# Diagnosis

Sarah Collin
Research and Development support unit

# Outline

- Design
- Sensitivity
- Specificity
- ROC curves
- Predictive values
- Eliminating bias
- Appraising studies
- Summary

# Ideal design of diagnostic studies

All participants

↓

Index test

↓

Reference test

↓

2 x 2 Table of results

**2 x 2 table of results**

| Index test | Reference test + | Reference test - |
|---|---|---|
| + | 90 | 60 |
| - | 10 | 240 |

Slide from Susan Mallett, Cancer Research UK Medical Statistics Group

# Hypothetical example

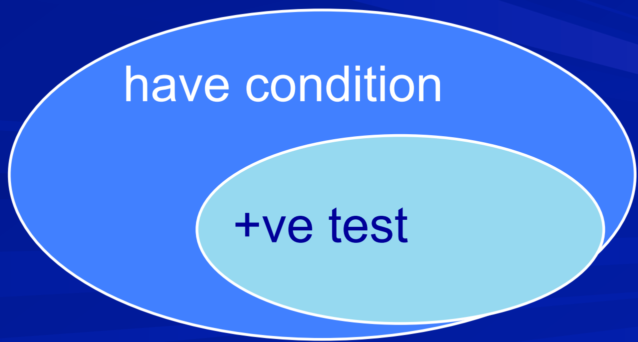
| | Actual +ve | Actual –ve | Total |
|---|---|---|---|
| Test +ve | 95 | 45 | 140 |
| Test -ve | 5 | 855 | 860 |
| Total | 100 | 900 | 1000 |

- What proportion of people tested had the disease?
- How 'accurate' is the test? (proportion of correct results)

# Sensitivity

| | Actual +ve | Actual –ve | Total |
|---|---|---|---|
| Test +ve | 95 | 45 | 140 |
| Test -ve | 5 | 855 | 860 |
| Total | 100 | 900 | 1000 |

- What proportion of people who have the condition are identified as positive by the test?



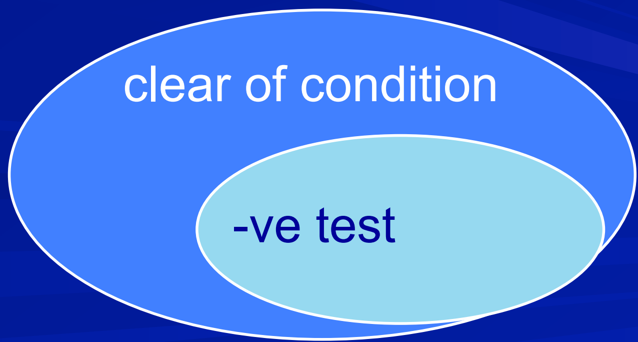
# Sensitivity

- If a test has very high sensitivity
  - –most people with the condition are picked up by the test

# Specificity

| | Actual +ve | Actual –ve | Total |
|---|---|---|---|
| Test +ve | 95 | 45 | 140 |
| Test -ve | 5 | 855 | 860 |
| Total | 100 | 900 | 1000 |

- What proportion of people who do not have the condition are identified as negative by the test?



# Specificity

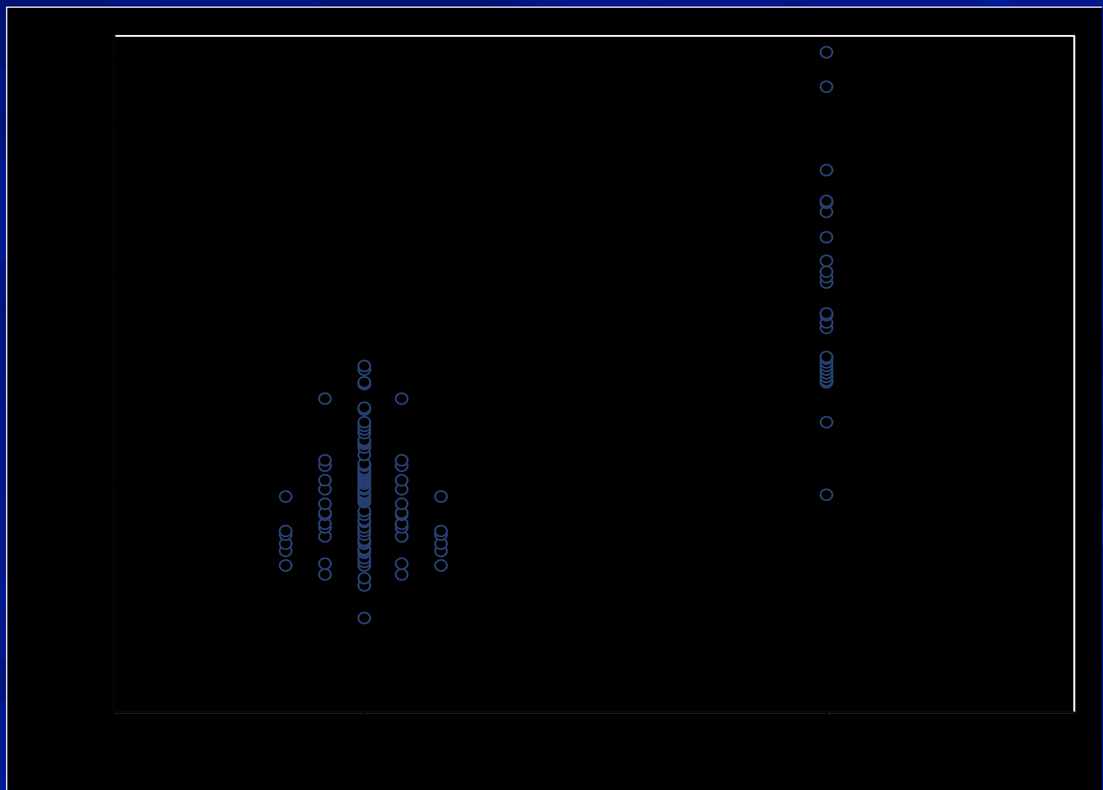
- If a test has very high specificity
  - –most people without the condition are ruled out by the test

# Notes

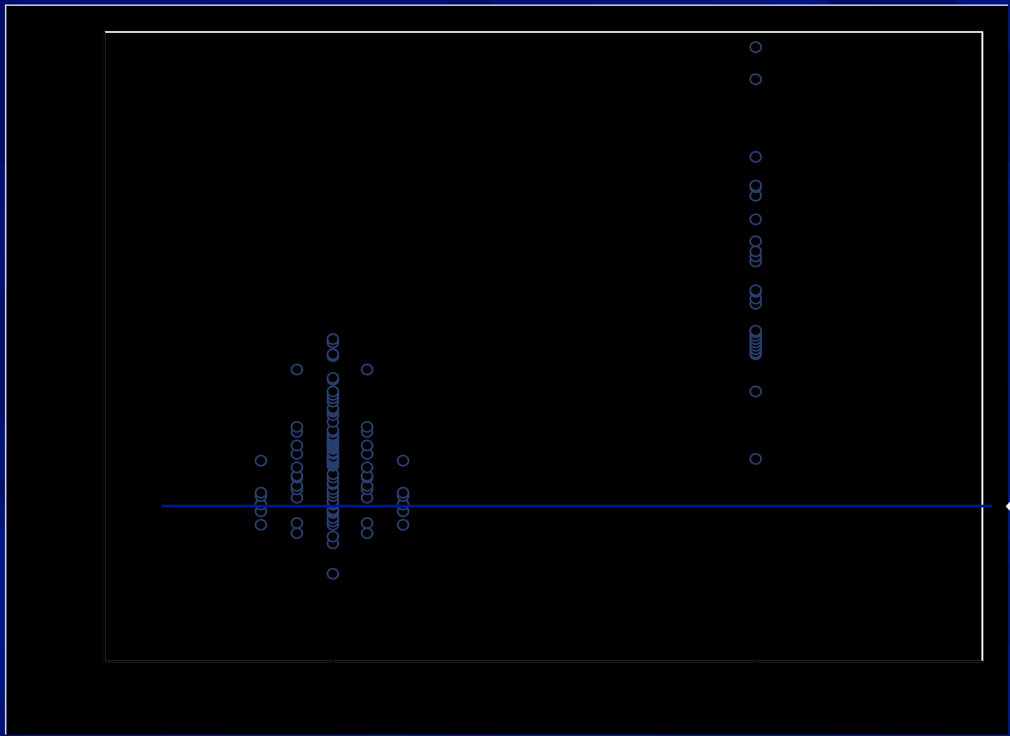
- It is essential to have a confirmed true diagnosis (+ve/-ve) for every patient to be able to judge the accuracy of a test (e.g. gold standard, long term follow up)
- Sensitivity and specificity should be accompanied by confidence intervals to convey the amount of uncertainty (simple proportions – use StatsDirect)

# Tests based on continuous variables:

## e.g. creatinekinase in patients with unstable angina or acute myocardial infarction



Data of Frances Boa, from ‘An introduction to Medical Statistics’ by Martin Bland

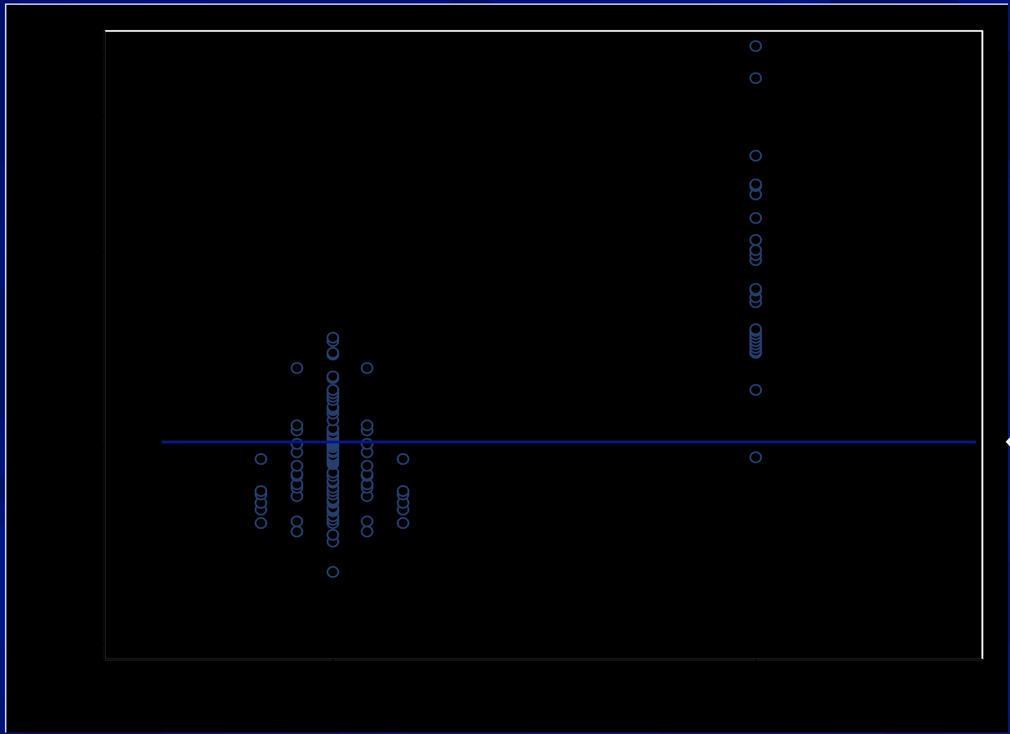


cut-off at 80

| | Actual +ve | Actual –ve | |
|---|---|---|---|
| Test +ve | 27 | 54 | 81 |
| Test -ve | 0 | 39 | 39 |
| | 27 | 93 | 120 |

Sensitivity=27/27=100%

Specificity=39/93=42%



cut-off at 100

| | Actual +ve | Actual –ve | |
|---|---|---|---|
| Test +ve | 26 | 35 | 61 |
| Test -ve | 1 | 58 | 59 |
| | 27 | 93 | 120 |

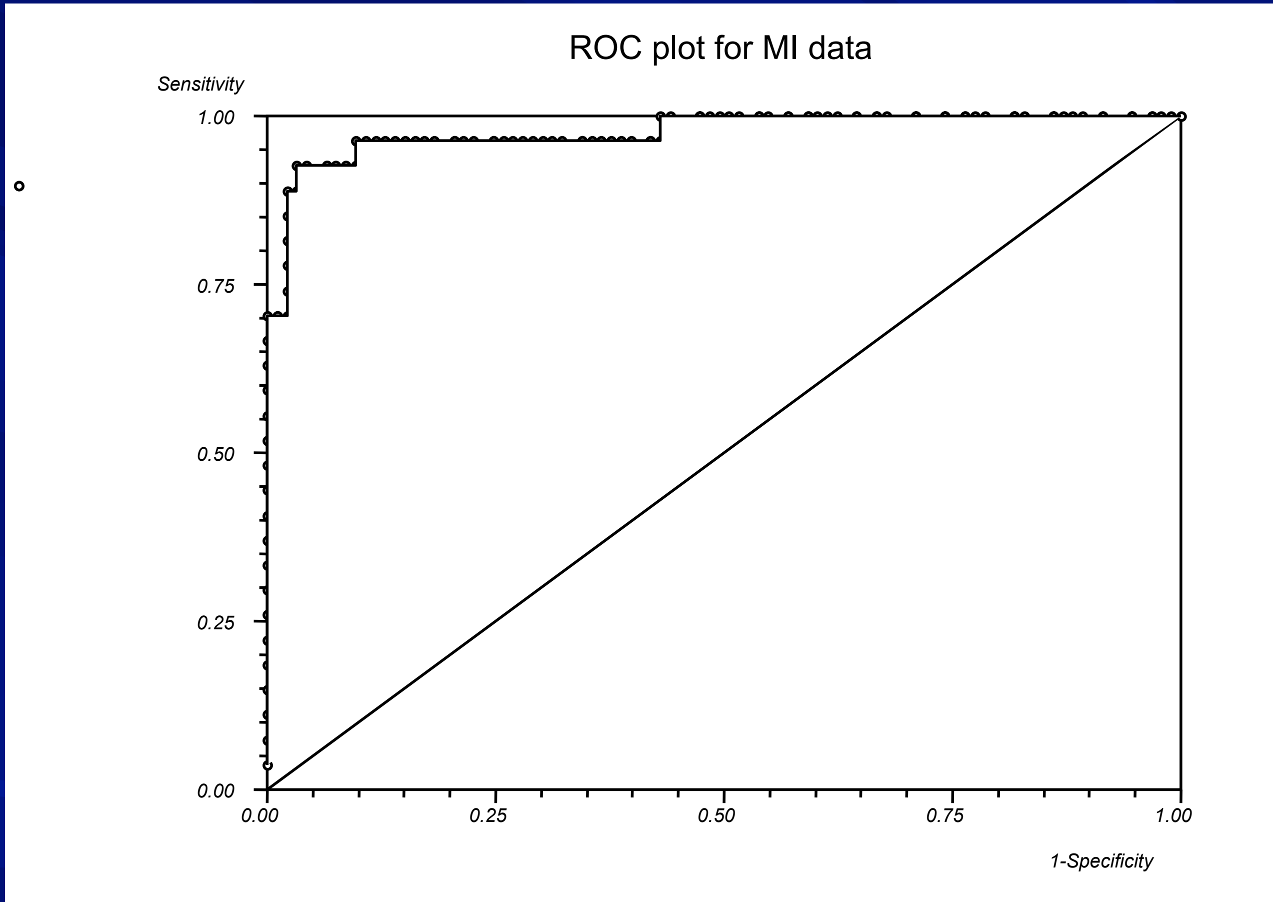
Sensitivity=26/27=96%

Specificity=58/93=62%

# Investigating the trade off between sensitivity and specificity

- Generally plot sensitivity v (100%-specificity)
- ROC curve (receiver operating characteristic)
- Look for cut off that gives us both high sensitivity and high specificity
  - Increase in sensitivity is at expense of specificity and vice versa
- Should always check sensitivity and specificity of cut off in a different sample to be sure

# ROC curve




'Optimum' cut-off point selected = 302

sensitivity (95% CI) = 0.93 (0.76 to 0.99)

specificity (95% CI) = 0.97 (0.91 to 0.99)

Note: 'optimum' assumes sensitivity and specificity of equal concern

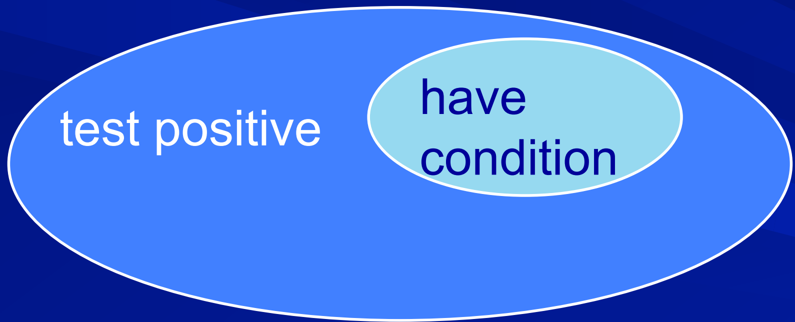
# Area under ROC curve

- Area under curve is an estimate of ‘probability that creatinekinase of random person with AMI will be higher than for random person with angina’
- Can be useful for comparing two tests

# Predictive values

- Positive predictive value = probability that a person has the condition, given that their test result is positive

i.e. the proportion of people with a positive result that actually have the condition




negative predictive value = probability that a person is free of the condition, given that their result is negative

# Hypothetical example

| | Actual +ve | Actual –ve | Total |
|---|---|---|---|
| Test +ve | 95 | 45 | 140 |
| Test -ve | 5 | 855 | 860 |
| Total | 100 | 900 | 1000 |

- Postitive predictive value?
- Negative predictive value?

# Note

- Sensitivity and specificity of a test should be constant
- Positive predictive values will vary depending on the prevalence

# Test with 95% sensitivity and 95% specificity

| Example of the effect of the prevalence of disease on the reliability of a diagnostic test | | |
|---|---|---|
| Prevalence (pre-test probability of disease) | Probability of having the disease given a positive test result | Probability of having the disease given a negative test result |
| 1% | 16% (84% false positive results) | 0.053% (99% true negative result) |
| 10% | 68% (32% false positive results) | 0.58% (99% true negative result) |
| 25% | 86% (14% false positive result) | 1.74% (98% true negative result) |

MeReC Briefing: supplement to issue 30

# Notes

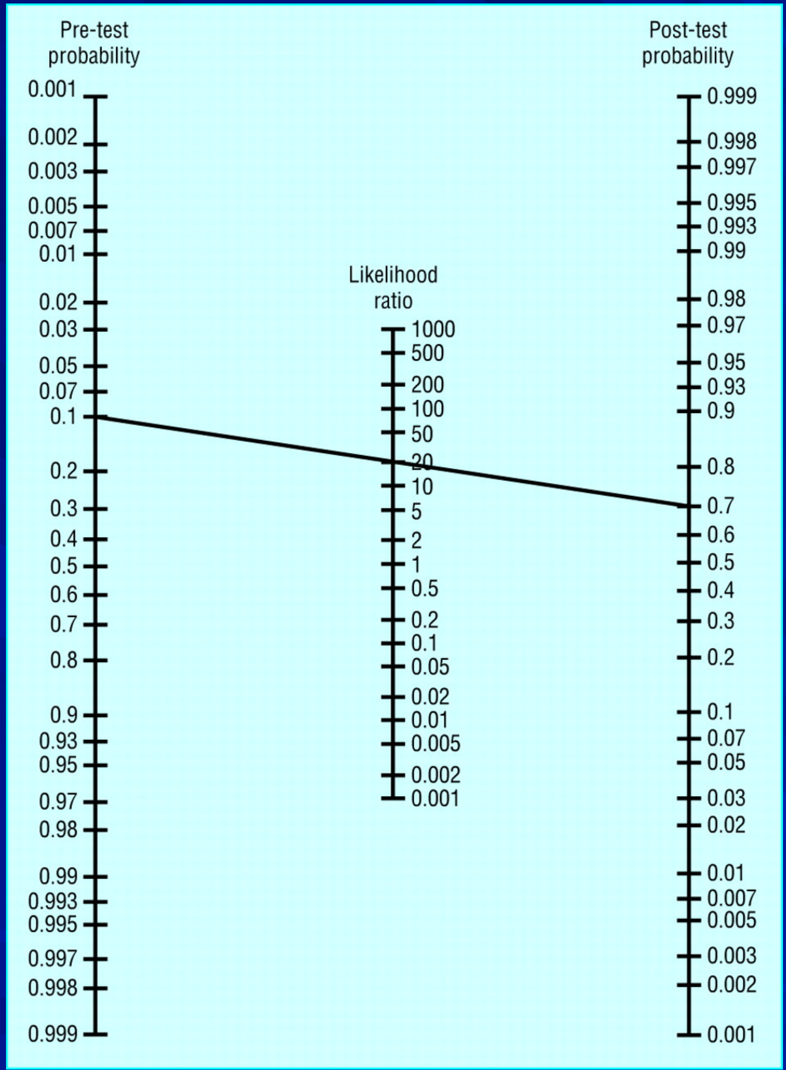
- Important that tests developed in population for which they will be used
- Good diagnostic test not necessarily a good screening test

# Likelihood ratio

- How many times more (or less) likely patients with the condition are to have that particular result than patients without the disease
- Can be used to calculate the probability of individual patient having condition based on test results

See Diagnostic test 4: likelihood ratios by Deeks and Altman; BMJ 2004 329 p168

## Use of Fagan's nomogram for calculating post-test probabilities7




**Deeks, J. J et al. BMJ 2004;329:168-169**

bmj.com

# Bias in studies

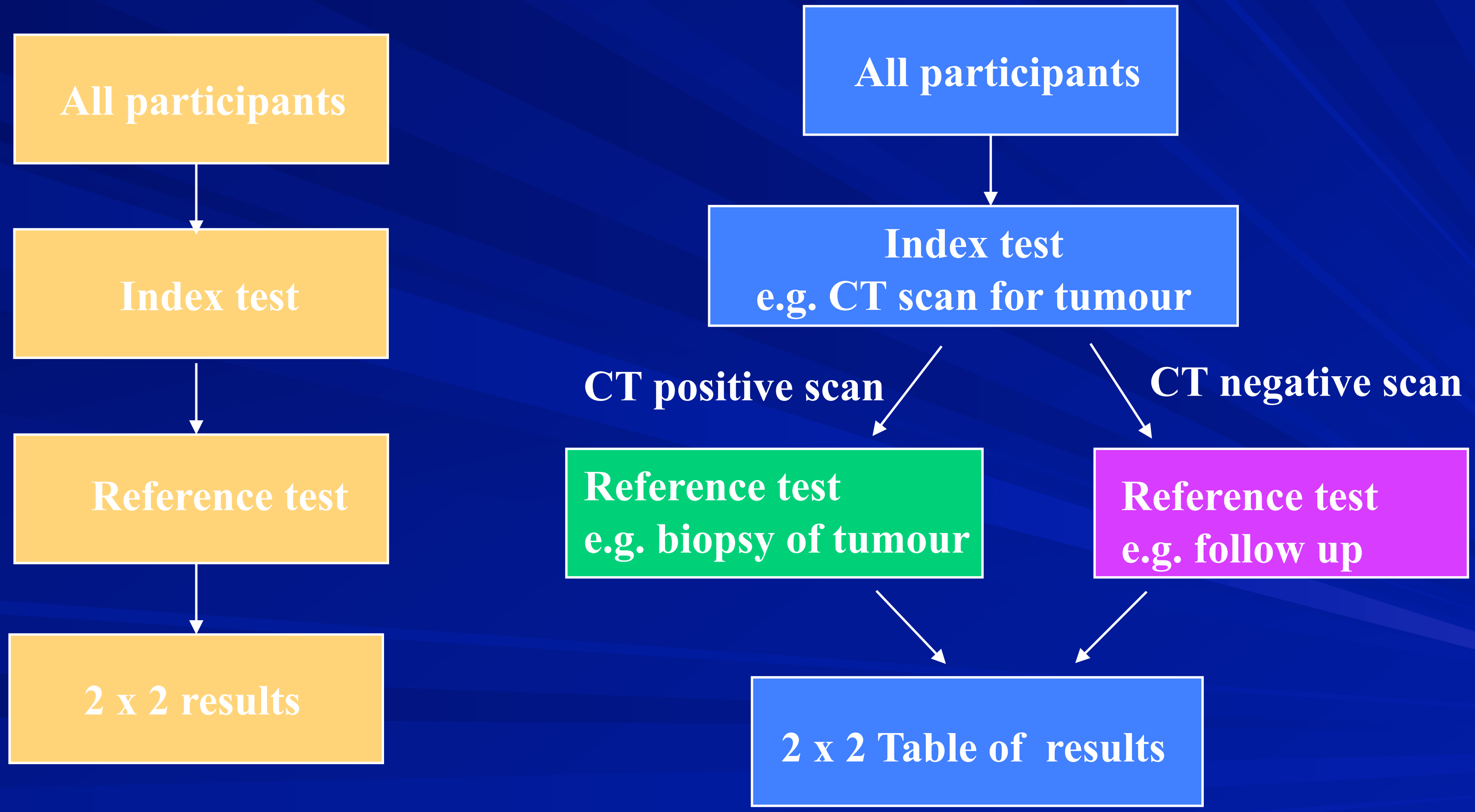
- Is the reference appropriate?
- Was the same reference used for all patients (verification bias)?
- Were assessors blind to case details?
- Was it a ‘diagnostic case-control study’?

See How to read a paper: Papers that report diagnostic or screening tests by Trisha Greenhalgh; BMJ 1997 315 p540

# Differential Verification

**Differential verification often inevitable**

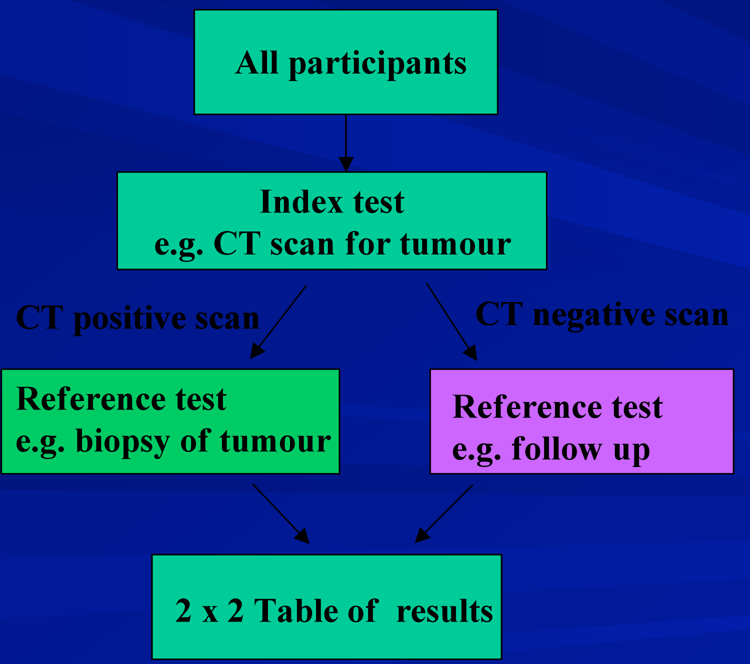
- **biopsy on detected lumps, but follow-up if normal**




Slide from Susan Mallett, Cancer Research UK Medical Statistics Group

# Verification bias

- Are the two reference tests as accurate as each other?
- If not, then get verification bias.
- Different accuracies can be due to different time frames e.g. biopsy today vs follow-up over 2 years. Same cancer?




Slide from Susan Mallett, Cancer Research UK Medical Statistics Group

# Best evidence

- Reporting using STARD guidelines (Standards for Reporting of Diagnostic Accuracy)
- Systematic reviews (Cochrane)
- Use of QUADAS quality checklist
- RCTs that look at effect of test on patient outcome (rare)

# Summary

- All patients must have both new test + reference (gold standard)
- Give sensitivity, specificity with precision
- Test cut-offs in independent sample
- Predictive values vary according to prevalence
- Consider all potential sources of bias

# SCOFF study

- Target population?
- Sample? Representative?
- Reference?
- Cut-off used?
- Potential bias?
- Reporting?