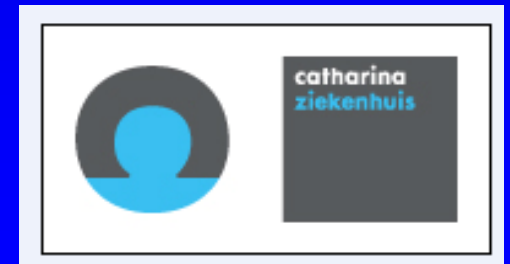*Imaging & Physiology Summit*

ON THE ACCURACY AND REPRODUCIBILTY OF PHYSIOLOGIC INDEXES

Seoul, Korea, december 6th, 2013

Nico H. J. Pijls, MD, PhD
Catharina Hospital,
Eindhoven, The Netherlands

Disclosures:

Dr Pijls received institutional research grants
from St Jude Medical and Maquet
and is consultant for St Jude Medical
and for Heartflow

Cutt-off or Threshold values *in clinical medicine are important for binary decision making:*

- ICD if EF < 35%
- Coronary Stent if FFR < 0.80
- Prehydration if MDRD < 60 ml/min
- Metformin if plasma glucose > 8.0 mmol/l

But…..nothing is perfect, mostly a gray zone exists, and sometimes a decision might be inappropriate.

The issue is how often that happens and how small a gray zone is.

➡ ***Accuracy, Reproducibility, Coefficient of Variation***

Cutt-off or Threshold values *in clinical medicine are important for binary decision making:*

- ICD if EF < *30-35%*
- Coronary Stent if FFR < 0.75-0.80
- Prehydration if MDRD < 45-60 ml/min
- Metformin if plasma glucose > 6.5-8.0 mmol/l

But…..nothing is perfect, mostly a gray zone exists, and sometimes a decision might be inappropriate.

The issue is how often that happens and how small a gray zone is.

➡ ***Accuracy, Reproducibility, Coefficient of Variation***

Focus on Physiologic Indexes in the Cath Lab

- The rationale of using a physiologic index in the catheterization laboratory, is *facilitating decision-making,* like placing a stent or not

- That is a *binary decision: yes or no; "go or no-go"*

- Therefore, a sharp cut-off or threshold value should exist for such index, with minimal gray zone

- Such index should be *accurate* and *reproducible*

- Finally, one threshold value should be applicable to all different types of populations

Accurate:

The decision, taken upon the measured value, should be the correct decision

Reproducible

At repeated measurements, values should be obtained leading to similar decision

→ *Coefficient of variation* (= SD / mean)

How to search for a threshold ?

In most studies:

*Analysis of **ROC curve** in a particular population and "cherry-picking" the best value*

(e.g. all resting indexes like Pd/Pa at rest, iFR, bSVR but also some hyperemic indexes like hSVR)

*Such studies are often called "prospective" but **in fact** are based upon a retrospective analysis of data (that is inherent to ROC analysis)*
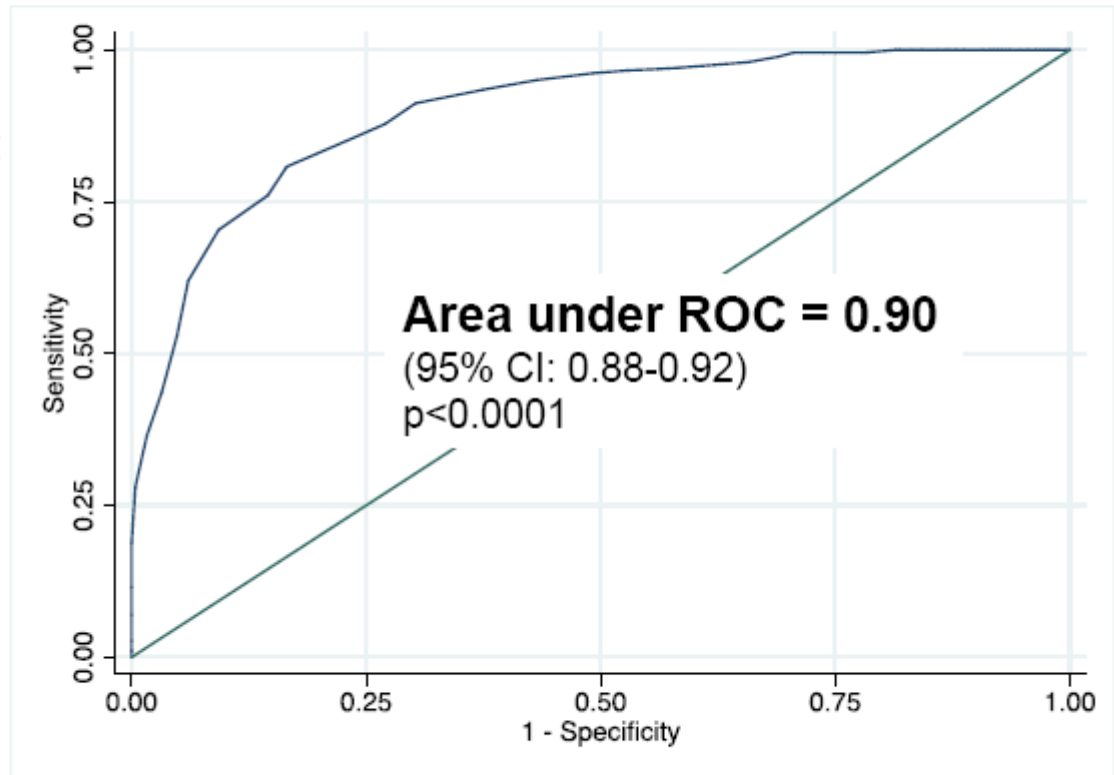
iFR

- Best iFR cut-off:
  ≤0.89

- Properly classified by iFR:
  82.46%

- Specificity:
  87.78%

- Sensitivity:
  72.98%

- Positive predictive value:
  77.02%

- Negative predictive value:
  85.27%

Area under ROC = 0.90
(95% CI: 0.88-0.92)
p<0.0001
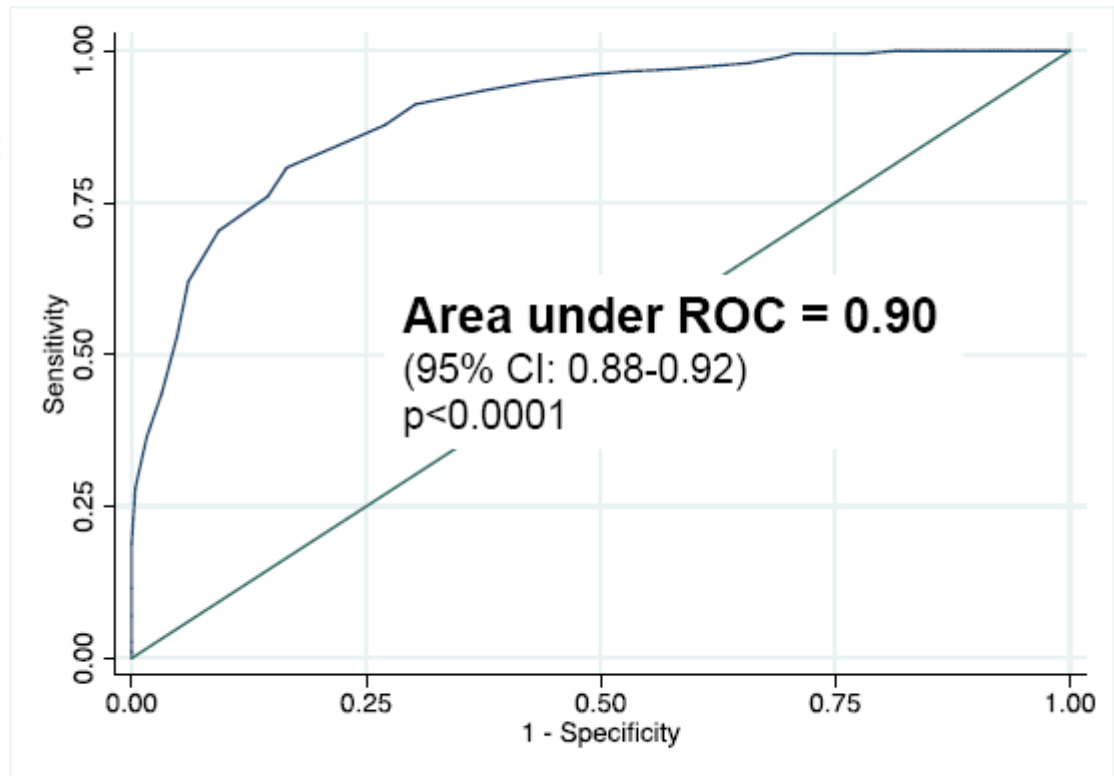
- *ROC analysis in Petraco et al, EuroIntervention 2013*
- *gold standard used is FFR*

iFR

- Best iFR cut-off:
  ≤0.89

- Properly classified by iFR:
  82.46%

- Specificity:
  87.78%

- Sensitivity:
  72.98%

- Positive predictive value:
  77.02%

- Negative predictive value:
  85.27%

Area under ROC = 0.90
(95% CI: 0.88-0.92)
p<0.0001

- *ROC analysis in Petraco et al, EuroIntervention 2013*
- *gold standard used is FFR*
- *best cut-off = 0.89*

iFR

- Best iFR cut-off:
  ≤0.89

- Properly classified by iFR:
  82.46%

- Specificity:
  87.78%

- Sensitivity:
  72.98%

- Positive predictive value:
  77.02%

- Negative predictive value:
  85.27%

Area under ROC = 0.90
(95% CI: 0.88-0.92)
p<0.0001
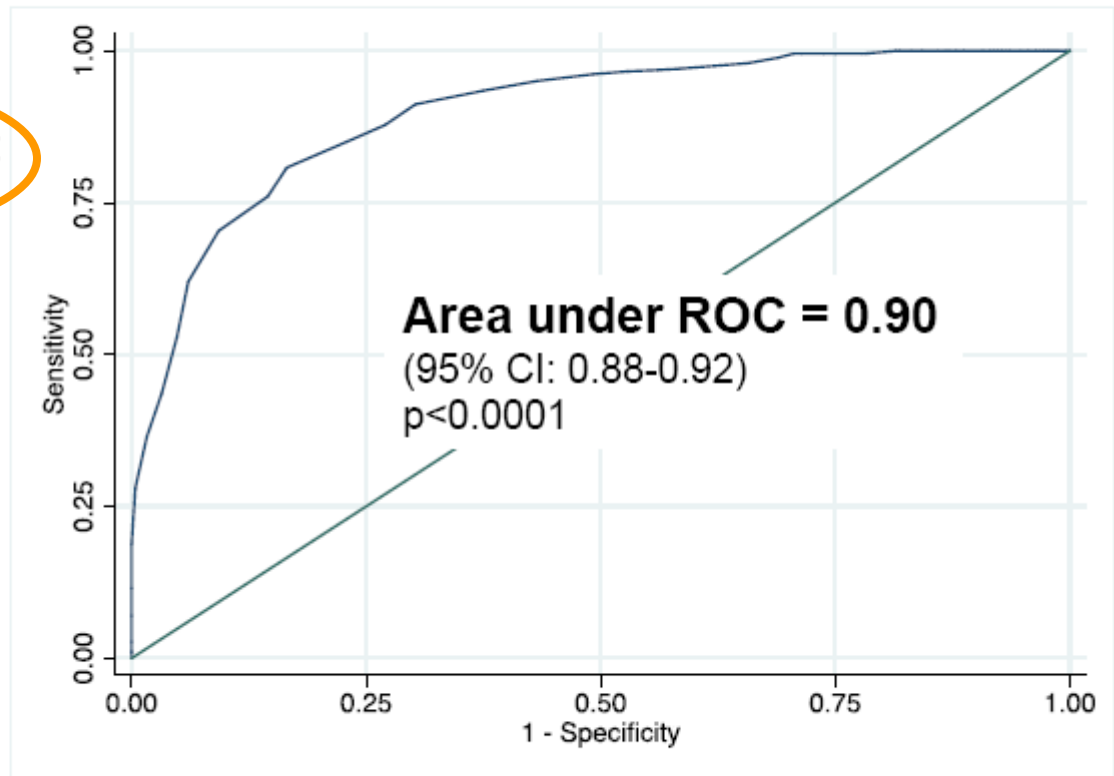
- *ROC analysis in Petraco et al, EuroIntervention 2013*
- *gold standard used is FFR*
- *accuracy = 82 %*

How to search for a threshold ?

In most studies:

Analysis of ROC curve in a particular population and "cherry-picking" the best value

(e.g. all resting indexes like Pd/Pa at rest, iFR, bSVR but also some hyperemic indexes like hSVR)

But……in another population, another ROC and another "best cut-off point" and "accuracy" will be found !!!

Author	Meeting or Citation	Date	N	iFR cutoff*
Davies	TCT	2011 November		none**
Sen	JACC 59:1392	2011 December	157	0.83
Park	EuroPCR	2012 May	238	0.89
Petraco	EuroIntervention	2012 August	339	0.89
Jeremias	TCT	2012 October	1548	0.90
Indolfi	TCT	2012 October	71	0.93
Johnson	JACC 61:1428	2013 February	1129	0.89
Sen	JACC 61:1409	2013 April	51	0.86

➡ *Value of iFR best corresponding to FFR of 0.80 varies from 0.83 -0.93*

How to search for a threshold that can be truly used as gold standard ?

The right way to go is a 2-step approach:

1. Exploration of range where a true threshold is expected:
   in a population in whom you can definitely conclude if there is disease or not

2. Truly prospective validation *of that particular threshold* in an arbitrary population, using a combined gold standard
   (prospective multitesting Bayesian approach; NEJM 1996; 334:1703-08)

How to search for a threshold that can be truly used as gold standard ?

The right way to go is a 2-step approach:

1. Exploration of range where a true threshold is expected:
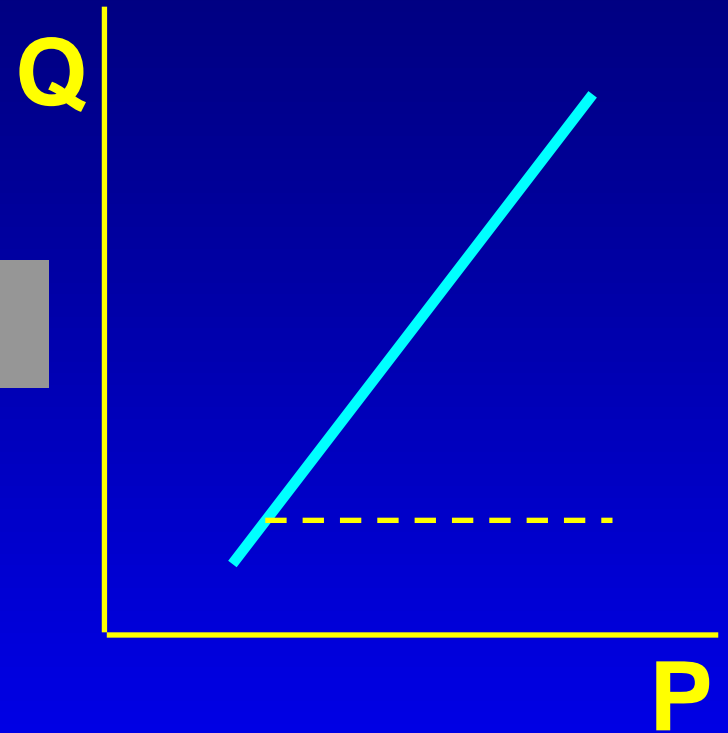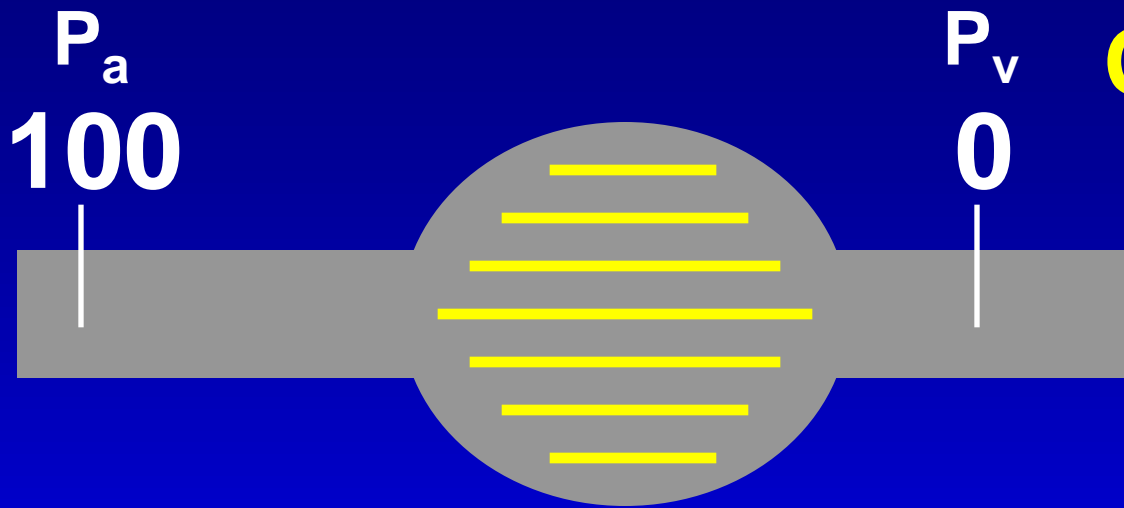   *in a population in whom you can definitely conclude if there is disease or not*

2. Truly prospective validation *of that particular threshold* in an arbitrary population, using a combined gold standard

   (prospective multitesting Bayesian approach; NEJM 1996; 334:1703-08)

→ **Fractional Flow reserve**

During Maximal Vasodilatation

P_a
100

P_v
0

Q

P_a
100

P_d
70

P_v
0

P
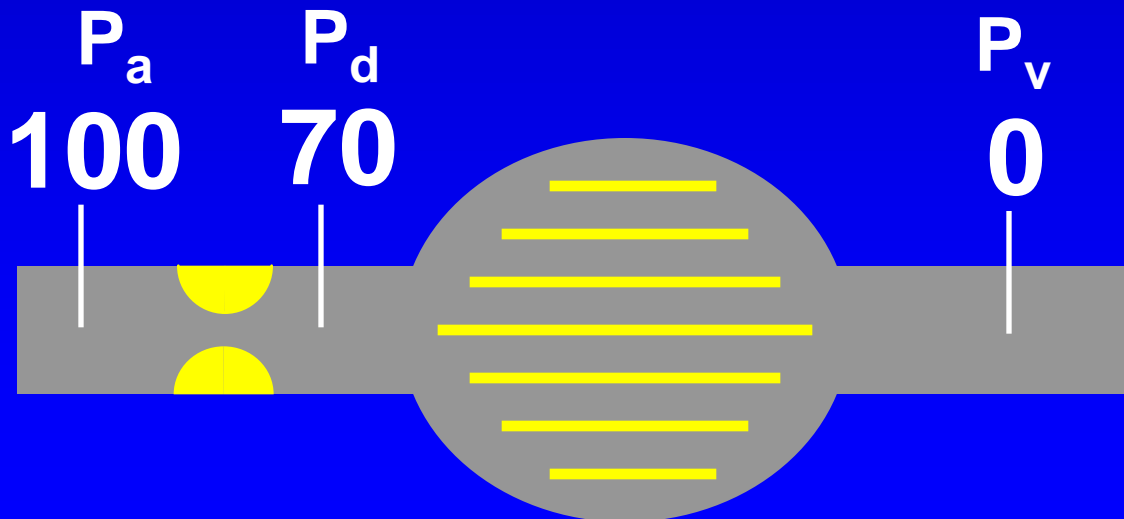
$$FFR_{myo} = \frac{P_d}{P_a}$$

$$= 0.70$$

Threshold value of FFR to detect significant stenosis in humans

FFR

non-signif.		stenosis significant

1.0   0.80   0.75   0

Pijls et al, N Engl J Med 1996; 334:1703-1708
Oldroyd et al, Circulation 2010

Threshold value of FFR to detect significant stenosis in humans

FFR

non-signif.		stenosis significant

1.0   0.80   0.75   0

Why did FFR become the gold standard ??!! How was that validated??
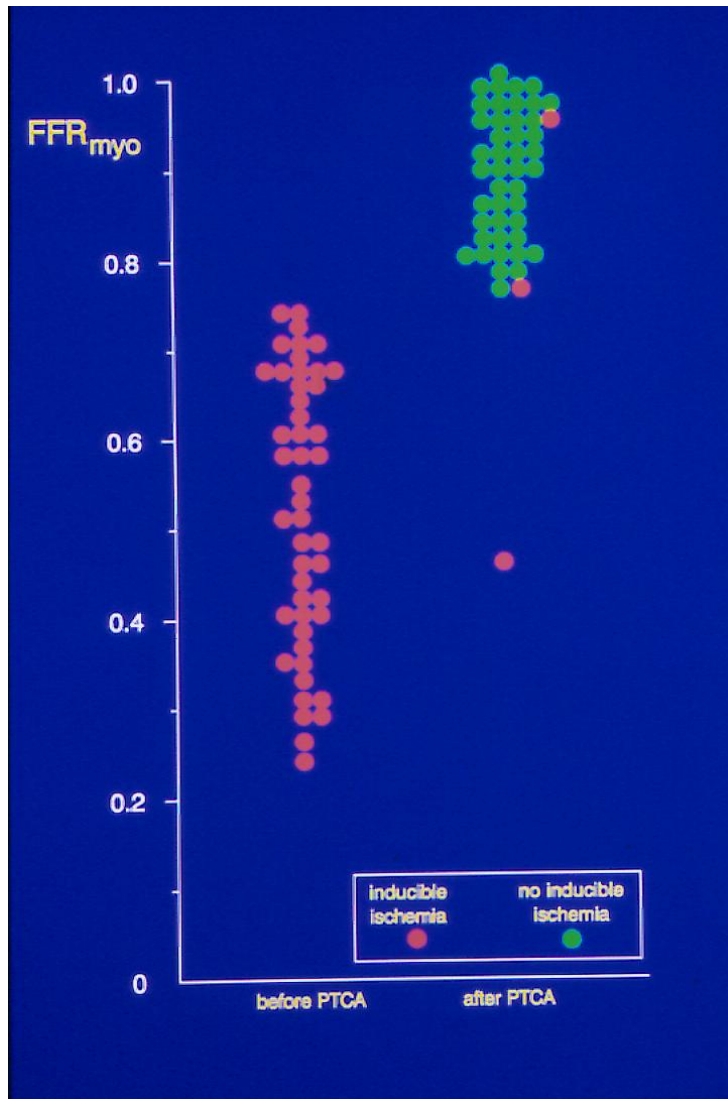
Pijls et al, N Engl J Med 1996; 334:1703-1708
Oldroyd et al, Circulation 2010

Validation of FFR in humans (step 1)
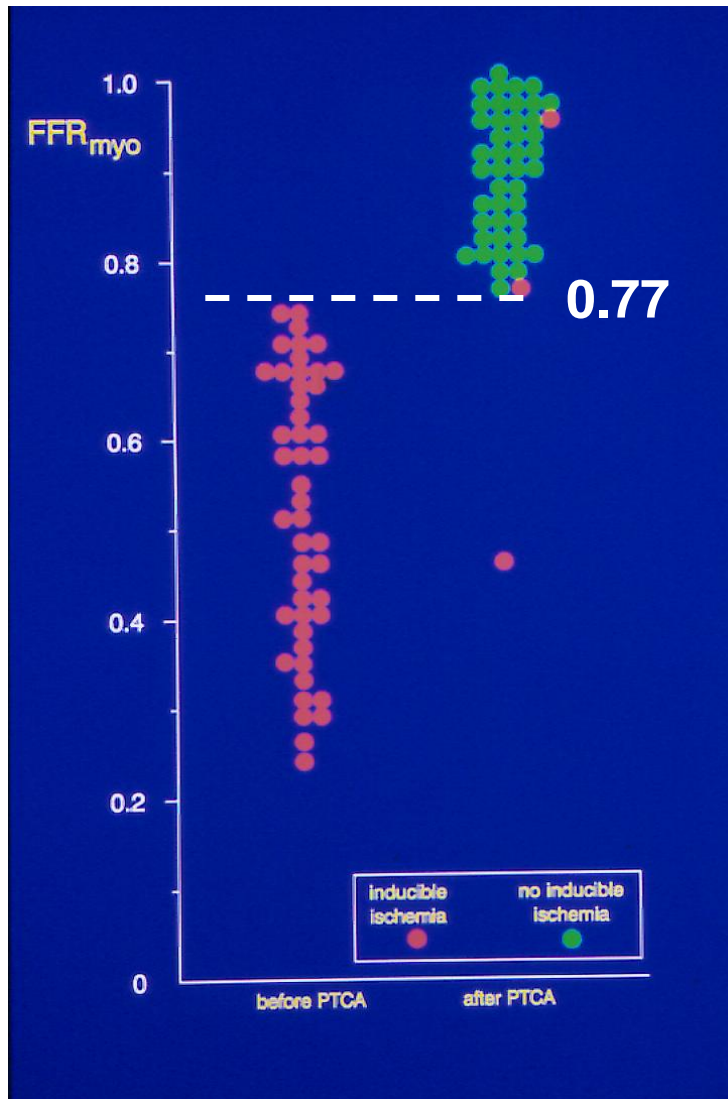

Proper validation of any index needs **2 steps:**

1. *Searching for the threshold value in a selected population (sens, specif, NPV, PPV, ROC analysis)*

2. Prospective validation in a population with unknown characteristics

Pijls et al, Circulation 1995
De Bruyne, Circulation 1996

Validation of FFR in humans (step 1)

Proper validation of any index needs
2 steps:

1. Searching for the threshold value in a selected population
   (sens, specif, NPV, PPV, ROC analysis)

2. Prospective validation in a population with unknown characteristics

Pijls et al, Circulation 1995
De Bruyne, Circulation 1996

Testing of FFR versus True Gold Standard

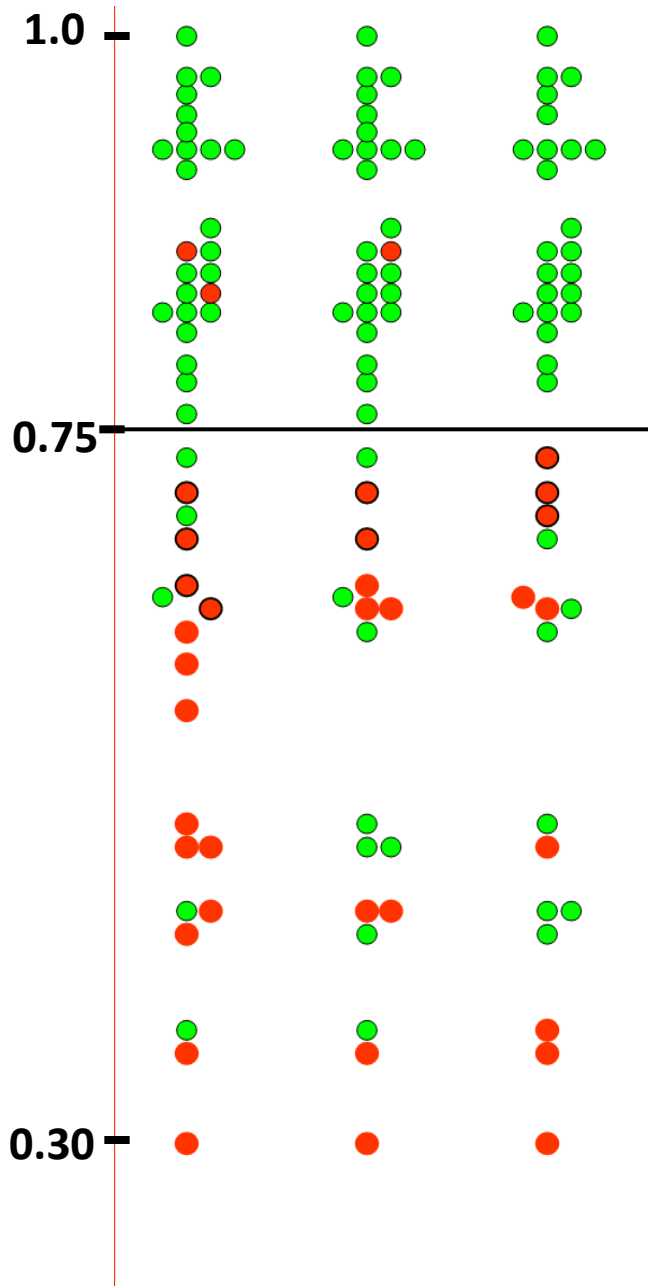Creating a gold standard by *Prospective Multitesting Sequential Bayesian Approach:*

- Exerc testing = electrical index of ischemia

- MIBISpect = perfusion index of ischemia

- Dobutrex Echo = contractile index of ischemia

- *reversal from positive before to negative after intervention, proves true positivity before and true negativity after test*

Diagnostic accuracy of FFR =

$$\left[(1\text{-}0.75) \times (1\text{-}0.8) \times (1\text{-}0.8)\right]^{-1} = 99\ \%$$

3 unclassifiable patients (no intervention)
⟶ worst case scenario for FFR ⟶ **93 %**

Pijls et al, NEJM 1996

Threshold value of FFR to detect significant stenosis in humans

FFR	non-signif.		stenosis significant

1.0   0.80   0.75   0

FFR is the *only* functional index which has ever been validated versus a **true gold standard**. *(Prospective multi-testing Bayesian methodology)*
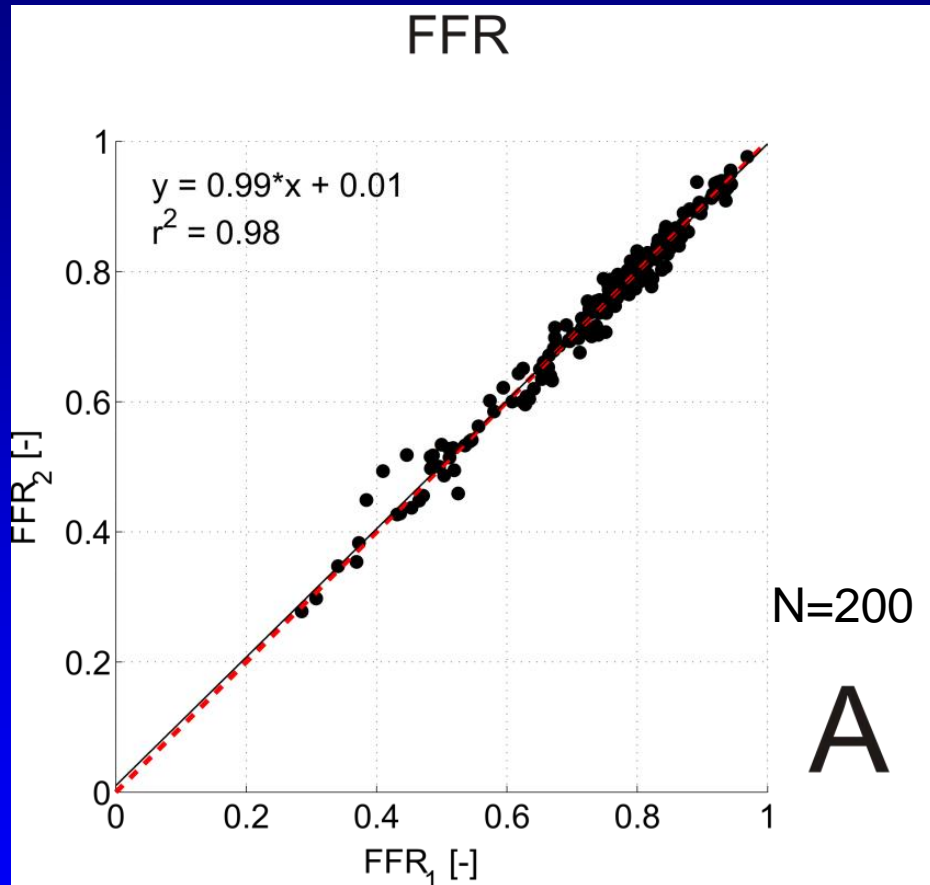
ALL studies ever performed in a wide variety of clinical & angiographic conditions, found threshold between 0.75 and 0.80

Diagnostic accuracy ≥93%

Pijls et al, N Engl J Med 1996; 334:1703-1708
Oldroyd et al, Circulation 2010

Reproducibility and Coefficient of variation

Reproducibility of FFR

(VERIFY study: all 205 consecutive arbitrary patients done in jan 2012)

FFR

$y = 0.99*x + 0.01$
$r^2 = 0.98$

N=200

A

VERIFY study, Berry et al, JACC 2013 (published februari 2013)
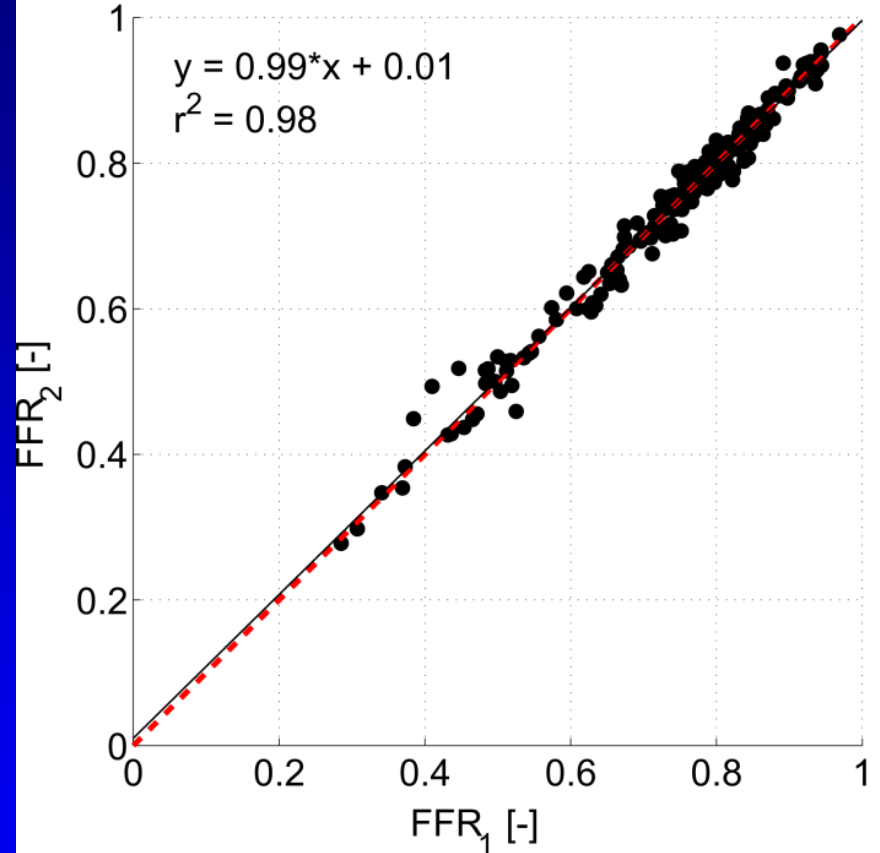
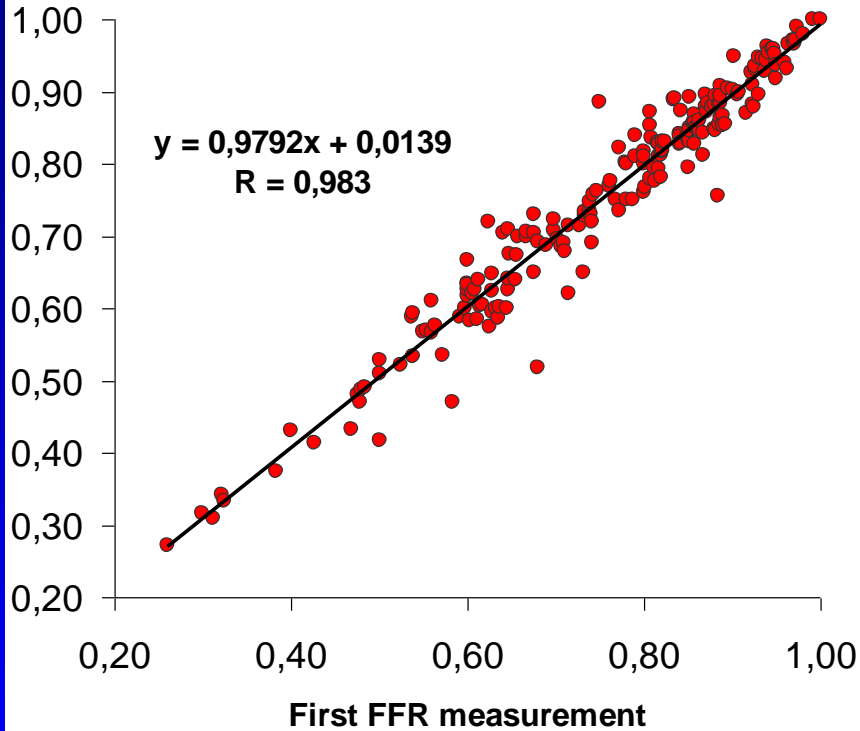*There is not any other index in physiology so reproducible as FFR*

Nevertheless, some doubt has been spread about reproducibily and variability of FFR, influencing its value for decision-making

(Petraco et al, EuroIntervention 2012 & JACC CVI 2013)

…….Let's have a closer look to these papers
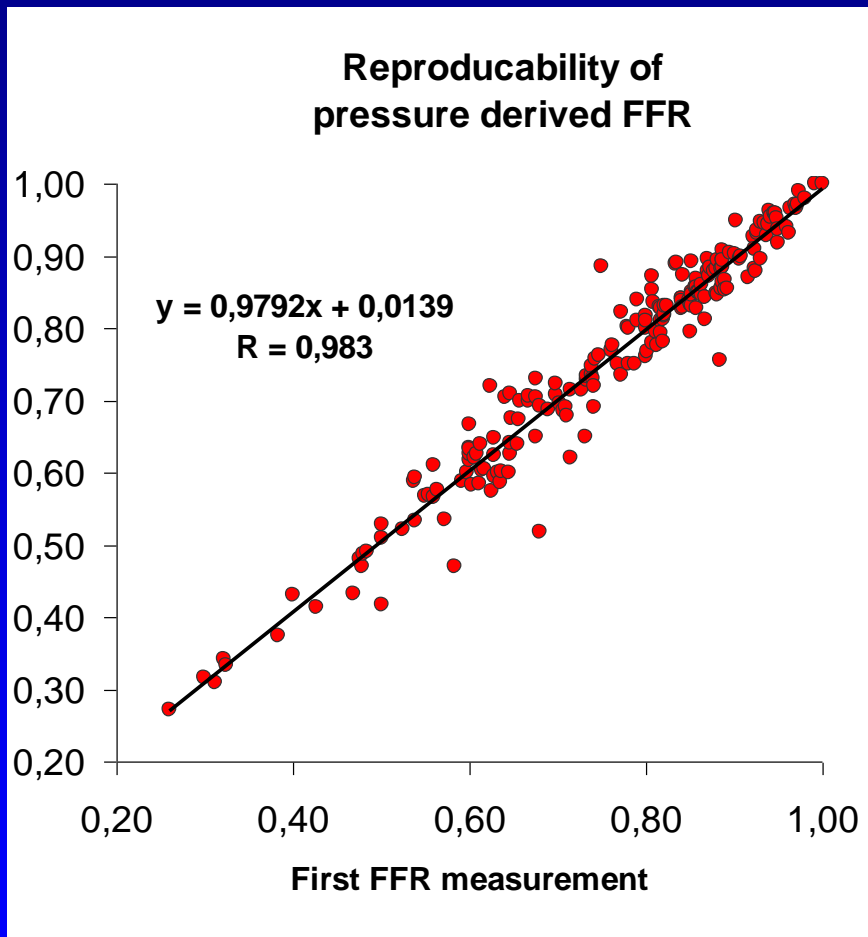
Reproducibilty of FFR

"Defer" 1996-1997

↑

(data used by Petraco)

Verify 2013

*Study by Petraco, Davies, and Escaned, published
in EuroIntervention and in JACC CardiovInterv (2012-2013)
to "demonstrate" limited reproducibility of FFR*

**Reproducability of
pressure derived FFR**

$$y = 0.9792x + 0.0139$$
$$R = 0.983$$

First FFR measurement

This figure is ***claimed*** to be taken
from DEFER study.

These are *scanned* data,
not from original database

Only *178 dots* counted, whereas
Defer study had 325 patients

Data with old *equipment*, almost
20 years old, from an era where
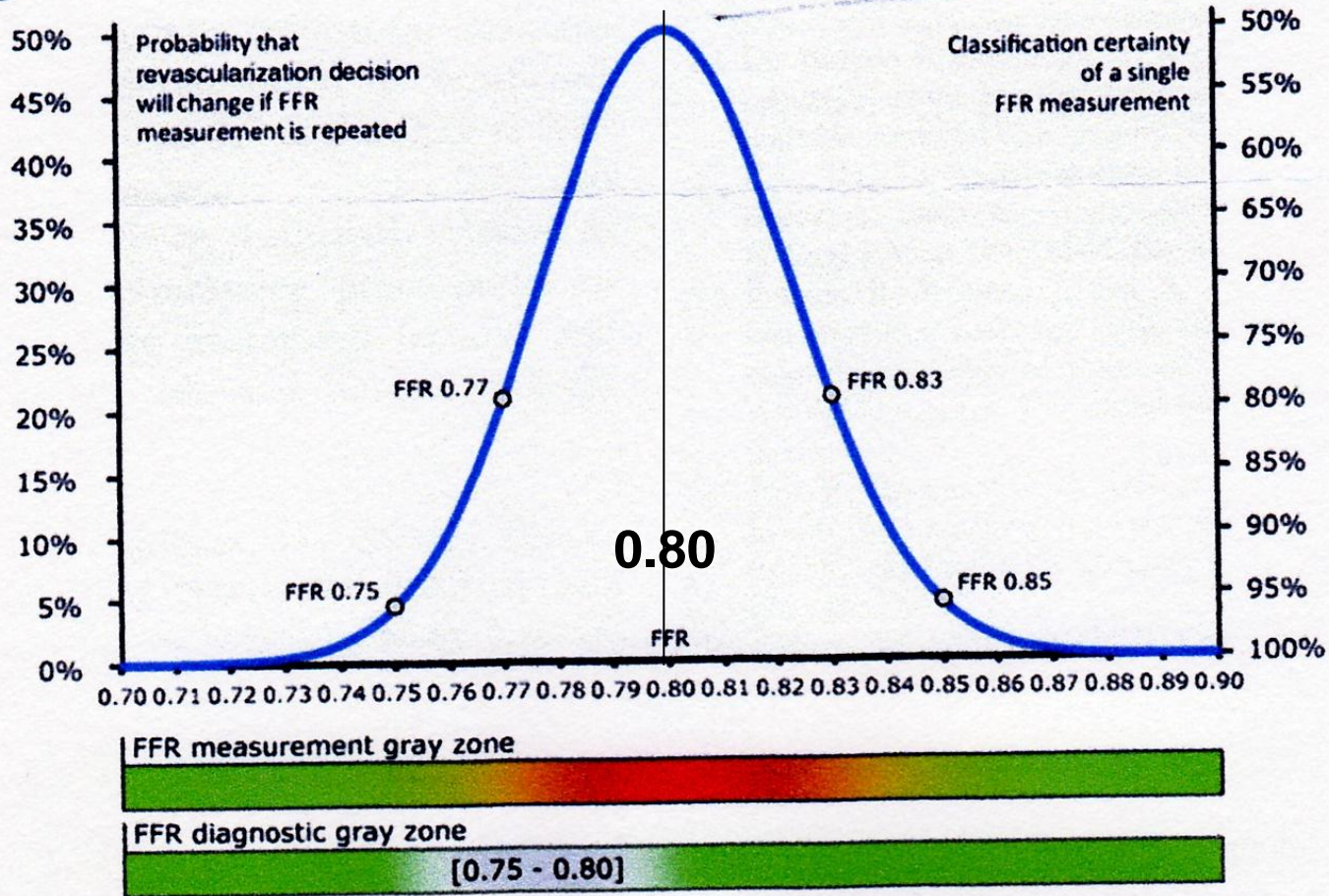FFR was seldom performed

Nevertheless, reproducibility is not bad at all for a clinical index !

Not only the data in these papers are disputable, also the interpretation is misleading:

Reproducability of pressure derived FFR

$y = 0,9792x + 0,0139$
$R = 0,983$

First FFR measurement
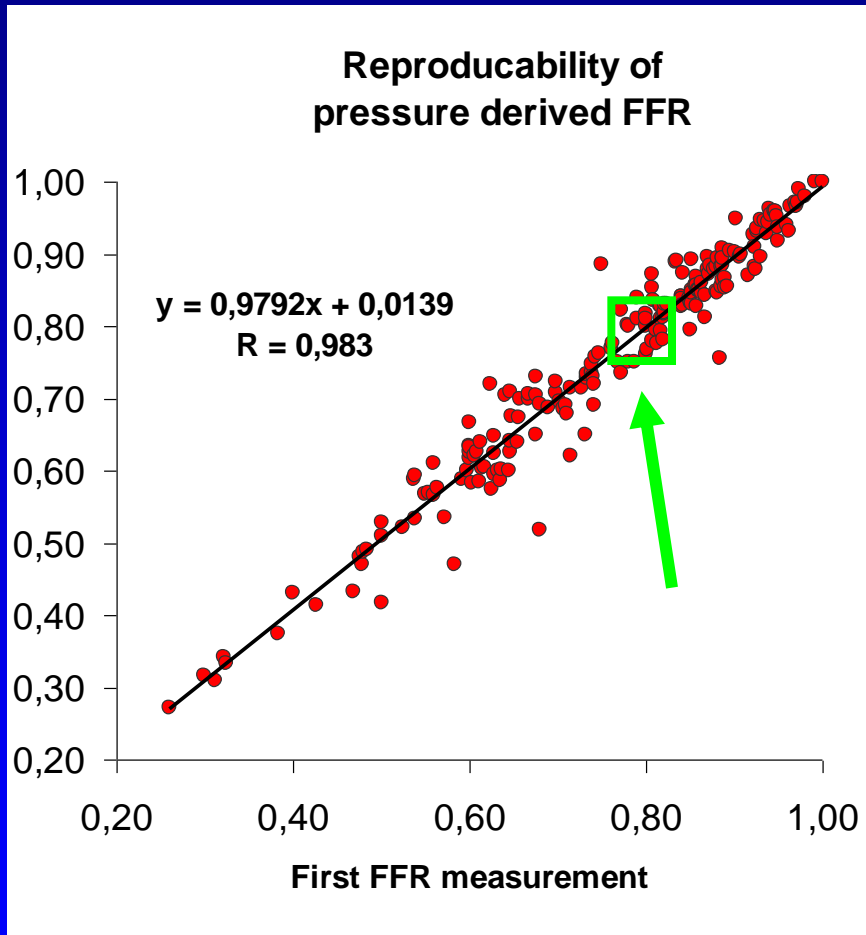
"Within the green box (patients with FFR close to the cut-off value), many jump from < 0.80 to > 0.80 at repeated measurement"

"therefore, especially close to the cut-off value, FFR is unreliable"

B

Probability that revascularization decision will change if FFR measurement is repeated

Classification certainty of a single FFR measurement

FFR 0.77 — FFR 0.83

0.80

FFR 0.75 — FFR 0.85

FFR

0.70 0.71 0.72 0.73 0.74 0.75 0.76 0.77 0.78 0.79 0.80 0.81 0.82 0.83 0.84 0.85 0.86 0.87 0.88 0.89 0.90

FFR measurement gray zone

FFR diagnostic gray zone

[0.75 - 0.80]

such curve (with a "peak" of 50%) exist by definition for any index, any cut-off value, and any reproducibility or coefficient of variation.

The coeff of variation determines the width of the "Gaussian"curve

For *ANY* index and *EVERY* cut-off value, *irrespective how accurate and how small its variabilty*, exactly 50% of all patients will cross the cut-off value at repeated measurement if the sub-population is studied *within one "coefficient-of-variation interval" from the cut-off value*

This has nothing to do with FFR, but is a general statistical law.

(cf: In a normal distribution, 68% of all points is within 1 SD of the mean)

Figure scanned from Petraco et al, JACC 2013

FFR, "DEFER study" *(Bech, Circulation 2001)*

FFR, VERIFY study *(Berry, JACC 2013)*

Probability that treatment decision will change if the respective index measurement is repeated
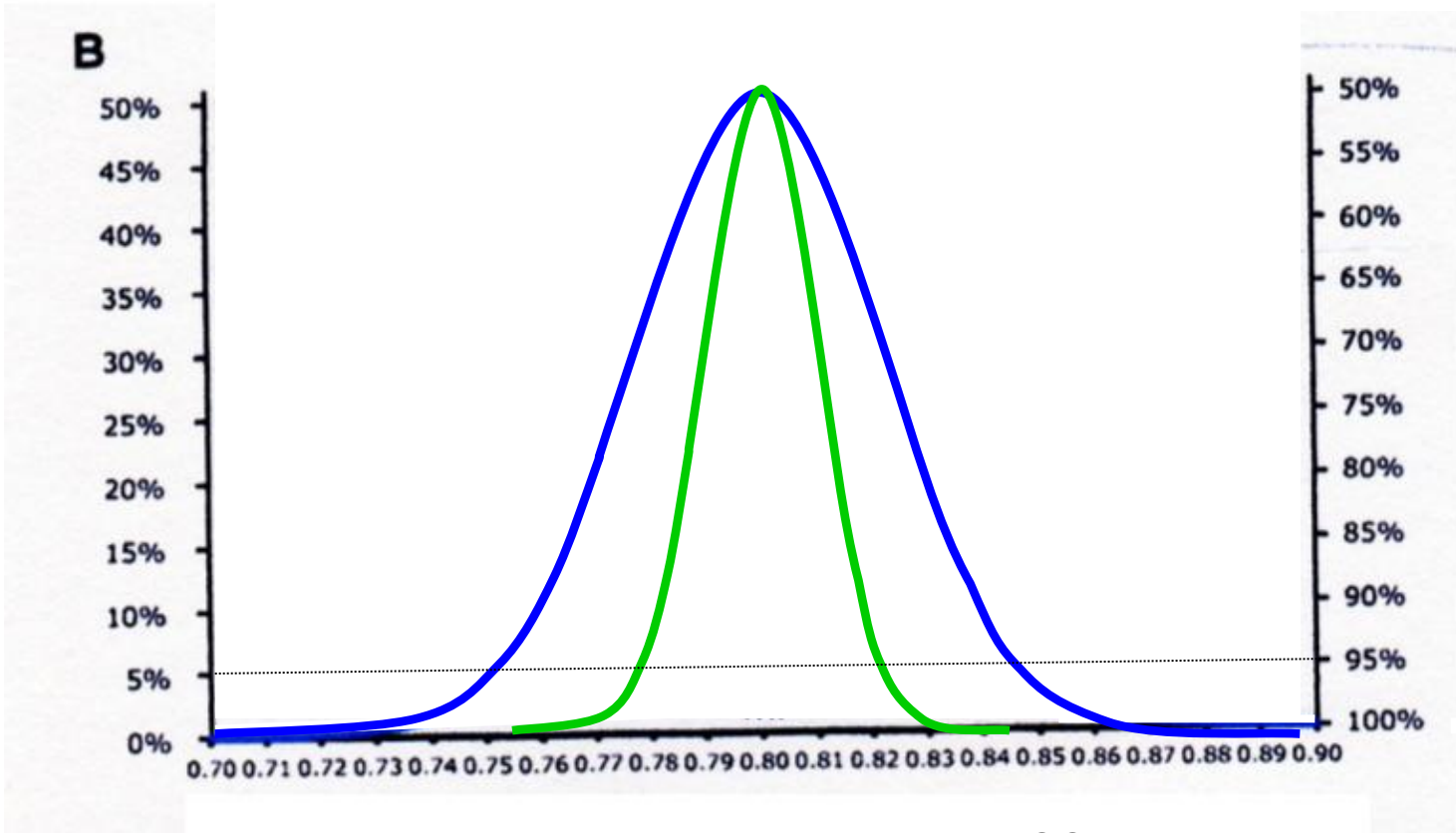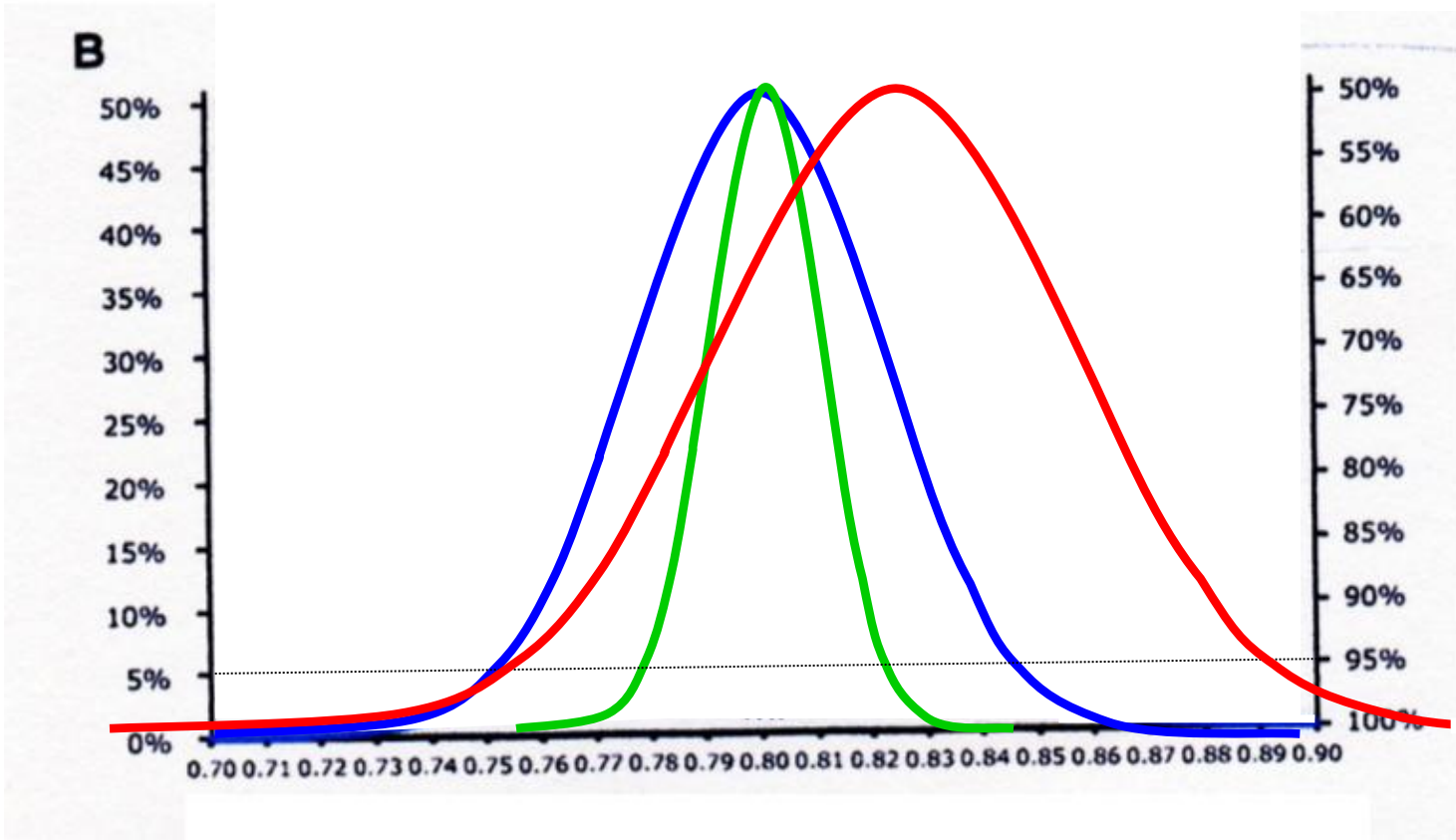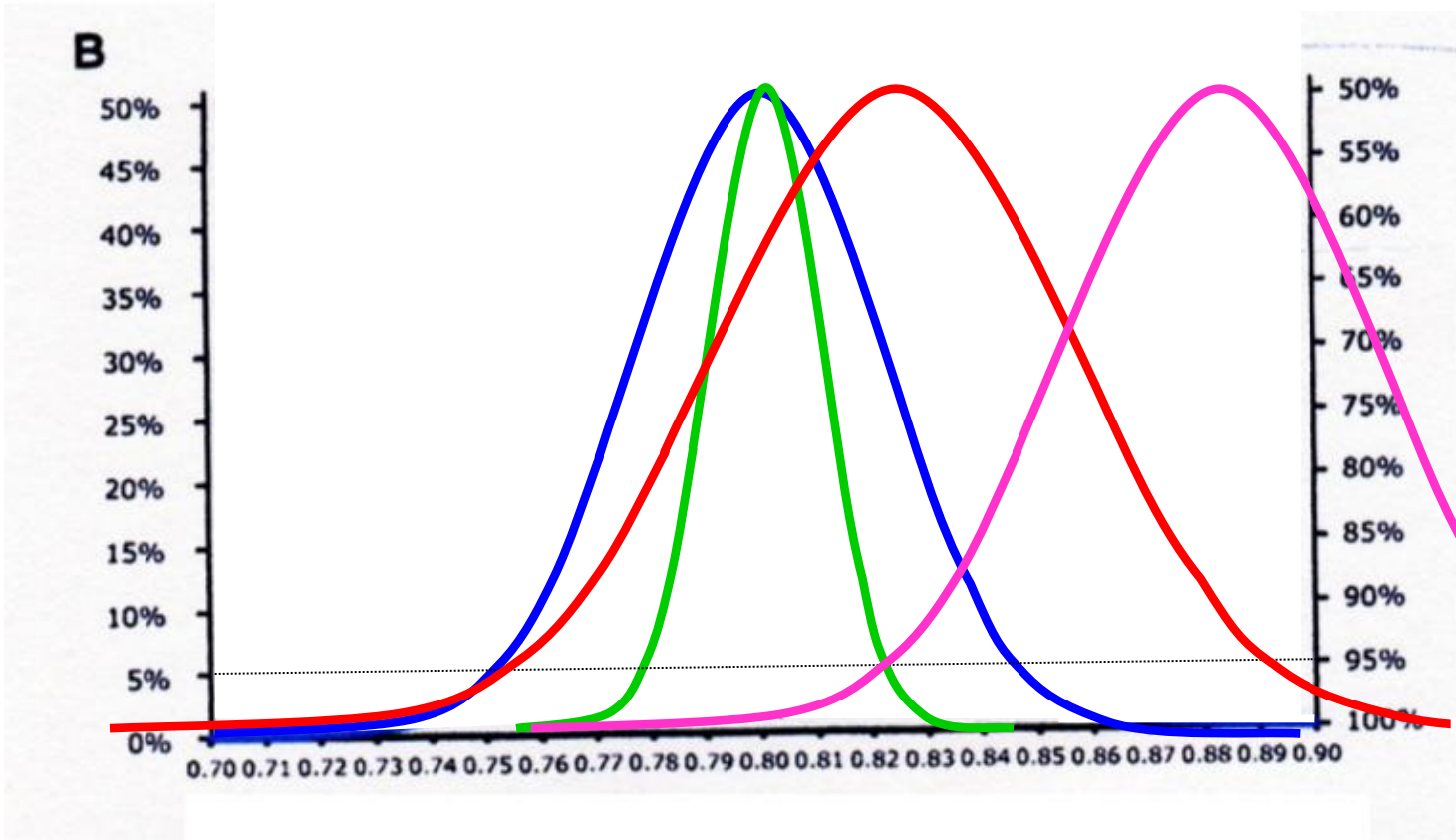
Classification certainty of single measurement

FFR, "DEFER study"

FFR, VERIFY study

iFR, ADVISE study

Probability that treatment decision will change if the respective index measurement is repeated

Classification certainty of single measurement

FFR, "DEFER study"

FFR, VERIFY study

iFR, ADVISE study

iFR, VERIFY study

Coefficients of variation for some frequently used clinical indices:

Fractional Flow Reserve	**2.4 %**	*Berry, JACC 2013; 61: 1421*
Fasting plasma glucose	9 %	*Mooy, Diabetologia 1996; 39:298*
ambulatory systol blood pr	11 %	*Eguchi, J Hypertension 2010;28:918*
Ejection Fraction by MRI	12 %	*Grothues, Amer J Card 2002; 90:29*
Percent stenosis by QCA	17 %	*Reiber, Circulation 1985; 71: 280*
C-Reactive Protein	46 %	*Bower, Arch Intern Med 2012; 172:1519*

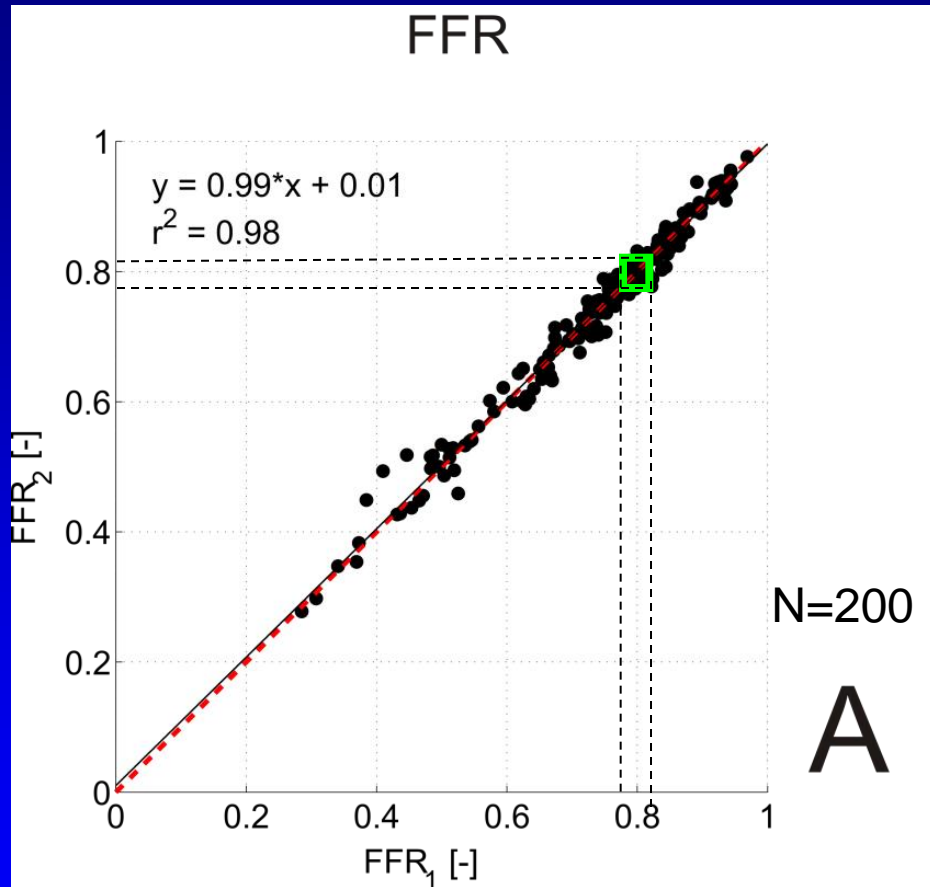**Probability that treatment decision will change if the respective index measurement is repeated**

Classification certainty of single measurement

FFR < 0.75

FFR > 0.80

0.75

0.80
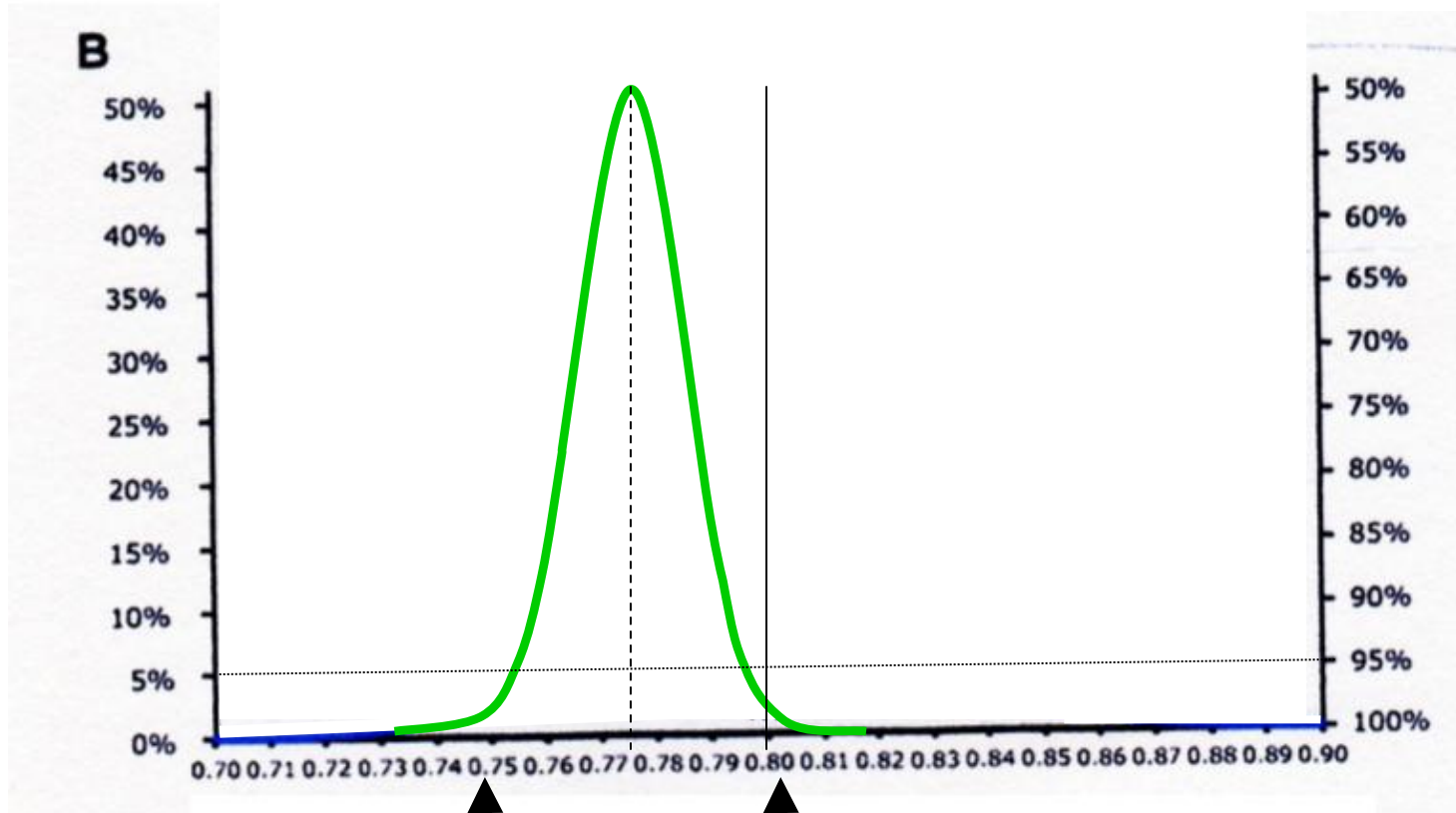
FFR, VERIFY study

2.4 % of patients go from green to gray or v.v. and 2.4 % from red to gray Almost nobody ever crosses from red to green or v.v.
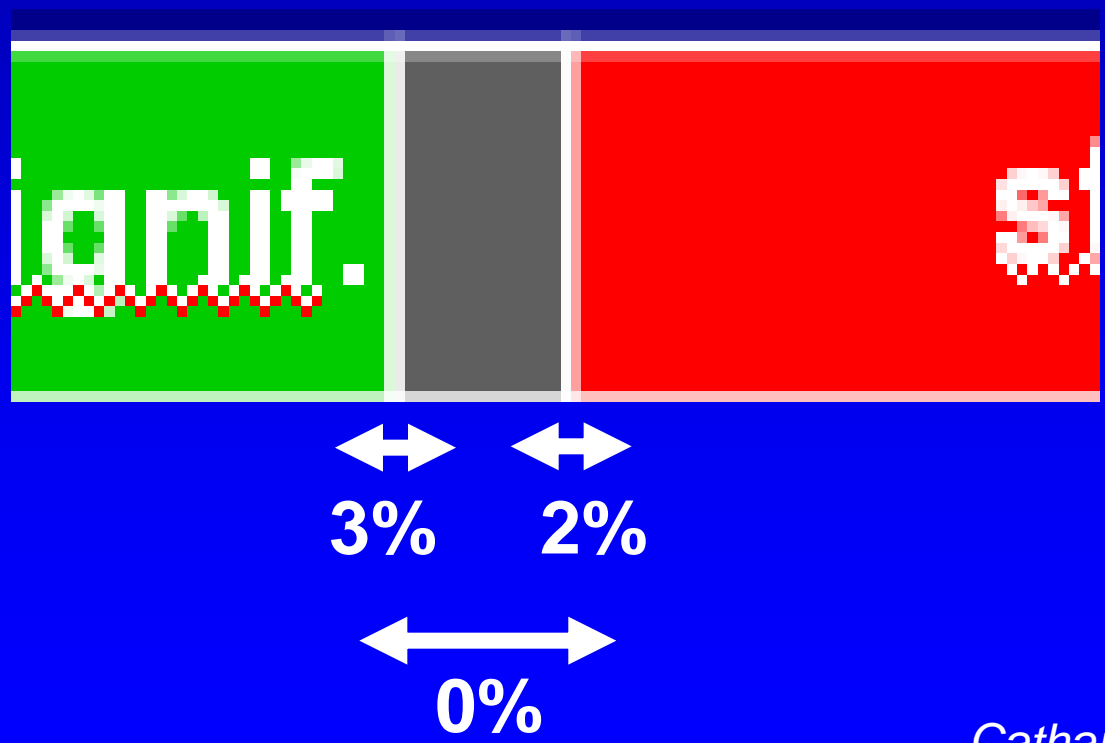
Reproducibility of FFR

FFR

$y = 0.99 \cdot x + 0.01$
$r^2 = 0.98$

N=200

A

FFR$_1$ [-]

FFR$_2$ [-]

2 x 2.4 % of patients within the green box

VERIFY study, Berry et al, JACC 2013 (published februari 2013)

There is not any other index in physiology so reproducible as FFR

At <u>1200</u> consecutive in-duplo measurements of FFR, there was <u>NOT ANY cross-over</u> across the gray zone

FFR | non-signif. | | stenosis significant

1.0   0.80   0.75   0

↔ 3%   ↔ 2%

↔ 0%

CONCLUSIONS:

- when interpreting (studies to) the accuracy & reproducibility of (physiologic) indexes used in the catherization lab, some critical attitude and understanding of statistics is mandatory

- simple ROC analysis is insufficient to validate any index. A two-step Bayesian approach is mandatory

- So far, such approach has only been applied to FFR

- Therefore, it is justified to use FFR as gold standard