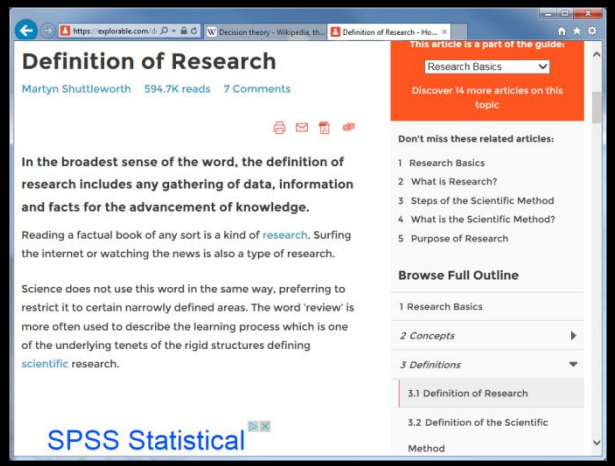
The Science of Quality Assurance Indicators & Technique Analysis

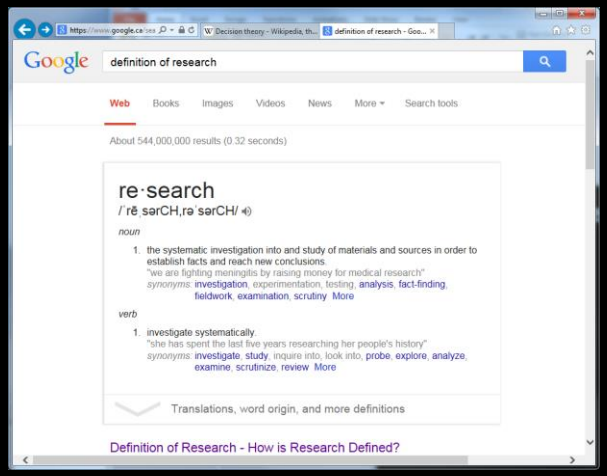
Marco Carlone, PhD, P.Eng, MCCPM
Medical Physicist, Princess Margaret Cancer Centre
Assistant Professor, University of Toronto

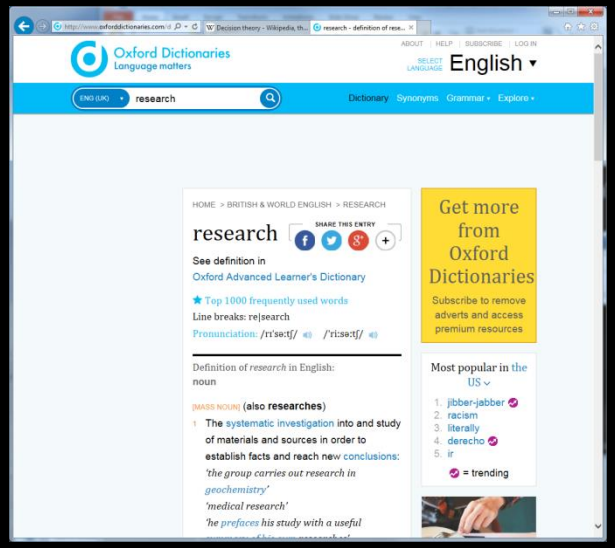
Radiation Oncology
UNIVERSITY OF TORONTO

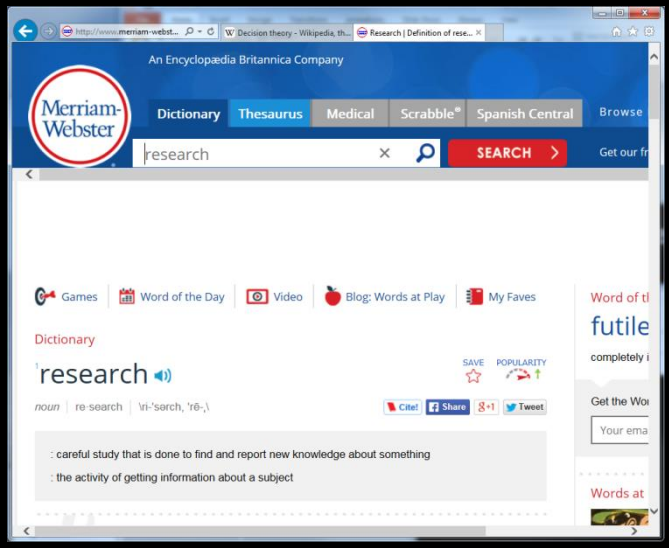
UHN Princess Margaret Cancer Centre

What is research?









Quality Improvement vs. Research

Quality Improvement vs. Research

Quality Improvements

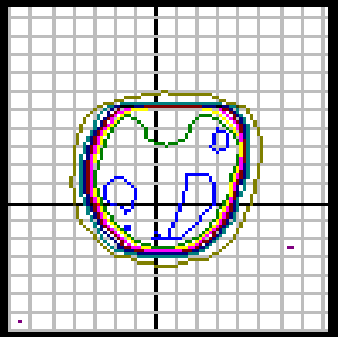
- Intent is to ***improve current practice***. For internal use only.
- Action is ***within existing standards of care***.
  - The knowledge is the same, but we can apply it in a better manner.

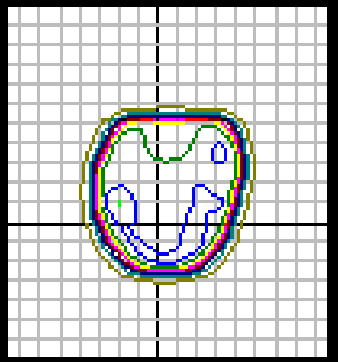
Research

- Intended to create ***generalized knowledge***.
- Desire to publish or present.
  - Data must be relevant *outside the institution*.
- Testing new methods.
  - *Perhaps new standards of care?*

What is Patient Specific QA?

- **Fundamentally, this is a beam diagnosis test.**
- **The purpose is to identify radiation beams that are “different” than the planned radiation beam.**
- **Most methods used to evaluate the result are physics & clinically based:**
  - **Dose & distance to agreement**
  - **Plan objectives, DVH, etc**
- **Are the usual metrics *generalizable*?**





Dosimetry vs. Imaging

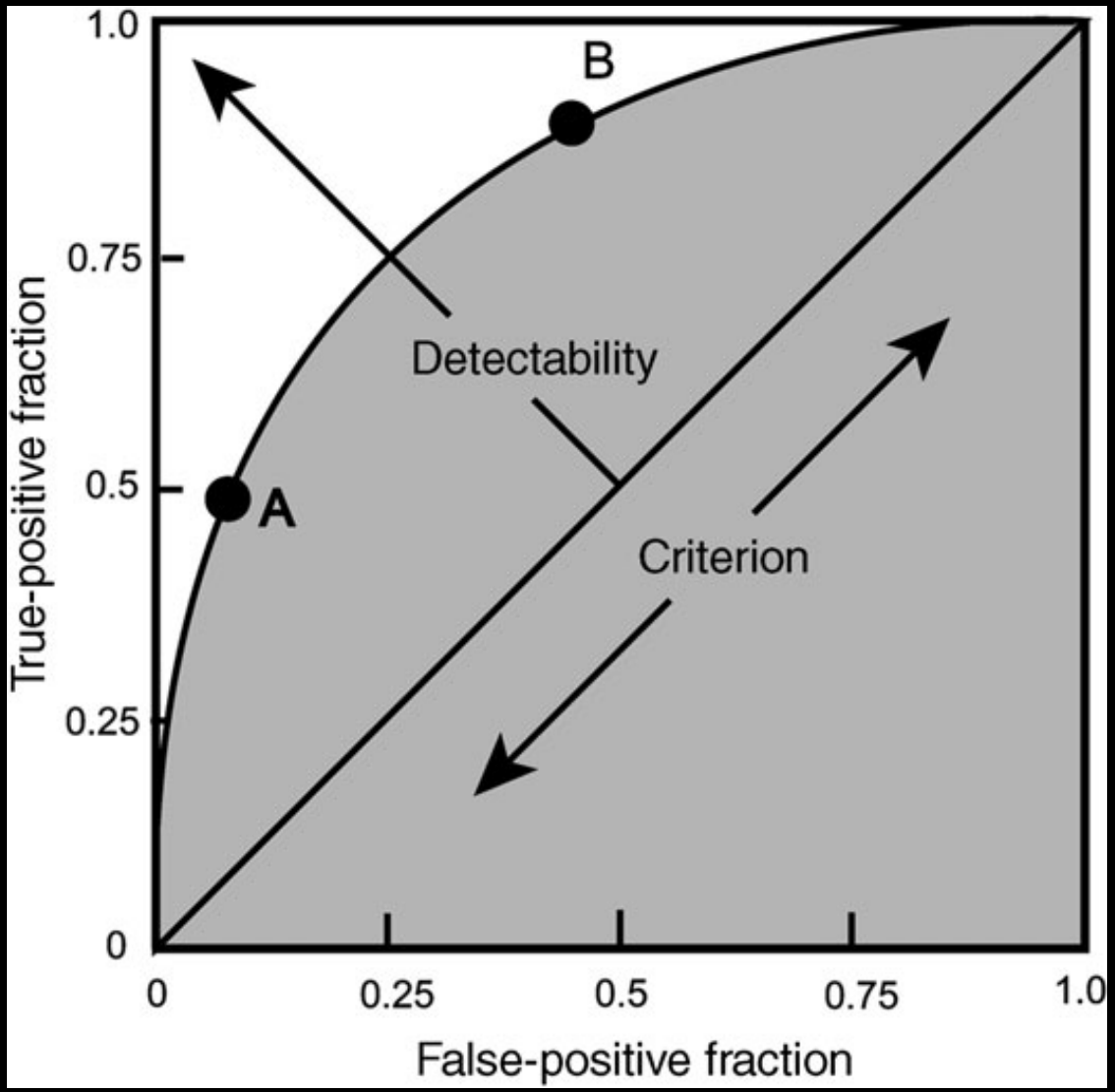
- **Patient specific QA has been viewed as a dosimetry problem.**
- **From dose point of view:**
  - **How different is the measured dose from the planned dose?**
  - **Detectors, spatial resolution, etc.**
- **From image analysis point of view:**
  - **Is the measured beam fluence abnormal?**

From a statistics point of view

- **Statistical tools exist to measure how different two (1-dimensional) distributions are.**
- **Extended to 2-dimensions, this problem becomes very difficult to solve explicitly.**
- **Measures of dose difference at a point and distance to agreement:**
  - **indicate conformance,**
  - **difficult to generalize**

Decision Theory

- **The ROC Decision model**
  - **Receiver Operator Curve (ROC) is a plot of performance of a binary classifier system.**
- **Graphical tool allows quantification of best cut-off point.**
- **Also offers insight into where gains in sensitivity and specificity can be obtained.**


1.0
0.75
0.5
0.25
0
True-positive fraction
B
A
Detectability
Criterion
0
0.25
0.5
0.75
1.0
False-positive fraction

Generalizing IMRT QA results

- **Evidence that IMRT/VMAT QA results don't translate well center to center.**
- **Letourneau, McNiven & Jaffray (IJROBP 2013)**
  - **Multi-institutional evaluation of IMRT/VMAT QA.**
  - **Results depend on the performance of many variables.**
  - **Variables interact with each other in ways that can be non-intuitive.**
- **Different centers may report IMRT/VMAT QA results that are similar, but the performance of their systems may be different.**
  - **Leaf calibrations, beam models, etc.**

ROC analysis in patient specific quality assurance

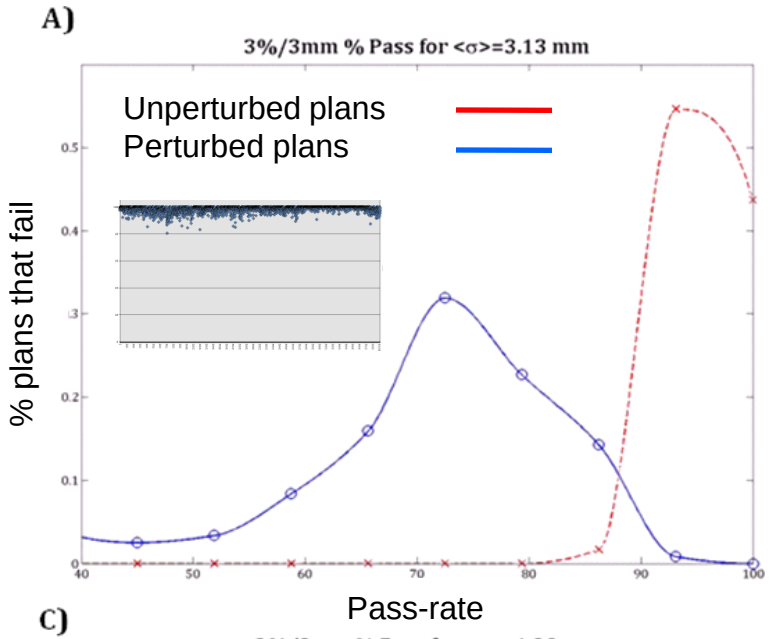
Marco Carlone[a)]
Department of Medical Physics, Trillium Health Partners, Mississauga, Ontario L5M 2N1, Canada; Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada; and Department of Radiation Oncology, University of Toronto, Toronto, Ontario M5S 3S2, Canada

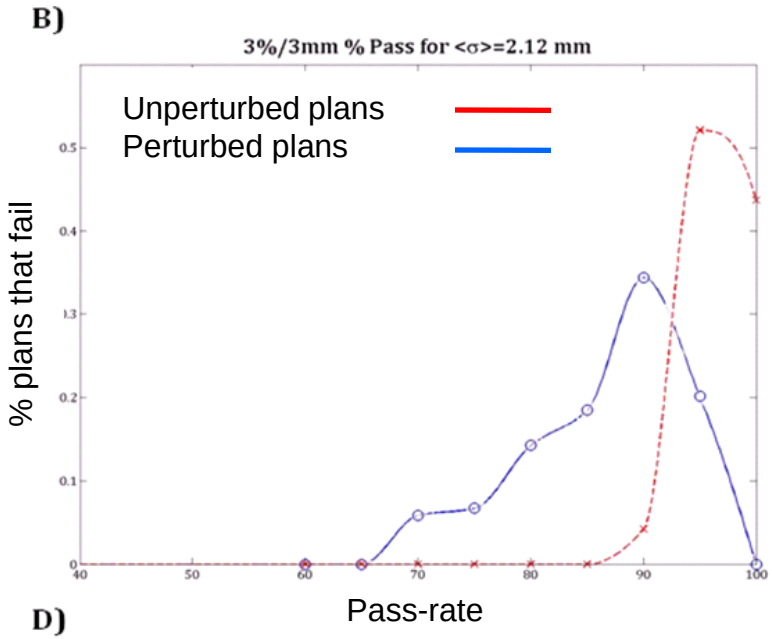
Charmainne Cruje, Alejandra Rangel, Ryan McCabe, and Michelle Nielsen
Department of Medical Physics, Trillium Health Partners, Mississauga, Ontario L5M 2N1, Canada

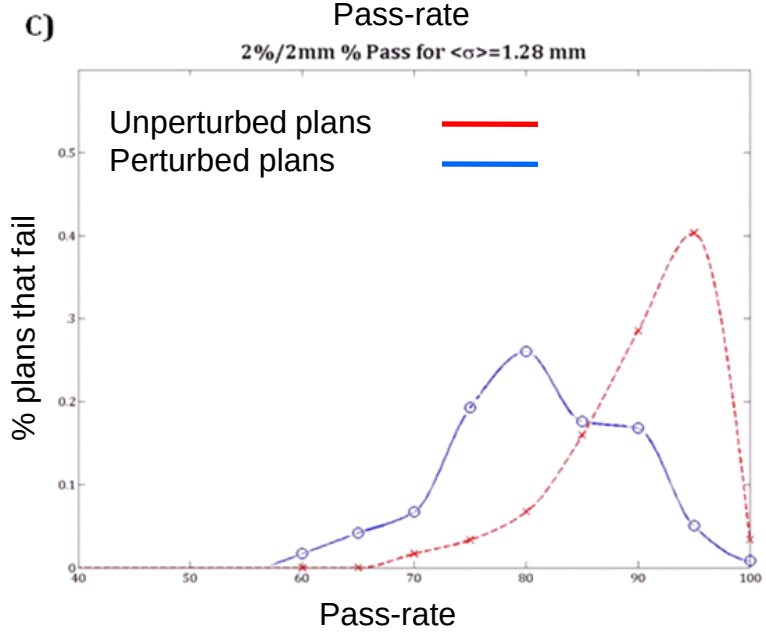
Miller MacPherson
Department of Medical Physics, Trillium Health Partners, Mississauga, Ontario L5M 2N1, Canada; Radiation Medicine Program, Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada; and Department of Radiation Oncology, University of Toronto, Toronto, Ontario M5S 3S2, Canada

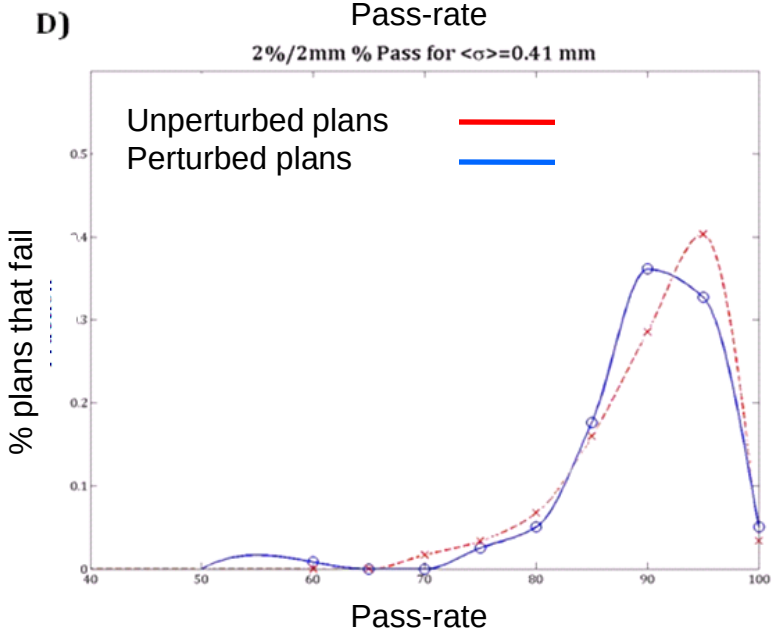
(Received 26 September 2012; revised 1 March 2013; accepted for publication 1 March 2013; published 2 April 2013)

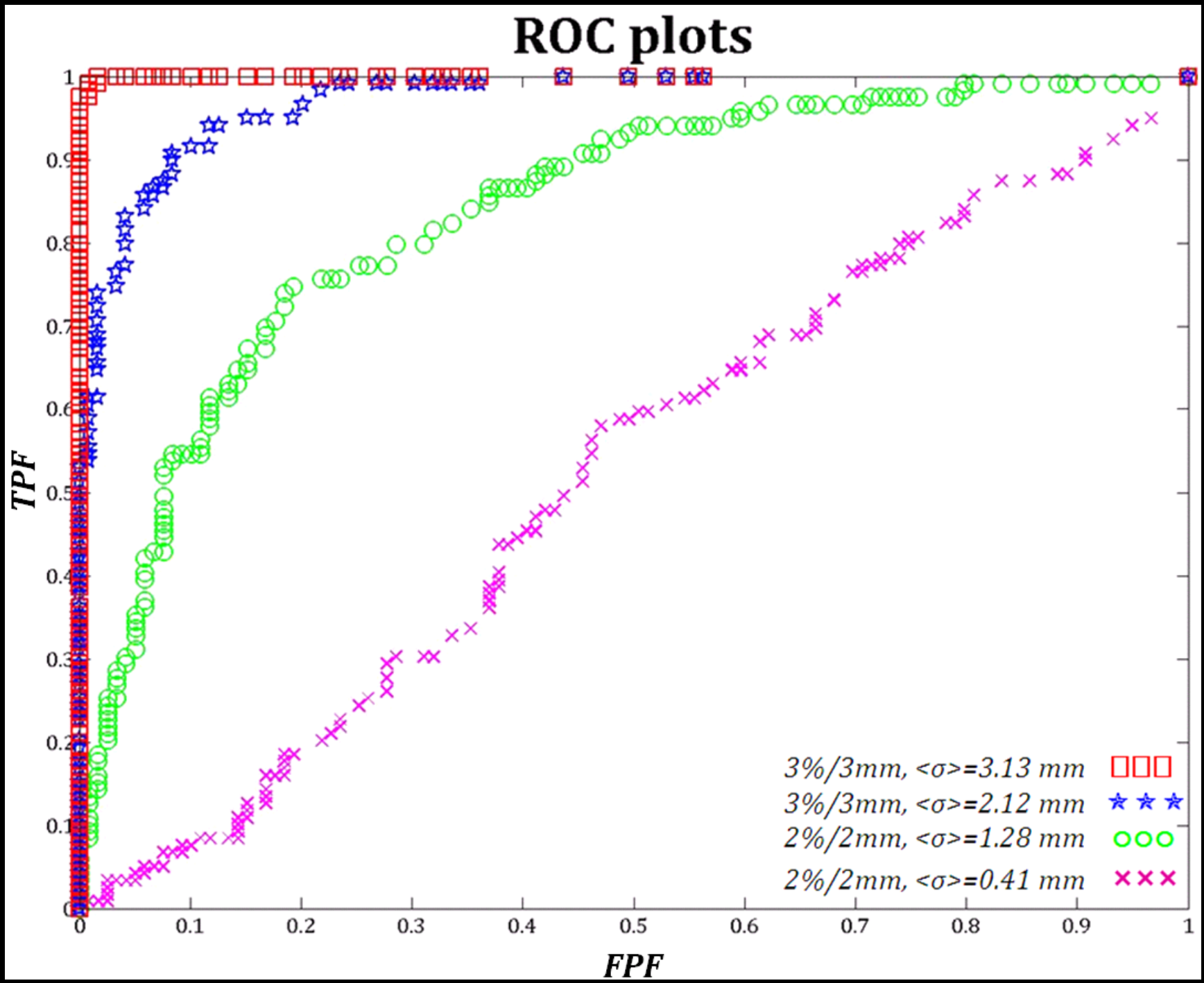
- 34 patients (single phase prostate cases)
- Half delivered normally.
- Other half delivered with known MLC errors
  - 1 mm, 2 mm, 3 mm
- Assume that the unperturbed delivery was “true” fluence pattern.
- Depending on passing criterion, measure rates of TP, TN, FP, FN.


A)
3%/3mm % Pass for <σ>=3.13 mm
Unperturbed plans
Perturbed plans
% plans that fail
Pass-rate

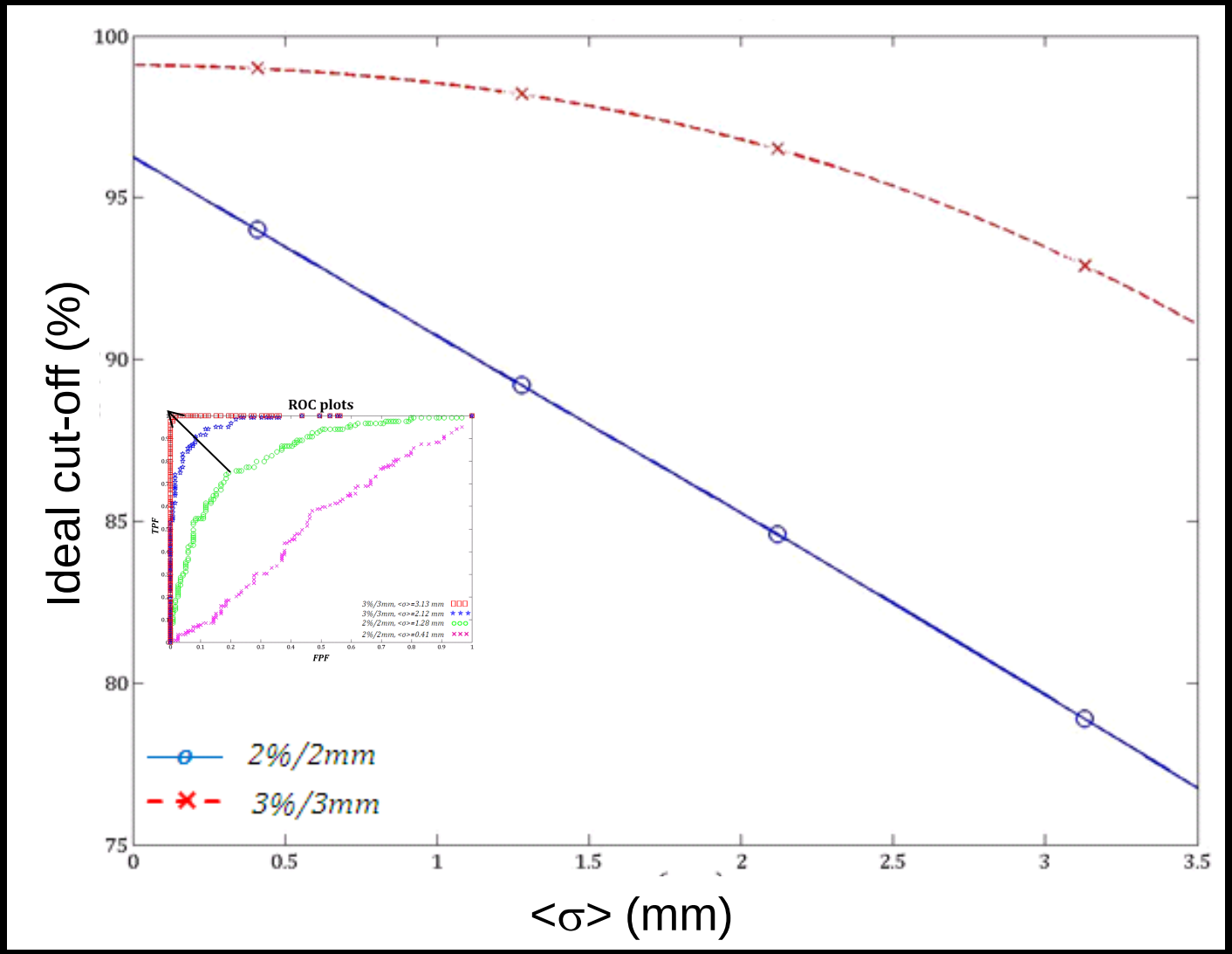

B)
3%/3mm % Pass for <σ>=2.12 mm
Unperturbed plans
Perturbed plans
% plans that fail
Pass-rate


C)
2%/2mm % Pass for <σ>=1.28 mm
Unperturbed plans
Perturbed plans
% plans that fail
Pass-rate

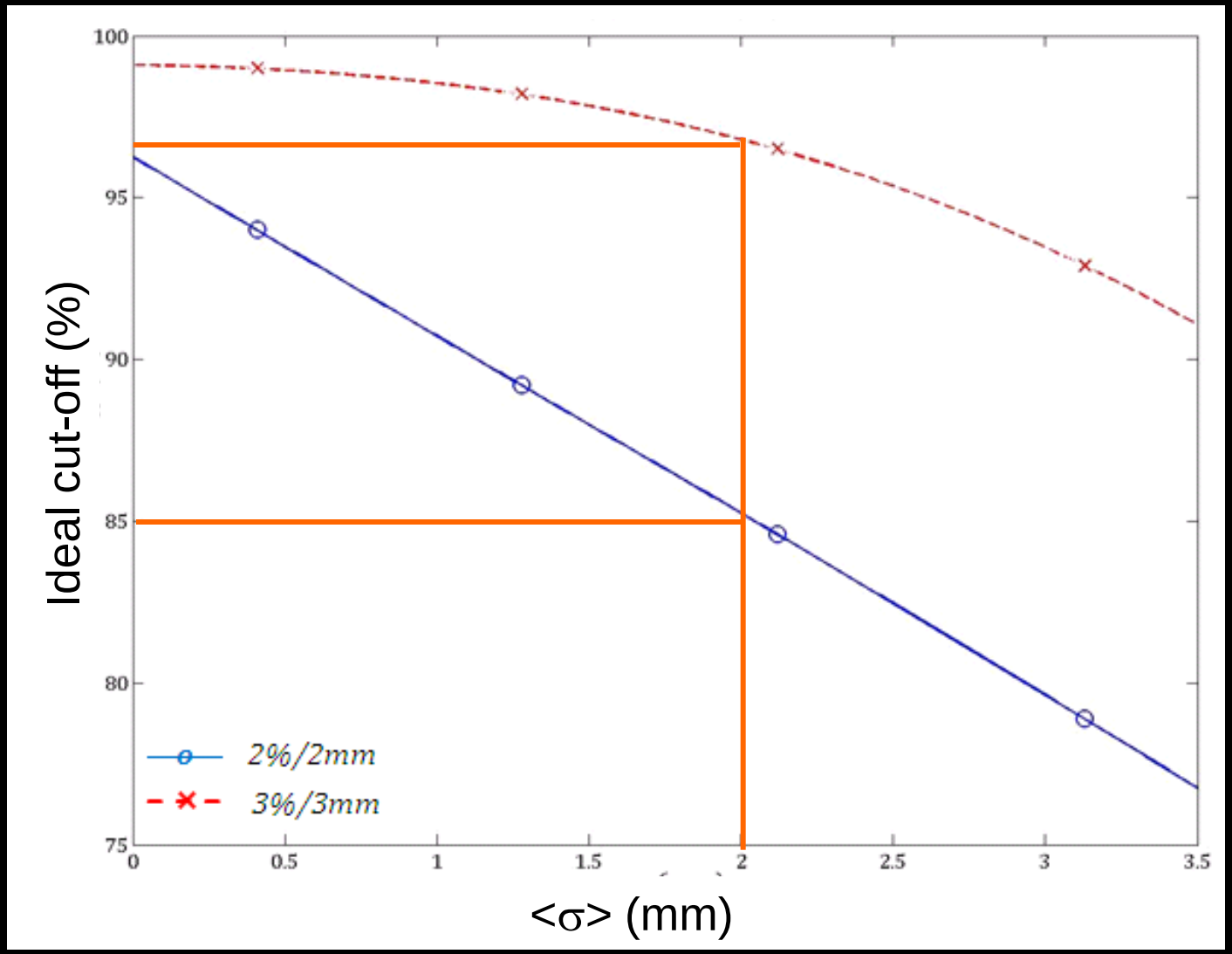

D)
2%/2mm % Pass for <σ>=0.41 mm
Unperturbed plans
Perturbed plans
% plans that fail
Pass-rate


ROC plots
TPF
FPF
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
3%/3mm, <σ>=3.13 mm
3%/3mm, <σ>=2.12 mm
2%/2mm, <σ>=1.28 mm
2%/2mm, <σ>=0.41 mm

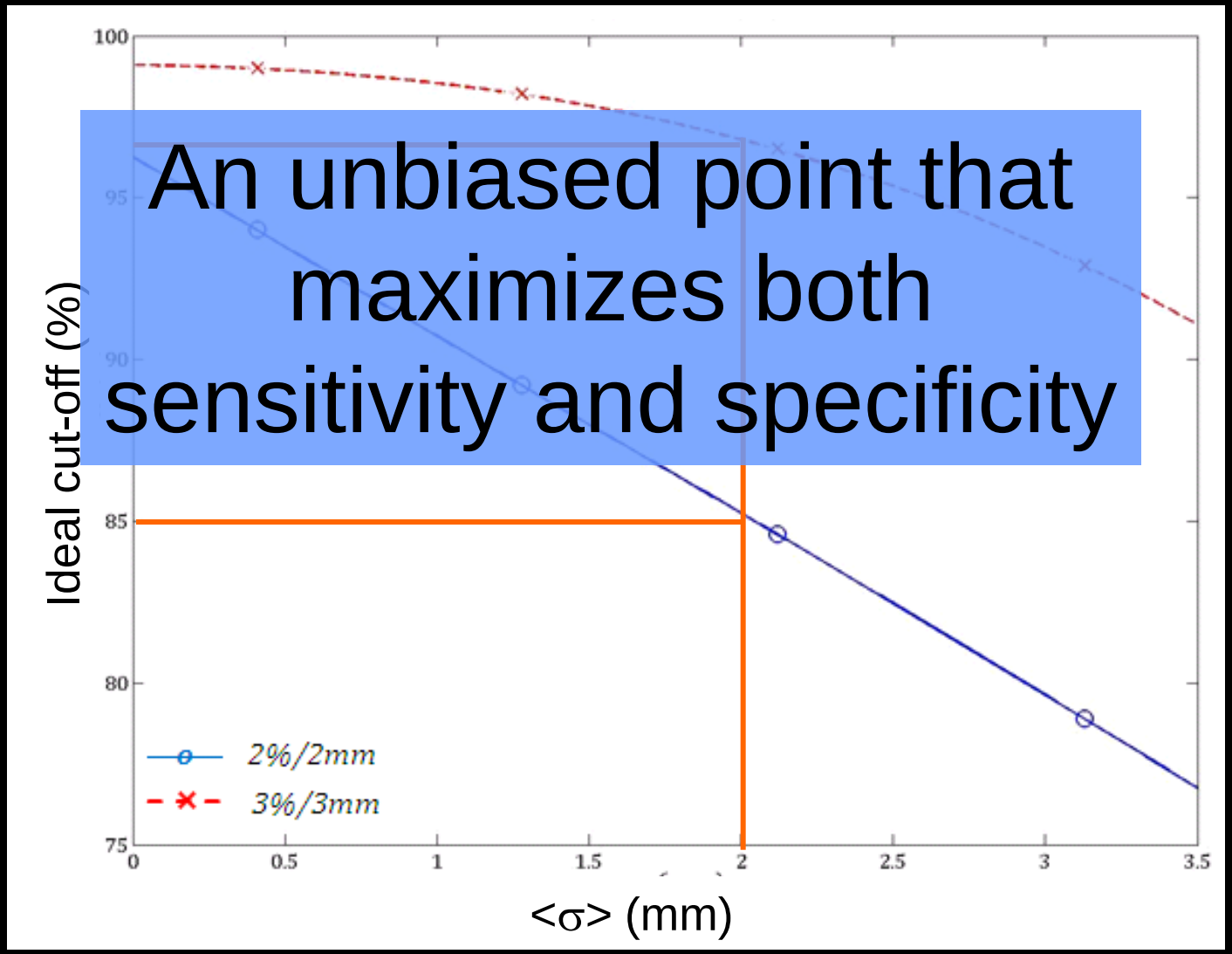
Setting the detector's operating point



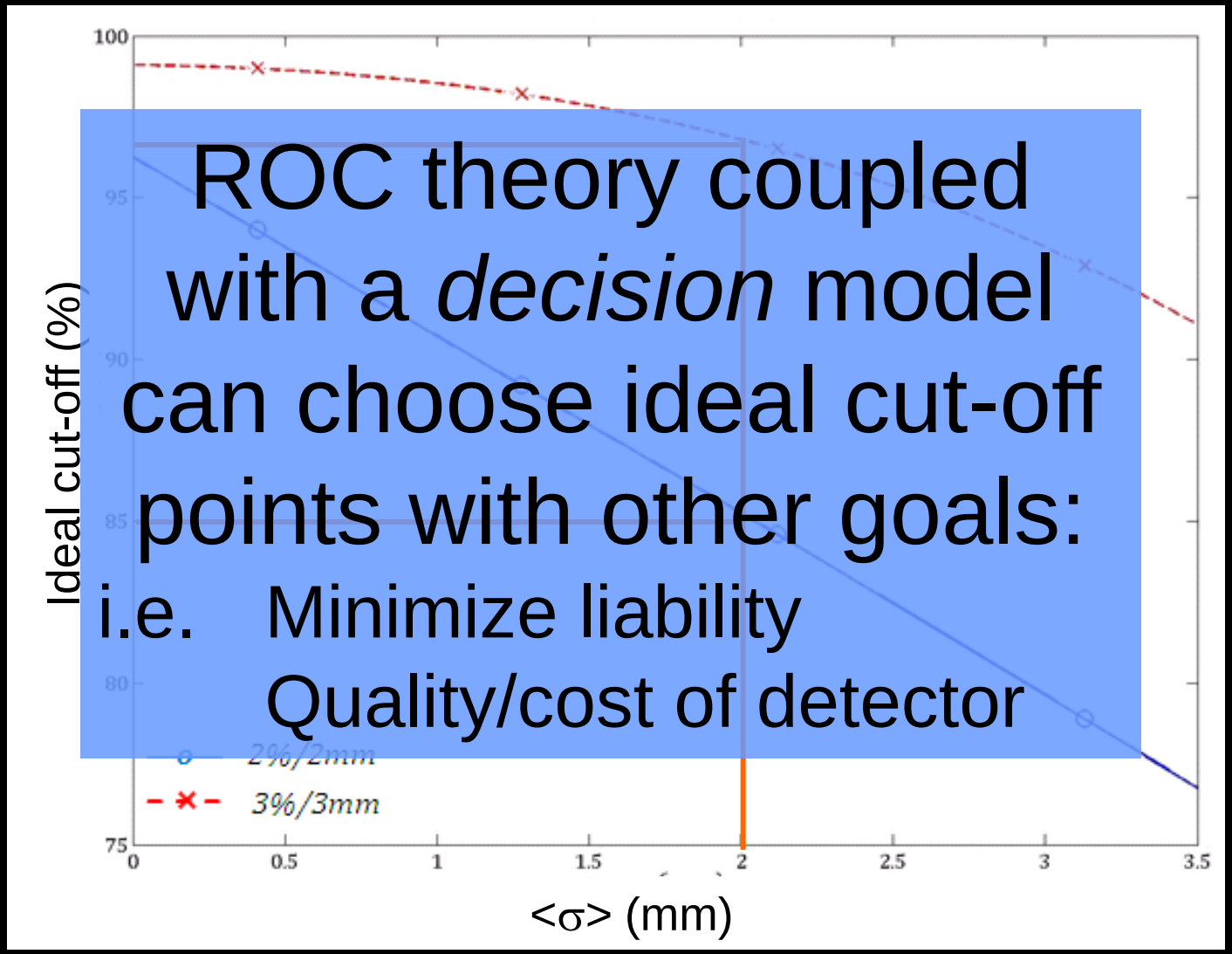
Setting the detector's operating point



Setting the detector's operating point



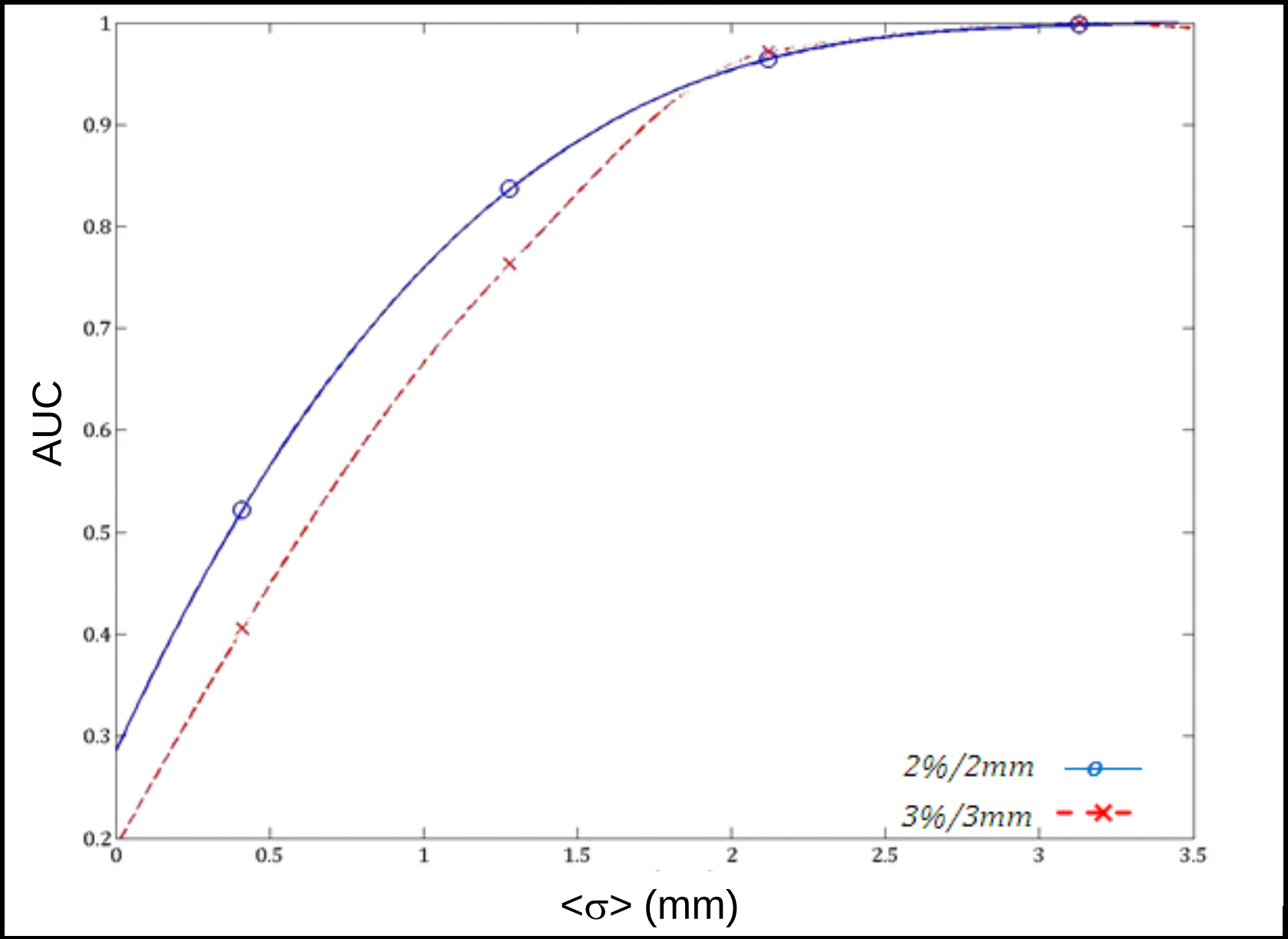
Setting the detector's operating point

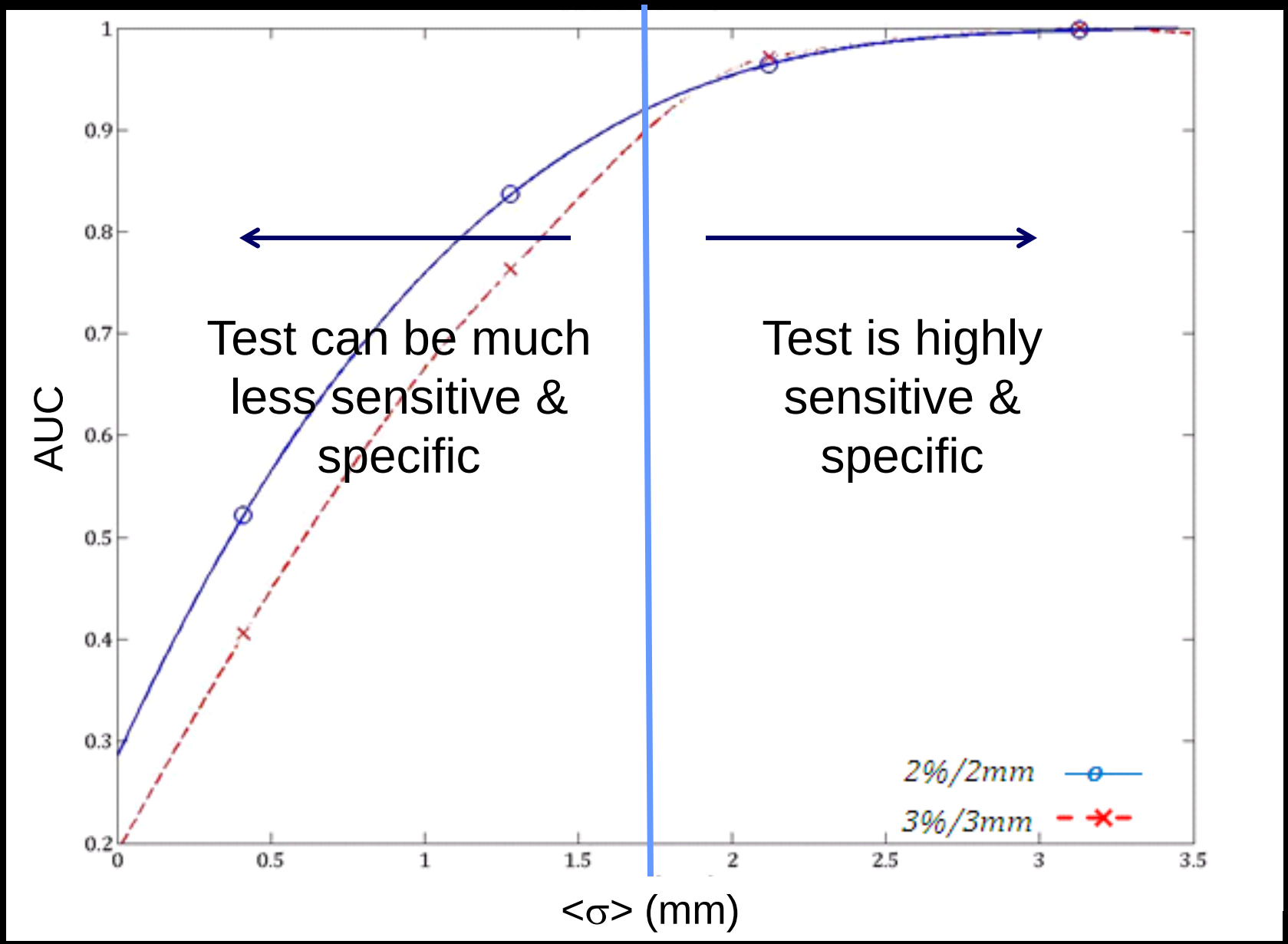


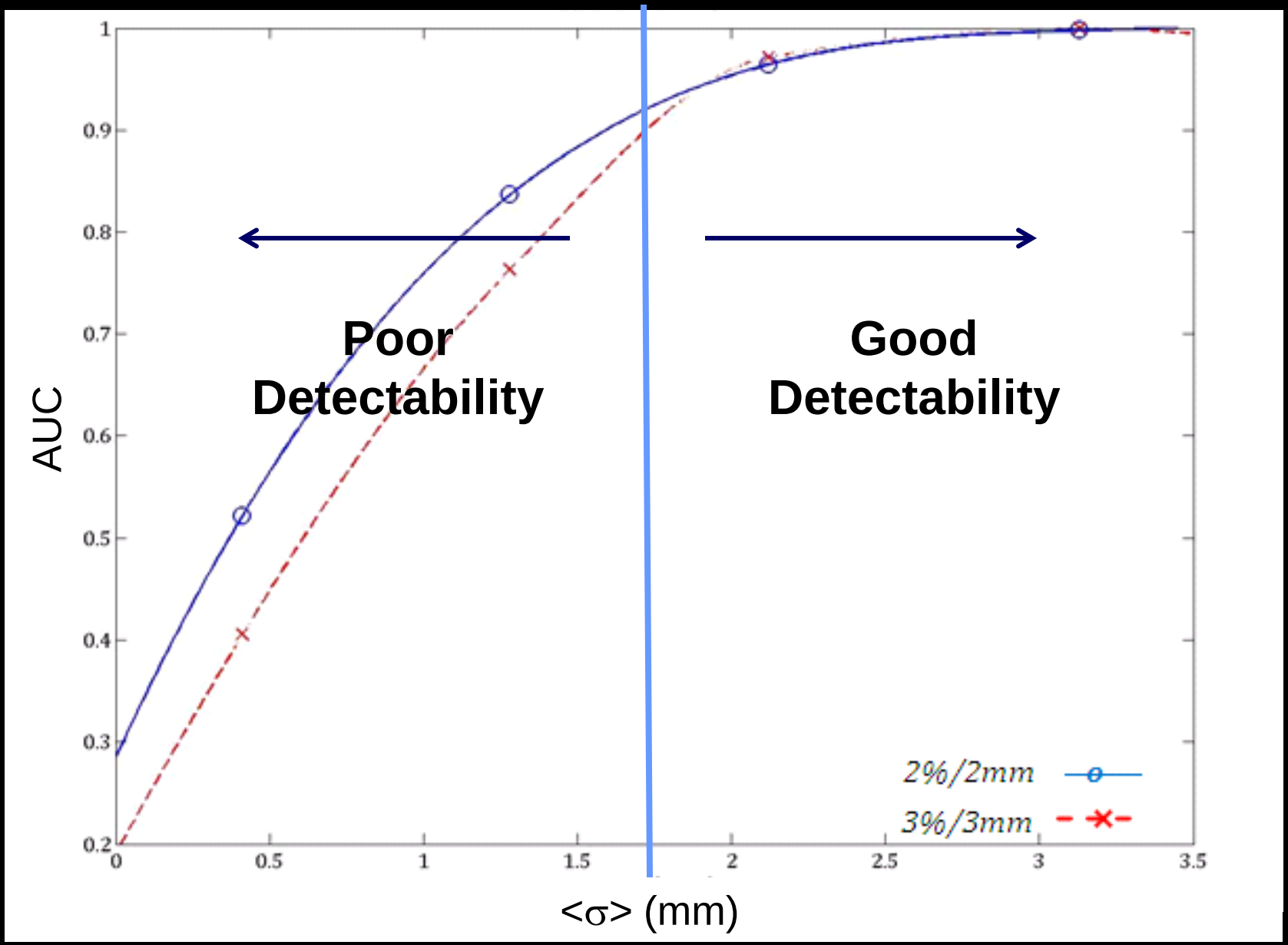
ROC theory coupled with a *decision* model can choose ideal cut-off points with other goals:

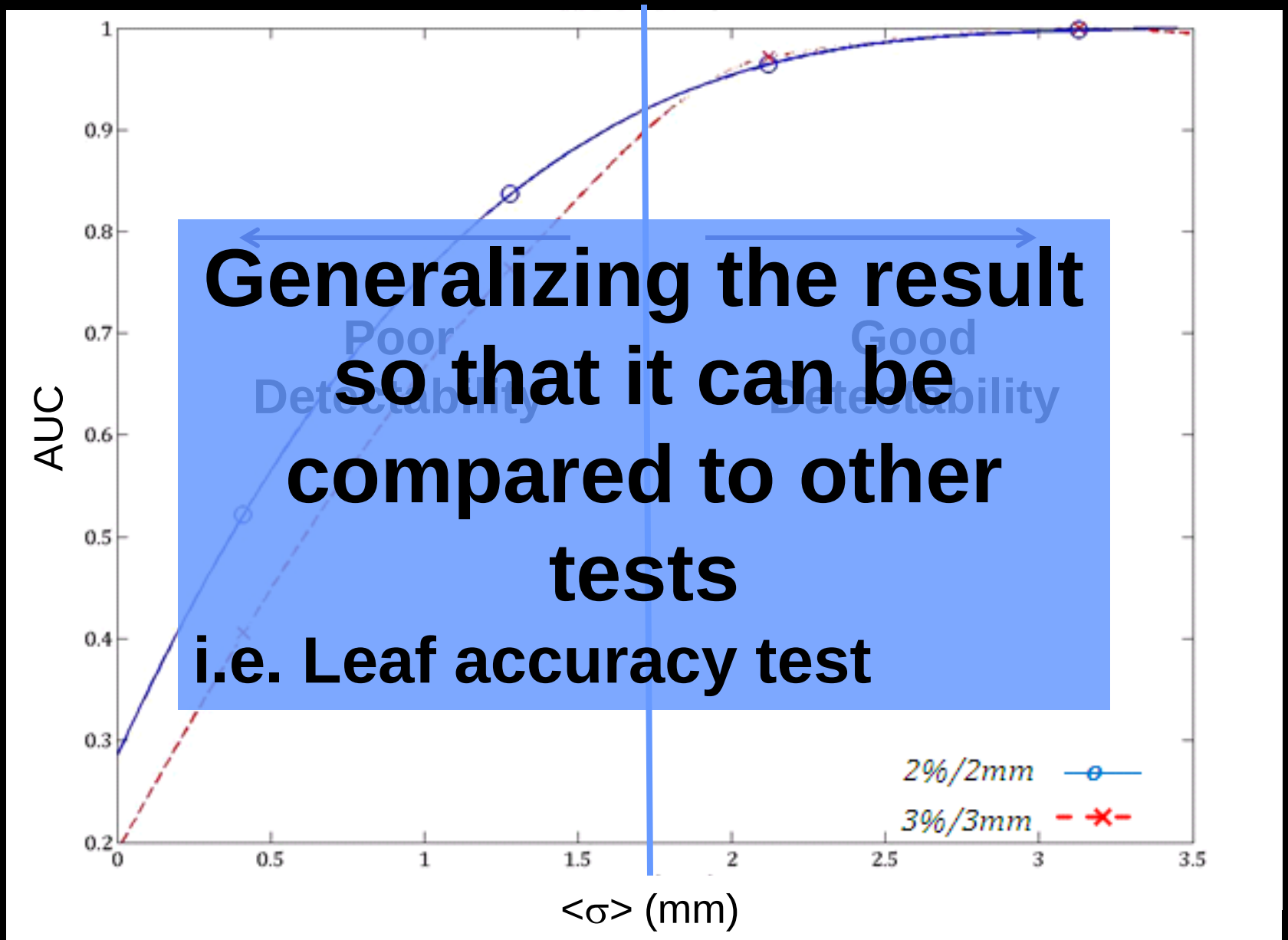
i.e. Minimize liability

Quality/cost of detector


AUC
<σ> (mm)
2%/2mm
3%/3mm
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.5
1
1.5
2
2.5
3
3.5


AUC
<σ> (mm)
Test can be much less sensitive & specific
Test is highly sensitive & specific
2%/2mm
3%/3mm
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.5
1
1.5
2
2.5
3
3.5


AUC
<σ> (mm)
Poor
Detectability
Good
Detectability
2%/2mm
3%/3mm
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
0.5
1
1.5
2
2.5
3
3.5



Generalizing the result so that it can be compared to other tests

i.e. Leaf accuracy test

The problem with γ

- **“Gamma” is not a physical measurement, nor does it have statistical meaning.**
- **It has local significance, but it is difficult to interpret in a broad, multi-institution sense.**
- **It’s interpretation will always be controversial.**

Summary

- **QA activities are meant to help a specific local problem.**
- **Local problems are often relevant with a ‘narrow’ set of constraining factors.**
- **For the knowledge gained from these activities to be helpful in a broad sense, efforts can be made to choose metrics that have a generalized context.**
- **In the example of patient specific QA, thinking of the problem from a ‘detectability’ point of view led to quantification of some parameter thresholds that may be helpful in more general problem solving.**