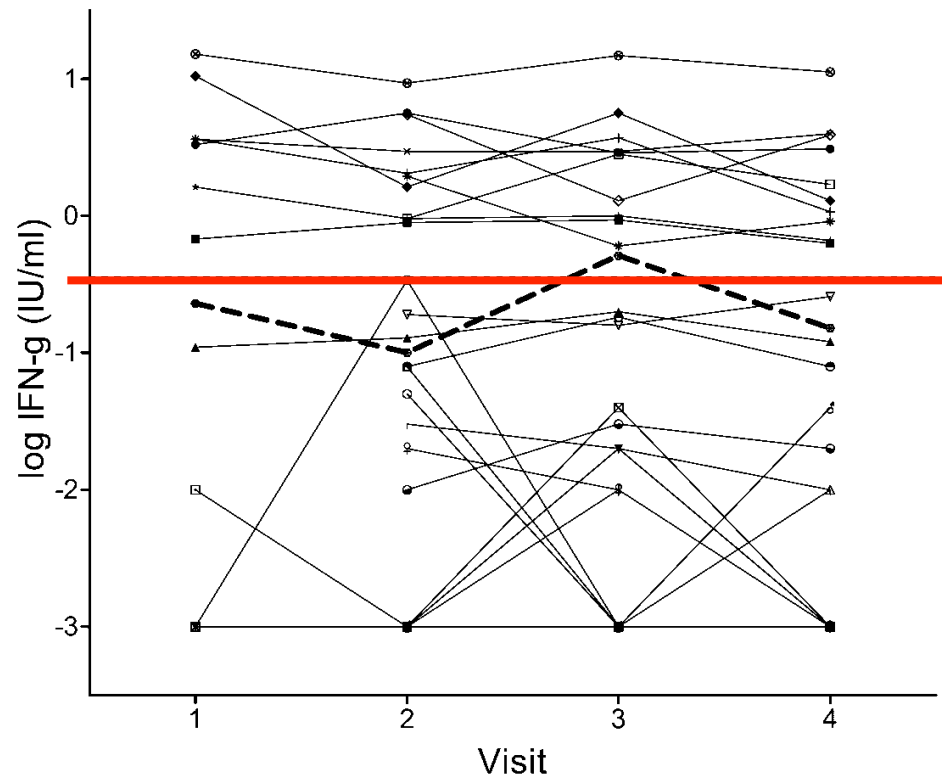
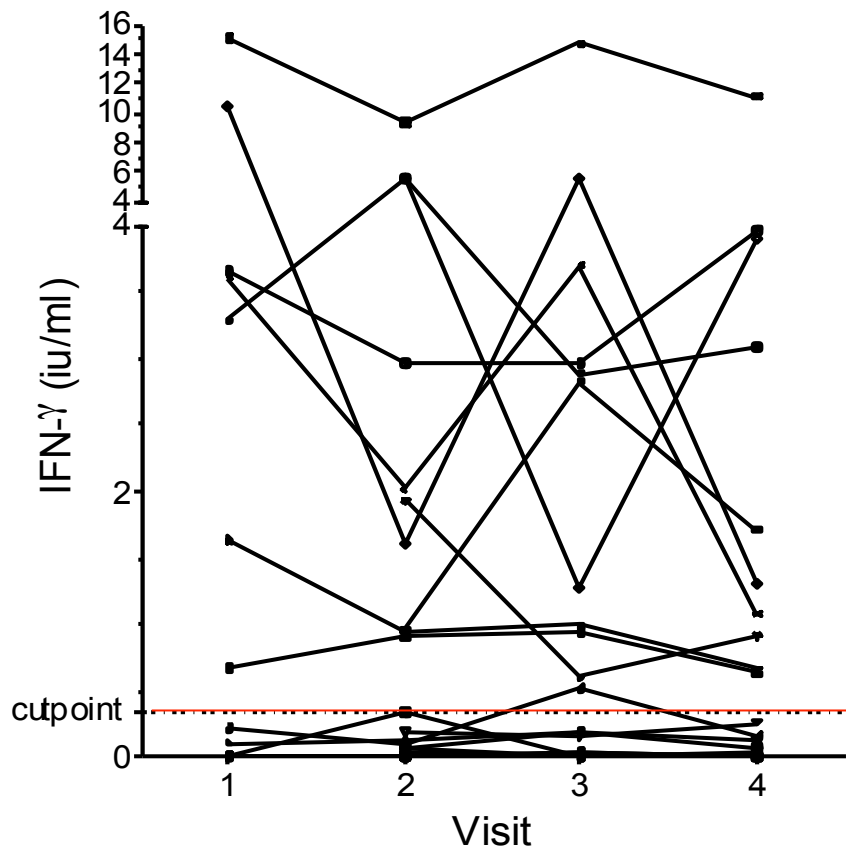


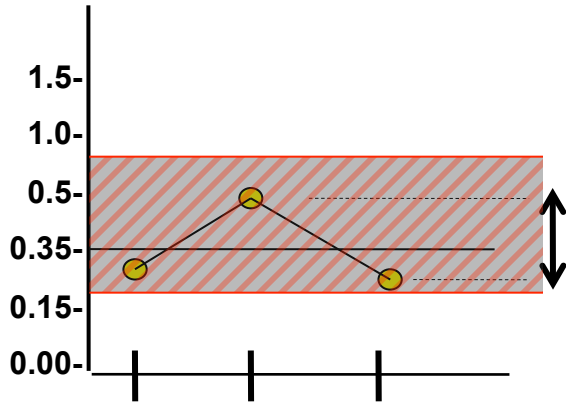
Study Design



Within-subject variability

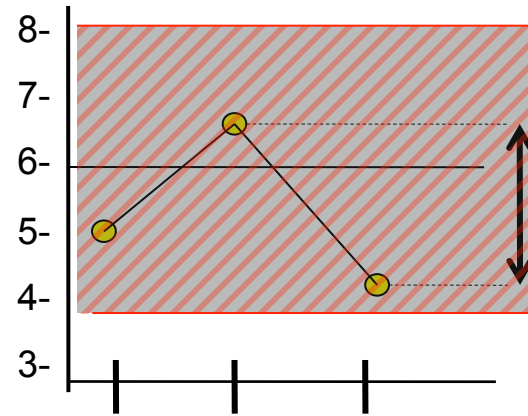
QuantiFERON® TB GIT






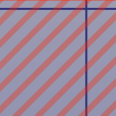
Within-subject variability

QuantiFERON®-TB
Gold-IT



Within-subject variability

T-SPOT.® TB

	QuantiFERON®TB GIT	T-SPOT.®TB
Manufacturer defined assay cut-point	> 0.35 IU/ml	≥ 6 Spots *
Within- subject short term variability 	± 80% of IFN-γ response	± 3 Spots
Borderline or uncertainty zone 	0.2-0.7 IU/ml	4 – 8 spots (inclusive)
Proposed conversion threshold	Increase from below 0.35 to above 0.7 IU/ml	Increase from below 6 to above 9 spots (inclusive)

So what is a positive test result?

- Defining the cut-off for positive and negative
 - Sensitivity
 - Specificity
 - Positive predictive value
 - Negative predictive value

Sensitivity (true positive rate)

	Disease +ve	Disease -ve
Test +ve	True positive	False Positive
Test -ve	False Negative	True Negative

Proportion of those with the disease who will test positive

$$\text{True pos} / (\text{true pos} + \text{false negative})$$

Specificity (true negative rate)

	Disease +ve	Disease -ve
Test +ve	True positive	False Positive
Test -ve	False Negative	True Negative

Proportion of those without the disease who will test negative

$$\text{True neg} / (\text{true neg} + \text{false positive})$$

Positive predictive value

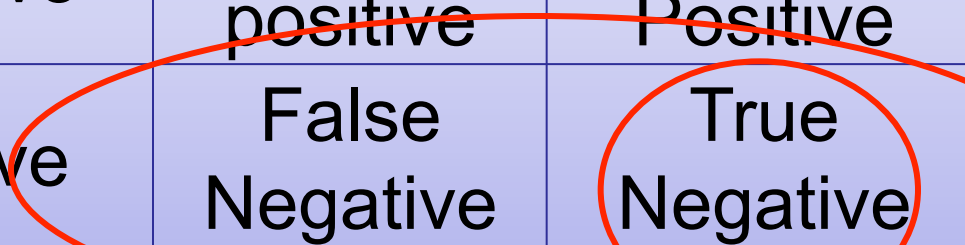
	Disease +ve	Disease -ve
Test +ve	True positive	False Positive
Test -ve	False Negative	True Negative

Proportion of subjects with a positive test who have the disease
(true positive / True positive + false positive)

Note: predictive values of a test will depend on the prevalence

Negative predictive value

	Disease +ve	Disease -ve
Test +ve	True positive	False Positive
Test -ve	False Negative	True Negative



Proportion of subjects with a negative test who do not have the disease
(true positive / True positive + false positive)

Cut offs....

- The manufacturer will define their cut off depending on what the test is for...
 - **Screening** – Highly sensitive (i.e. doesn't miss anyone)
 - **Confirmatory** – Highly specific (i.e. does not falsely give positive results)
- Tests such as haemoglobin, creatinine, lung function etc have “normal ranges” rather than positive or negative based on population tests of normal subjects...

Defining a cut off

- Decide what the test is used for...
 - “predefined by the manufacturer”
 - “decide on your own”!

Usually a balance between sensitivity and specificity

- What you gain on the one side you lose on the other!

Defining an experimental cut off

Eur Respir J 2009; 34: 1118–1126
DOI: 10.1183/09031936.00005309
Copyright ©ERS Journals Ltd 2009



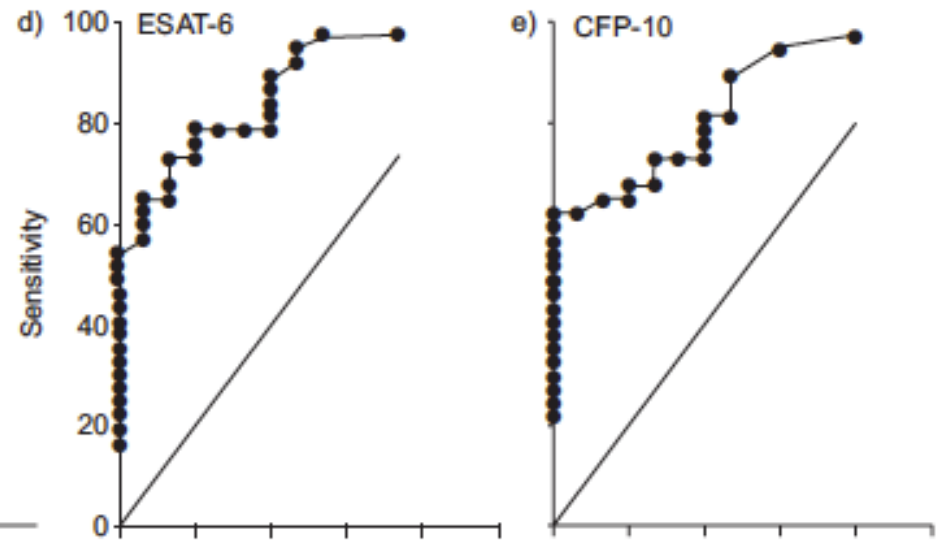
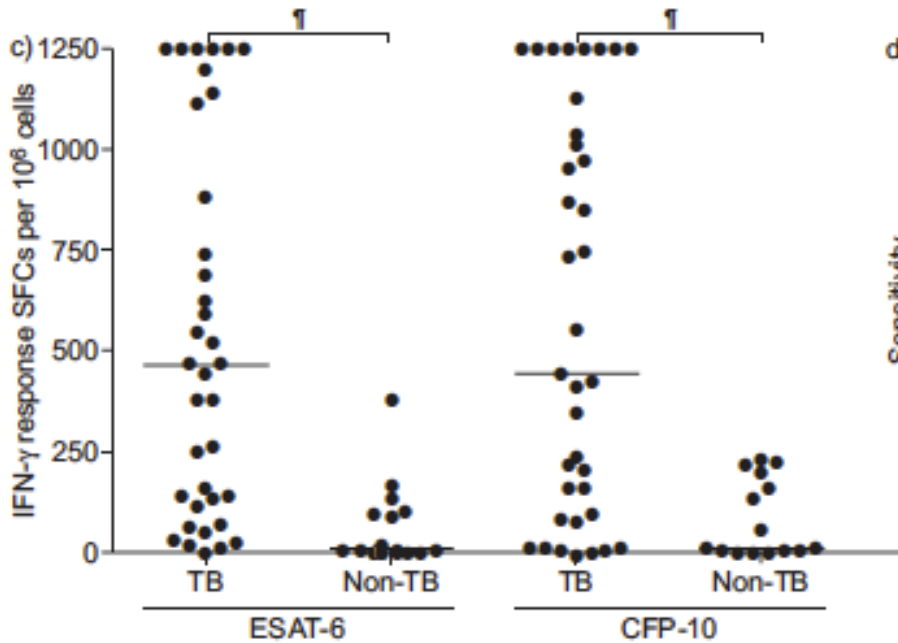
Utility of quantitative T-cell responses
versus unstimulated interferon- γ for the
diagnosis of pleural tuberculosis

K. Dheda^{*,#,†}, R.N. van Zyl-Smit^{*}, L.A. Sechi⁺, M. Badri^{*}, R. Meldau^{*}, S. Meldau^{*},
G. Symons^{*}, P.L. Semple^{*}, A. Maredza^{*}, R. Dawson^{*}, H. Wainwright[§], A. Whitelaw^f,
Y. Vallie^{*}, P. Raubenheimer^{*}, E.D. Bateman^{*} and A. Zumla[†]

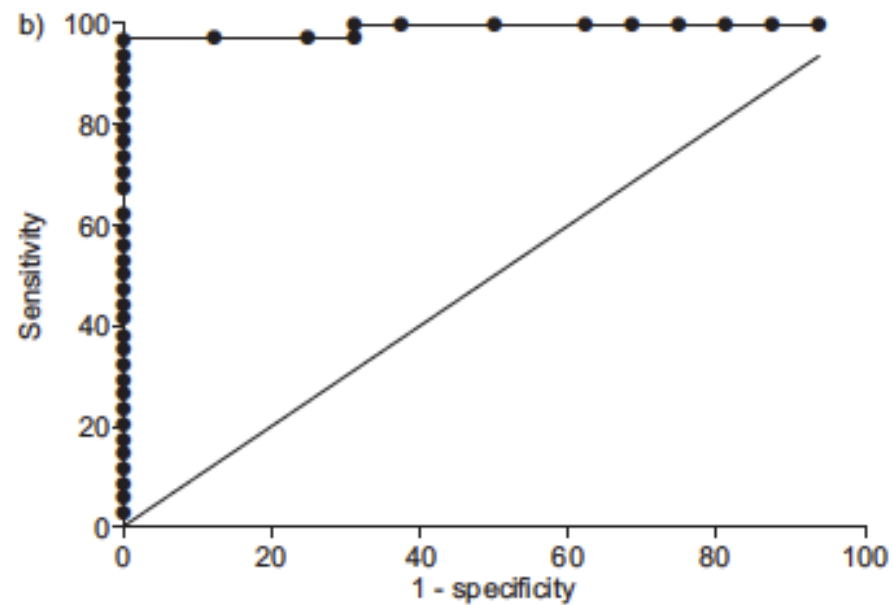
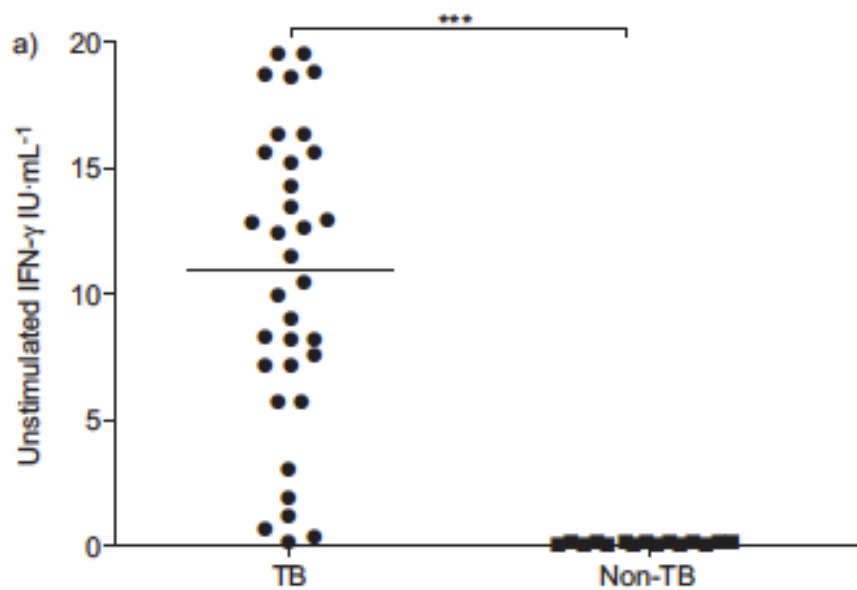
Can you use an IGRA to diagnose TB in the pleural space?

- **No manufacturer's cut point....**

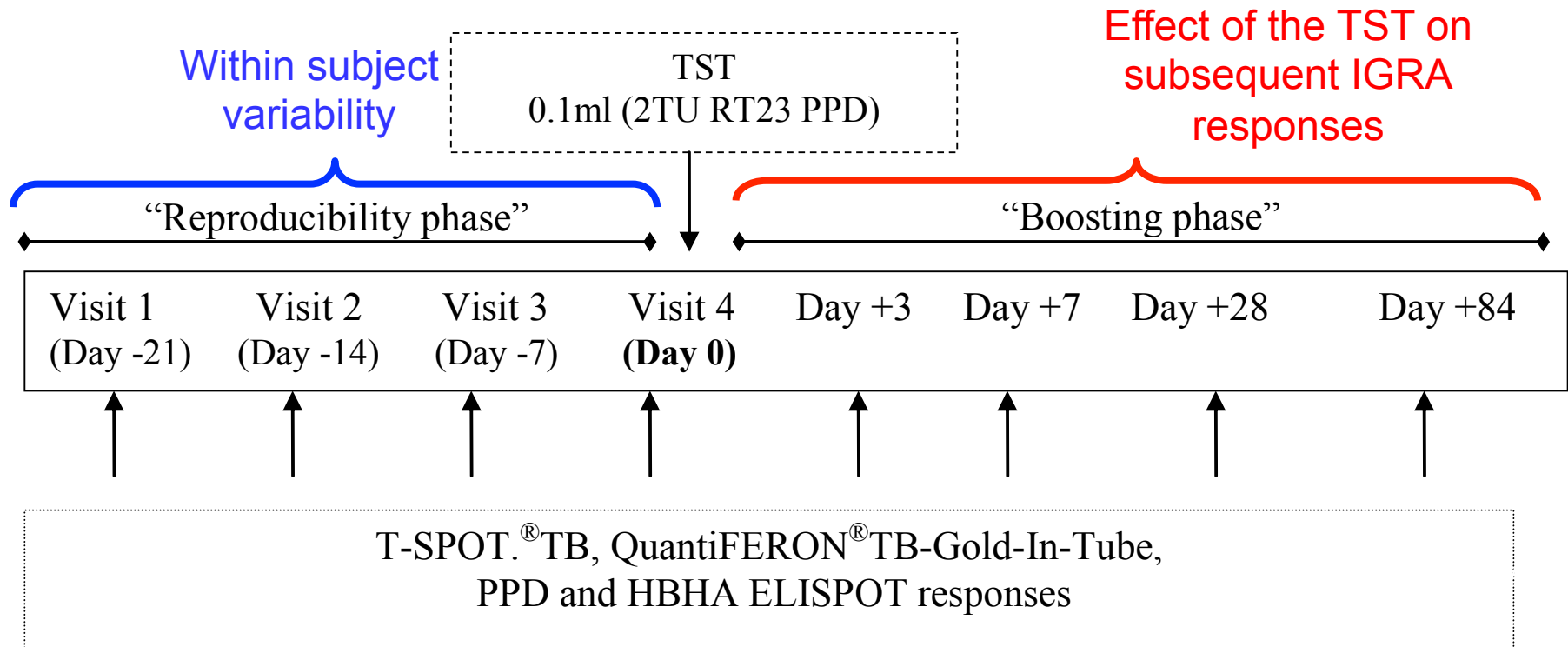
Pleural Fluid RD1-antigens (T-SPOT.[®]TB)



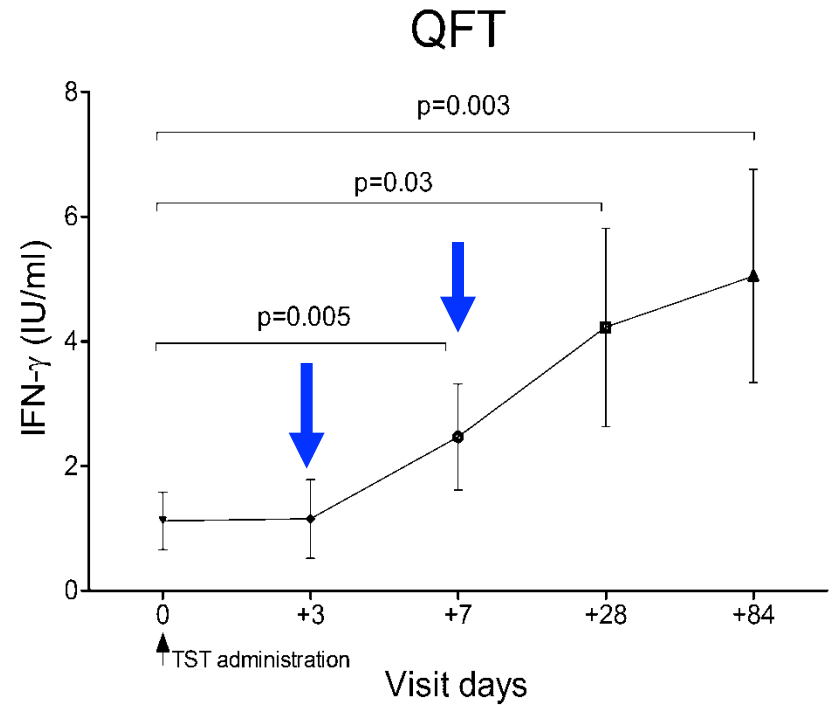
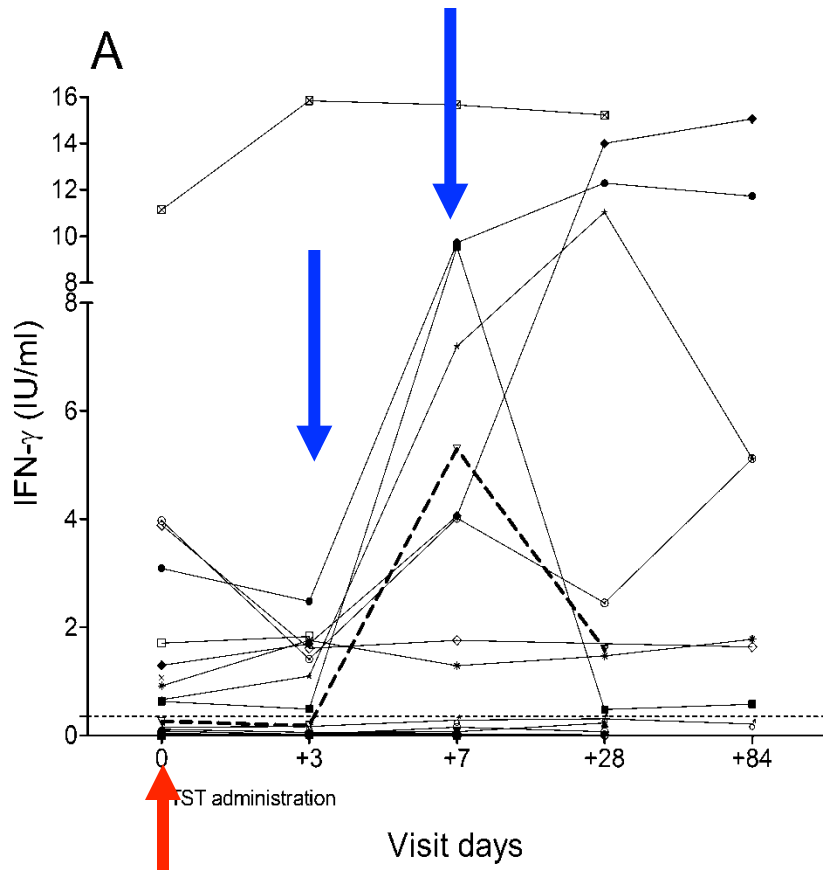
IFN-gamma



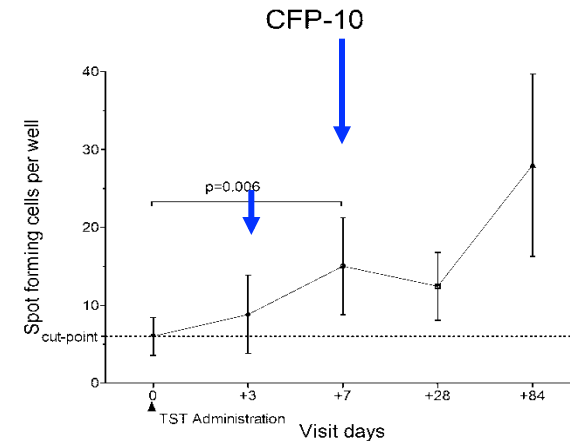
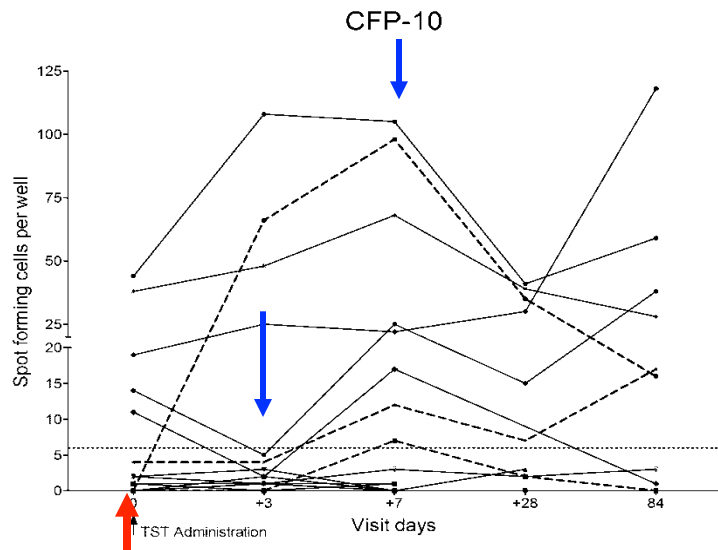
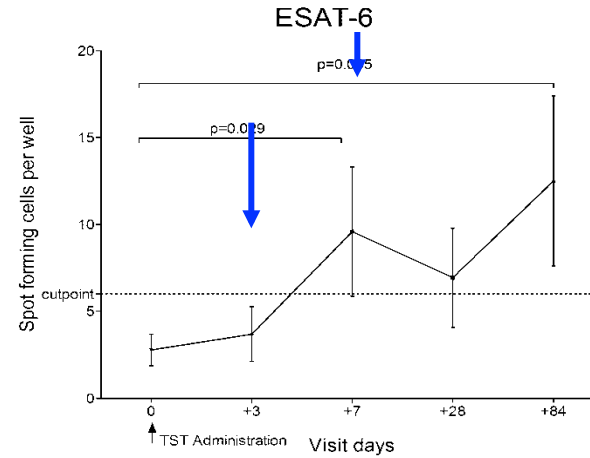
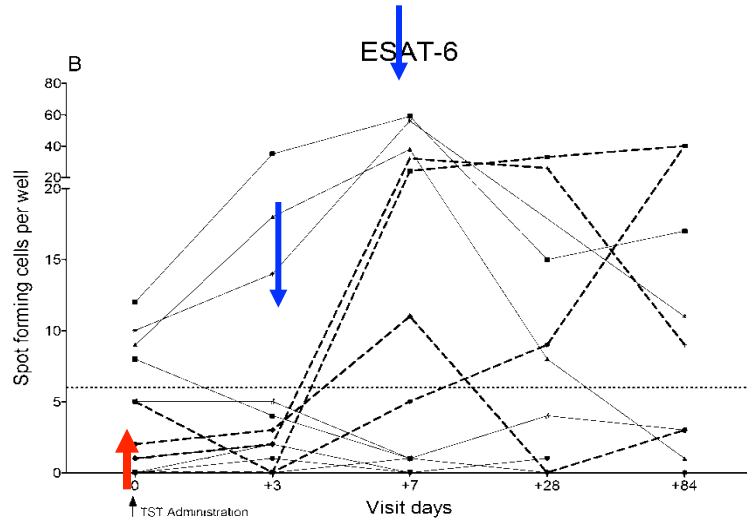
returning to:
Variability and TST mediated IGRA boosting



QFT-GIT boosting...



T-SPOT.[®]TB boosting



TST induced IGRA response boosting

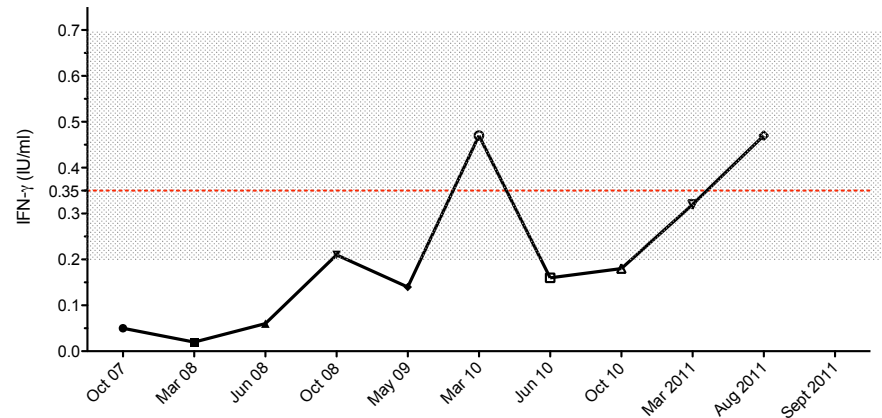
What does it mean to the clinician?

- Boosting occurs:
 - After day 3 post TST
 - In both T-SPOT.[®]TB and QFT-GIT assays
 - Predominantly in IGRA +ve subjects
 - Also occurs in IGRA negative subjects
 - Smaller but significant percentage (12.5%)

The saga continues...

- **Laboratorium onderzoek (Erasmus MC):**

- mei 2009: 0,14 INF IE/ml
- **mrt 2010: 0,47 INF IE/ml**
- Jun 2010: 0,16 INF IE/ml
- jul 2010: 0,16 INF IE/ml
- okt 2010: 0,18 INF IE/ml
- mrt 2011: 0,32 INF IE/ml
- **aug 2011: 0,47 INF IE/ml**



- Radiographic examination:
- CXR dated 09.27.2011: no abnormalities.
- **Conclusion:** re-conversion of the IGRA in a medical assistant working in a mycobacterial microbiological laboratory.
- In consultation with a second specialist it was decided that given the positive result a recent conversion could not be excluded **and to prescribe prophylaxis to the patient**

The saga goes on...

- The index patient decided to phone a friend...
- A microbiology registrar drew another blood sample: for QFT and TSPOT...
- **sept 2011: 0,25 INF IE/ml (T-spot.TB IGRA negatief)**
- She was going on holiday and would think about IPT.....
- On return from her holiday she refused IPT...

Occupational TB specialist advise:

To stop using IGRA and screen with CXR yearly...

In summary

- All tests are subject to variability...
- In diagnostic studies the performance characteristics of the test being evaluated need to be understood and defined.
- All factors affecting the assay/test performance need to be tightly controlled to generate “valid results”...

Questions and clarification...?