Evaluating diagnostic technologies: sensitivity, specificity & false reporting

Mr. Panagiotis Georgiou MD MRCS

Academic Clinical Fellow in G. Surgery

Imperial College, London, UK

Aims of Lecture

- Demonstrate the various methods of evaluating diagnostic technologies in terms of sensitivity, specificity and false reporting

- Discuss the use of diagnostic technologies for population-based screening and targeted screening of disease

- Construct an argument for imaging-based follow-up strategies of curative surgical procedures

Diagnostic technologies

- X-Rays
  - 1895
  - Dr Wilhem Rontgen

- USS
  - 1956
  - Prof Ian Donald
  - O&G
  - Endoscopic
  - Intraoperative

- CT
  - 1972
  - Dr Godfrey Hounsfield
  - Quality improved

- MRI
  - 1967
  - Nottingham
  - No Clinical Use
  - 1980's
  - DW-MRI

- Nuclear Medicine
  - 1896
  - Thyroid
  - PET
    - PET/CT
    - PET/MRI

Which technology?

- Evaluate

- Identify Strengths and Weaknesses

- Improve technology and knowledge

- Re-evaluate

Research

- Diagnostic Accuracy Studies
  - STARD Criteria (25)
  - Compare
  - Prospective or Retrospective

- Meta-analysis

Characteristics of Diagnostic Tests

- Test effectiveness measured as
  - **Sensitivity:** ability to confirm disease
  - **Specificity:** ability to identify disease absence
- Clinical importance related to predictive ability
  - **Positive Predictive Value:** proportion testing positive who actually have the disease
  - **Negative Predictive Value:** proportion testing negative who do not have the disease

1ˢᵗ Step

- Technology to be evaluated

- Disease/Condition
  - Specific question to be answered

- Study Population

- Reference Standard

Index Test

- Blinded radiologists (at least 2)
- Level of expertise
- Standardised report i.e. proforma

- ? Agreement/Interobserver agreement

Reference Test

- Defines the presence or absence of disease
- Ideally should be 100% accurate
- Applicable results to all patients within the group
- ?

- Degree of compromise
- Period of follow up may be necessary

Blinding

Blinding of Radiologists
- Blinded to Reference test outcome
- Blinded to Surgical outcome
-

Blinding of Reference Test assessor
- Blinded to Index test result
- Blinded to Surgical outcome

- Eliminates biased outcome assessment

Bias

- Case-Selection

- Performance of reference test relies on the performance of index test

- Incorporation bias

- Time Interval between index and reference test

- Not blinded radiologists/reference standard assessor

Avoid Bias

- Representative group of patients with disease

- Index test is compared to an independent reference test

- Radiologists and reporters of reference tests are all blinded to outcome (s).

Definitions
2x2 table

		Reference Test	
		Positive	Negative
Index Test	Positive	TP (a)	FP (b)
	Negative	FN (c)	TN (d)

Definitions

- **True Positive**
  - Correct diagnosis of presence of disease/condition

- **True Negative**
  - Correct diagnosis of absence of disease/condition

- **False Positive**
  - Incorrect diagnosis of presence of disease/condition

- **False Negative**
  - Incorrect diagnosis of absence of disease/condition

Sensitivity

- Proportion of those with disease who test positive in the study group. *(Positive in disease)*

- How good the index test is to ***pick up*** the disease.

- Sensitivity = TP / (TP+FN) = a / (a+c)

Sensitivity

		Reference Test	
		Positive	Negative
Index Test	Positive	TP (a)	FP (b)
	Negative	FN (c)	TN (d)
		Sensitivity TP/Presence	

Specificity

- Proportion of those without disease who test negative in study group. *(Negative in health)*
- How good the test is to ***exclude*** the disease.

- Specificity = TN / (TN+FP) = d / (d+b)

Sensitivity

		Reference Test	
		Positive	Negative
Index Test	Positive	TP (a)	**FP (b)**
	Negative	FN (c)	**TN (d)**
			Specificity TN/ABscence

The Ideal Situation

		Reference Test	
		Positive	Negative
Index Test	Positive	200	0
	Negative	0	800
		100%	100%

100% Agreement

Reality

		Reference Test	
		Positive	Negative
Index Test	Positive	170	30
	Negative	30	770
		85%	96.25

Consequences of a False Positive

- Follow-up tests
- Cost

- Potential harm
  - Surgery
  - More tests

- Anxiety

Consequences of a False Negative

- Disease undetected and progress

- At best, a false sense of security (screening)

- Might neglect future screening tests (screening)

The Tradeoff: Sensitivity vs. Specificity

- If missing cancers is a concern, **sensitivity** can be raised by adjusting the diagnostic cut point for a positive result

- But, the false positive rate will also increase

- Impact on screening program costs?

- Specificity may be the determining factor in the success of screening programs
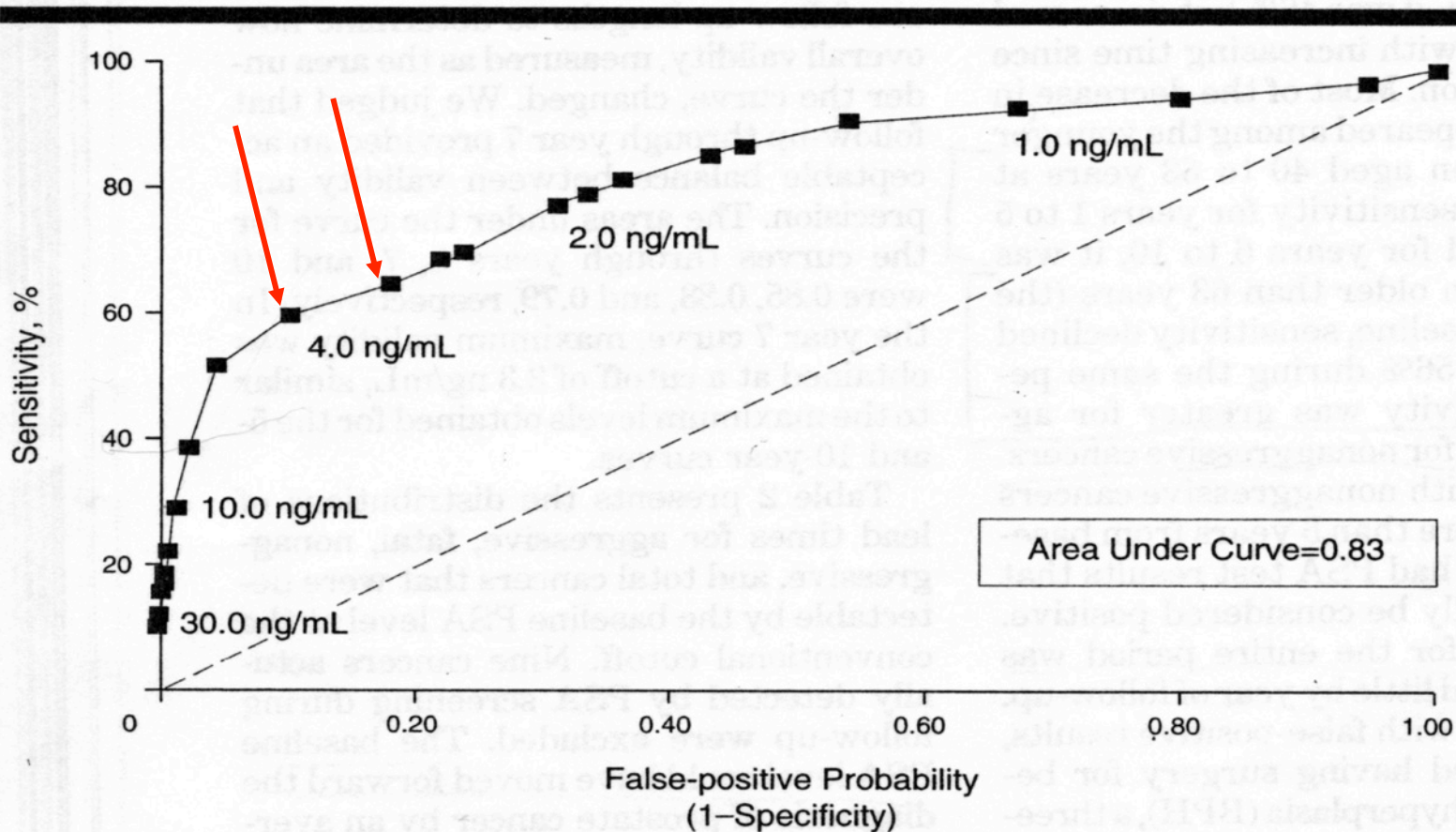
Changing a Diagnostic Cut Point

Figure 2.—Sensitivity and specificity for prostate-specific antigen and prostate cancer at various cutoff points, during 7 years of follow-up: Physicians' Health Study (203 cases and 609 controls).

Predictive Values

- Important for the Clinician
- If a test result is positive, how likely is it that this individual has the disease?
- Predictive value varies with the **prevalence** of the disease in the screened population

- *Bayes' theorem*: As the prevalence of a disease increases, the positive predictive value of the test increases (PPV) and its negative predictive value (NPV) decreases.

Positive Predictive Value

- Probability that a positive test indicated the presence of disease.
- Depends on prevalence

- PPV= TP / (TP+FP) = a / (a+b)

Positive Predictive Value

Index Test		Reference Test		
		Positive	Negative	
	Positive	TP (a)	FP (b)	PPV TP/All +ve's
	Negative	FN (c)	TN (d)	

Negative Predictive Value

- The probability that a negative test result indicated the absence of disease.

- Depends on prevalence

- NPV= TN / (TN+FM) = c / (c+d)

Positive Predictive Value

Index Test		Reference Test		
		Positive	Negative	
	Positive	TP (a)	FP (b)	
	Negative	**FN (c)**	**TN (d)**	NPV TN/ all –ve's

Predictive Values and Prevalence

Sensitivity = 99%; Specificity = 95%

Prevalence = 1%	Disease Yes	Disease No	PPV
Positive result	99	495	
Negative result	1	9405	
Total	100	9900	**17%**
Prevalence = 5%			
Positive result	495	475	
Negative result	5	9025	
Total	500	9500	**51%**

The Ideal Situation

		Reference Test		
		Positive	Negative	
Index Test	Positive	200	0	100%
	Negative	0	800	100%

Reality

		Reference Test		
		Positive	Negative	
Index Test	Positive	170	30	85%
	Negative	30	770	96.25%

Level of Agreement

- Importance
  - Reproducibility of outcomes

- Assessment

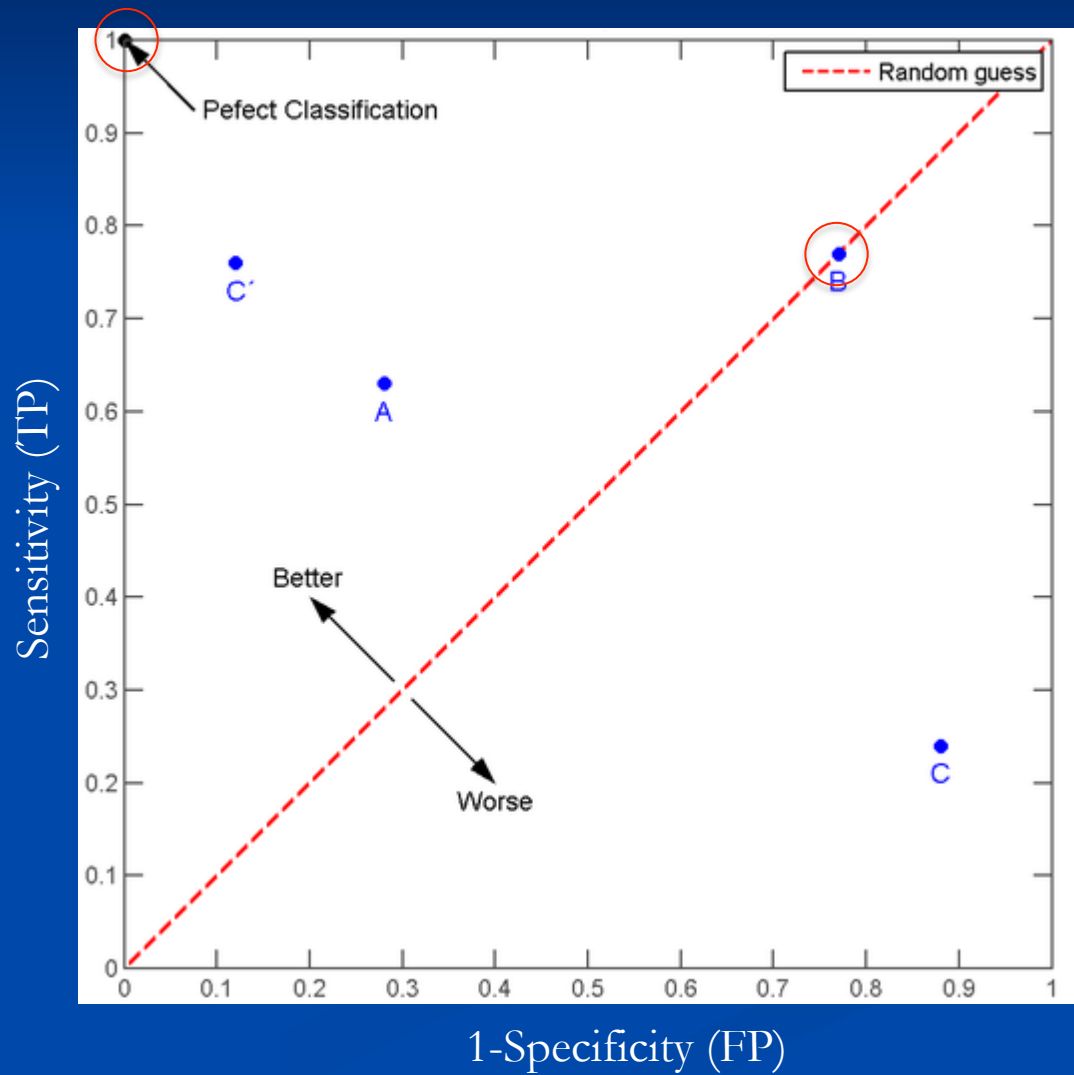- Cohen's kappa (k) coefficient

Cohen's Kappa Coefficient

- Inter-observer agreement
- Reproducibility of results

$$\kappa = \frac{Pr(a) - Pr(e)}{1 - Pr(e)}$$

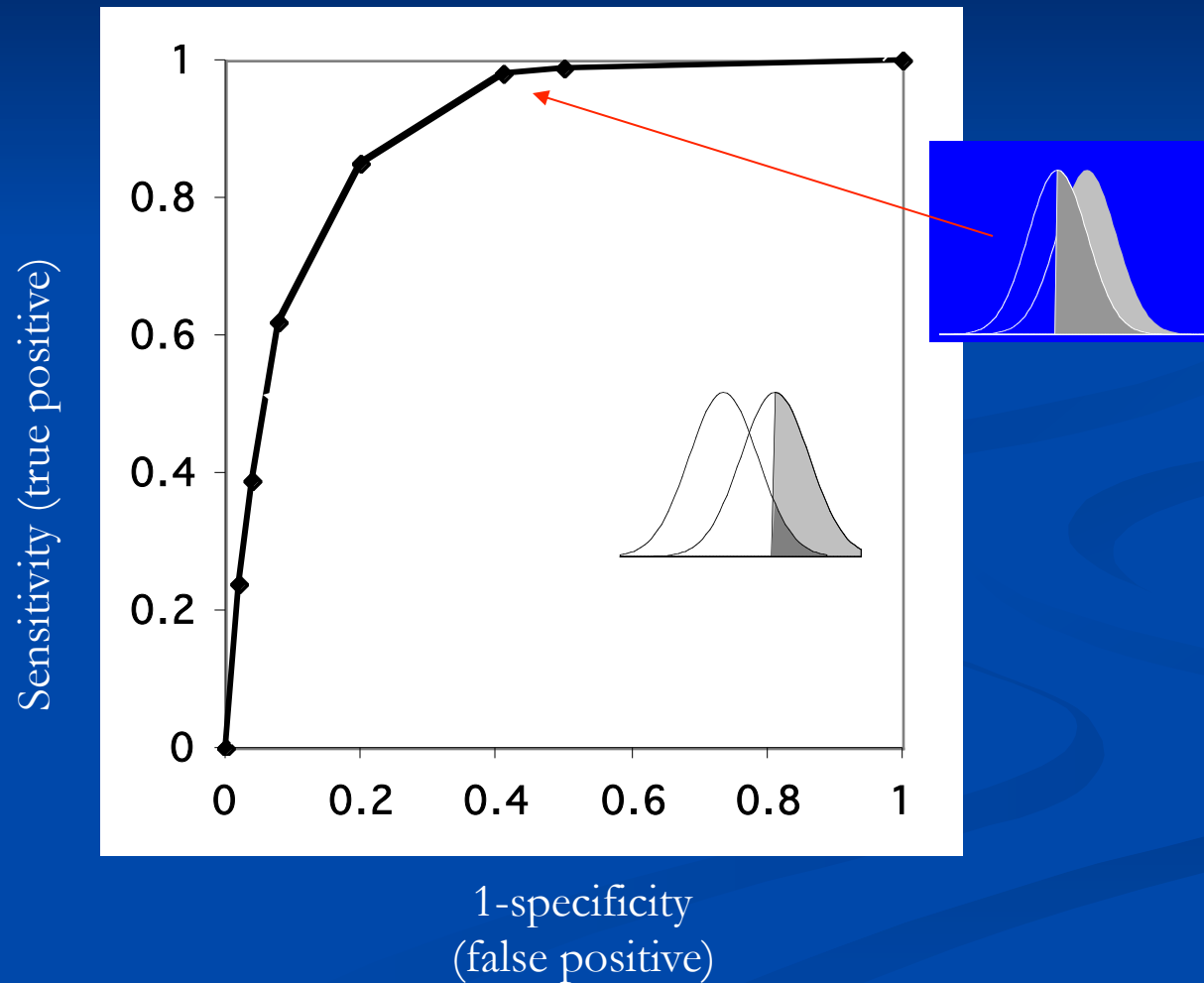| k	Agreement
< 0.20	Poor
0.20 - 0.40	Fair
0.40 - 0.60	Moderate
0.60 - 0.80	Good
0.80 - 1.00	Very Good

Receiver-Operating Characteristic (ROC) Curve

- Graphical Plot

- Summarize sensitivity and specificity as cutoff changes

- Set a cutoff

- Compare index tests

ROC Curve for endometrial thickness

Screening - WHO

- Important Health Problem
- Should be a treatment
- Facilities for diagnosis and treatment available
- Latent stage of disease
- Should be a test
- Natural History of disease understood
- Clear Guidelines for treatment
- Cost
- Continuous Process
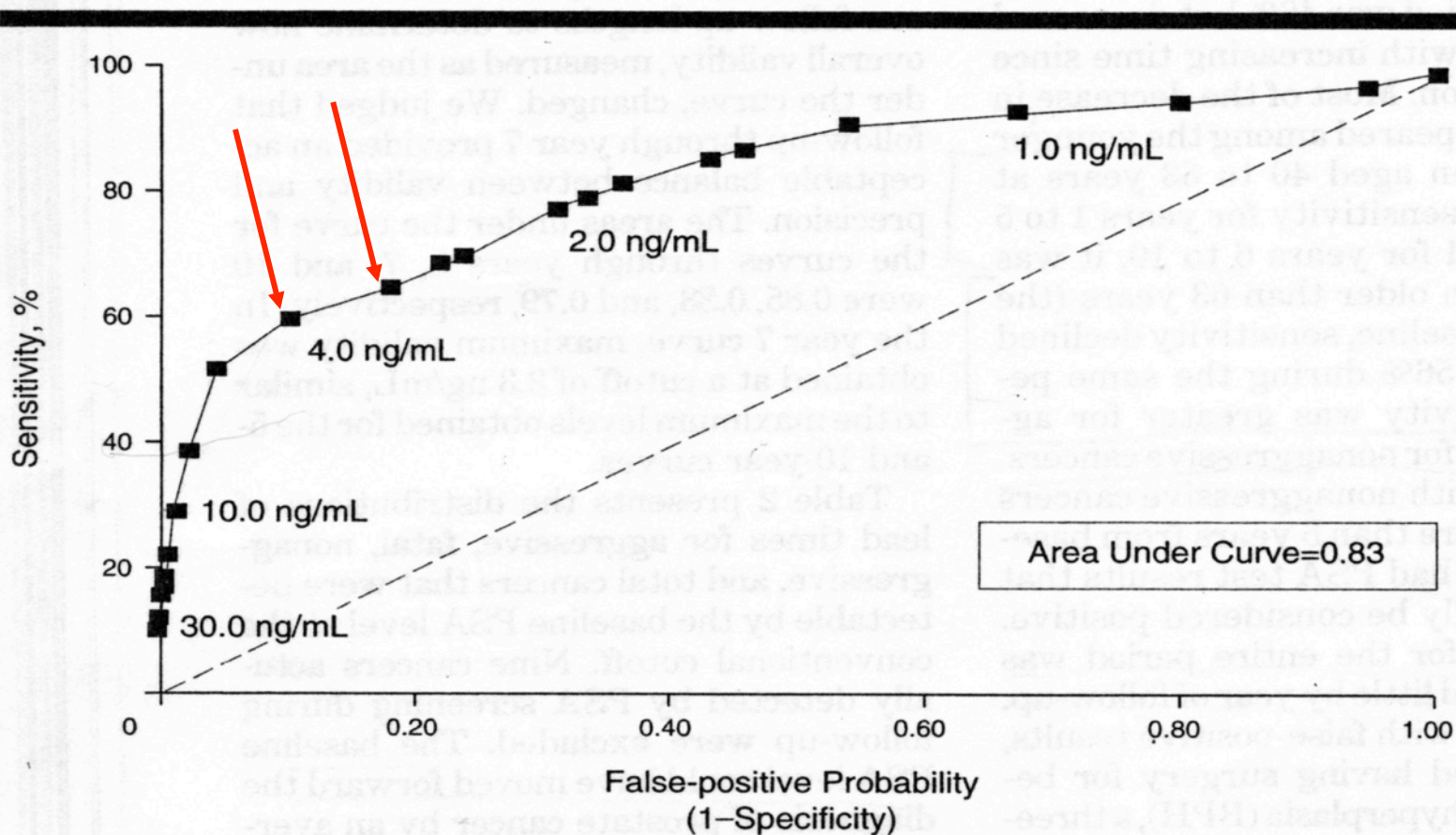
Changing a Diagnostic Cut Point

Figure 2.—Sensitivity and specificity for prostate-specific antigen and prostate cancer at various cutoff points, during 7 years of follow-up: Physicians' Health Study (203 cases and 609 controls).

Adjusting the Cutpoint

- Increase sensitivity when dealing with cancer

- This will increase FP

- Impact on Specificity

- Increase Cost of further investigation

- Use of a 2nd screening test with high specificity to filter FP

Faecal Occult Blood Test

- Age >60

- High Sensitivity
- Poor Specificity
  - Benign conditions

- Positive result
- Endoscopic studies

Cervical Smear – Pap Test

- Females 21-65
- To detect pre-cancerous conditions
- High sensitivity
- Low specificity

- If positive
- HPV test to exclude viral infection
- Further investigation
- Adjust Follow up

Limitations of Screening

- Consequences of FP and FN
- Resources
- Complications from Screening test
- Bias
  - Lead Time Bias
  - Length Time Bias
  - Overdiagnosis

Follow Up Post Curative Surgery

- The same as screening
- Changes from Surgery
  - Inflammation
  - Anatomy changes
- Length of Follow up
  - Natural History of disease
  - Stage of Disease
  - Timing - Majority of Recurrences – More often
  - When to stop

CRC Follow up

- Majority of recurrences within 2 years
- CEA
    - 3 monthly for first 2 years
    - 6 monthly for 3rd year
    - Then yearly

- CT and MRI
    - 6 monthly for 2 years
    - Then yearly
- PET scan on suspicion

- Follow up stops at 5 years

Thank you

Practice ?

Example 1 – Pelvis AB PR

		Histopathology		
		Positive	Negative	
MRI	Positive	40	3	93%
	Negative	2	19	90.5%
		95.2%	86.4%	

Example 2 – Pelvis Lateral

		Histopathology		
		Positive	Negative	
MRI	Positive	25	5	83.3%
	Negative	3	31	91.2%
		89.3%	86.1%	

Example 3 – Pelvis AB PR

		Histopathology		
		Positive	Negative	
MRI	Positive	17	3	85%
	Negative	1	43	97.7%
		94.4%	93.5%	