# Basic Statistics in Diagnostic Accuracy Research

Dr. Bethany Shinkins
Test Methodologist/Medical Statistician
23/09/2014
(Acknowledgement to Annette Pluddemann for use of her slides)

Using a brain scan, the researchers detected autism with over 90% accuracy…

You can't diagnose autism with a brain scan…


BBC Mobile News | Sport | Weather | iPlayer | TV

NEWS HEALTH

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Envir

10 August 2010 Last updated at 22:01

# New brain scan to diagnose autism

By Jane Hughes
Health correspondent, BBC News

**A brain scan that detects autism in adults could mean much more straightforward diagnosis of the condition, scientists say.**

Experts at King's College London said the scan - tested on 40 people - identified tiny but crucial signs of autism, only detectable by computer.

Current methods of diagnosis can be lengthy and expensive.

But some experts say further research will be needed before the new technique can be widely used.

The computer scan shows up a distinctive pattern associated with autism

# Assessing the Evidence

Ensuring results are valid

↓

- Appropriate spectrum of patients?
- Does everyone get the gold standard?
- Is there an independent, blind or objective comparison with the gold standard?

# Assessing the Evidence

Ensuring results are valid

↓

Interpreting the results

↓

- Appropriate spectrum of patients?
- Does everyone get the gold standard?
- Is there an independent, blind or objective comparison with the gold standard?

# What we're going to cover

- Key statistics:
  - Sensitivity and specificity
  - ROC curves
  - Likelihood ratios
  - Predictive values

What do all the numbers mean??

# Example

BMJ

*BMJ* 2012;345:e8012 doi: 10.1136/bmj.e8012 (Published 17 December 2012)



RESEARCH

CHRISTMAS 2012: RESEARCH

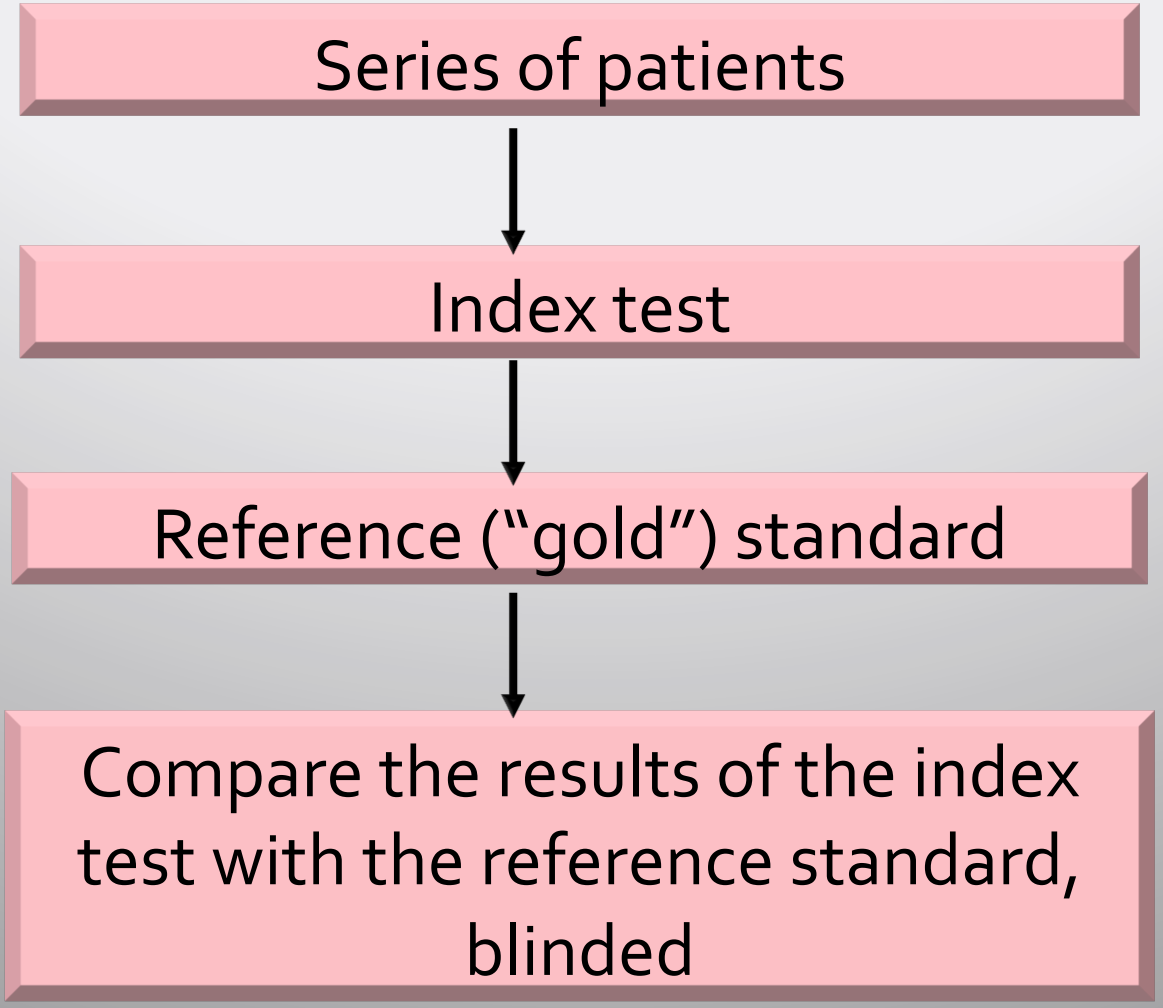
## Pain over speed bumps in diagnosis of acute appendicitis: diagnostic accuracy study



Helen F Ashdown *academic clinical fellow in general practice*[1], Nigel D'Souza *specialist registrar in general surgery*[2], Diallah Karim *foundation trainee*[2], Richard J Stevens *senior medical statistician*[1], Andrew Huang *consultant colorectal and general surgeon*[2], Anthony Harnden *university lecturer in general practice*[1]

[1]Department of Primary Care Health Sciences, University of Oxford, Oxford OX2 6GG, UK; [2]Department of Surgery, Stoke Mandeville Hospital, Aylesbury HP21 8AL, UK

# Study Design



# The Results

of appendicitis. Fifty four of 64 participants were "speed bump positive." Thirty four participants had a confirmed diagnosis of appendicitis, 33 of whom had worsened pain over speed bumps, giving a sensitivity of 97% (85% to 100%) and a specificity of 30% (15% to 49%). The positive predictive value was 61% (47% to 74%), and the negative predictive value was 90% (56% to 100%). The likelihood ratios were 1.4 (1.1 to 1.8) for a positive test result and 0.1 (0.0 to 0.7) for a negative result.

# The 2 x 2 Table

| Test | Disease + | Disease – |
|---|---|---|
| + | **True Positives** | **False Positives** |
| – | **False Negatives** | **True Negatives** |

# Sensitivity and Specificity

# Sensitivity

| Test \ Disease | + | − |
| --- | --- | --- |
| + | 84 a True positives | |
| − | 16 c False negatives | |

Proportion of people **WITH** the disease who have a **positive test result**

So, a test with 84% sensitivity….means that the test identifies 84 out of 100 people **WITH** the disease

Sensitivity = a / a + c

Sensitivity = 84/100

# Specificity

| Test \ Disease | + | − |
|---|---|---|
| + | | 25 b False positives |
| − | | 75 d True negatives |

Proportion of people **WITHOUT** the disease who have a **negative test result**

So, a test with 75% specificity will be **NEGATIVE** in 75 out of 100 people **WITHOUT** the disease

Specificity = d / b + d

Specificity = 75/100

# Speed Bump Example

Disease: Appendicitis

Test: Pain over speed bump

| | + | – | |
|---|---|---|---|
| **+** | 33 | 21 | 54 |
| **–** | 1 | 9 | 10 |
| | 34 | 30 | 64 |

*There were 34 people who had appendicitis… the speed bump test was positive in 33 of them*

*There were 30 people who did not have appendicitis… the speed bump test was negative in 9 of them*

Sensitivity = 33/34 = 0.97 (97%)

Specificity = 9/30 = 0.30 (30%)

# Tip

- **Sensitivity** makes sense
  - 'The new speed bump test was positive in 33 out of 34 people with appendicitis (sensitivity = 97%)'
- **Specificity** seems a bit confusing!
  - 'The new speed bump test was negative in 9 of the 30 people who did not have appendicitis (specificity = 30%)'
- So…the **false positive rate** is sometimes easier

**False positive rate = 1 - specificity**

  - 'There were 30 people who did not have appendicitis… the speed bump test was falsely positive in 21 of them'
  - So a specificity of 30% means that the new rapid test is wrong (or falsely positive) in 70% of people without the disease

# Ruling In and Ruling Out

**High Sensitivity** → A good test to help **Rule Out** disease

SnNOUT

High sensitivity means there are very **few false negatives** – so if the test comes back negative it's highly unlikely the person has the disease

**High Specificity** → A good test to help **Rule In** disease

SpPIN

High specificity means there are very **few false positives** – so if the test comes back positive it's highly likely the person has the disease

Disease: Appendicitis

Test: Pain over speed bump

| Test \ Disease | + | − |
|---|---|---|
| + | 33 | 21 |
| − | 1 | 9 |

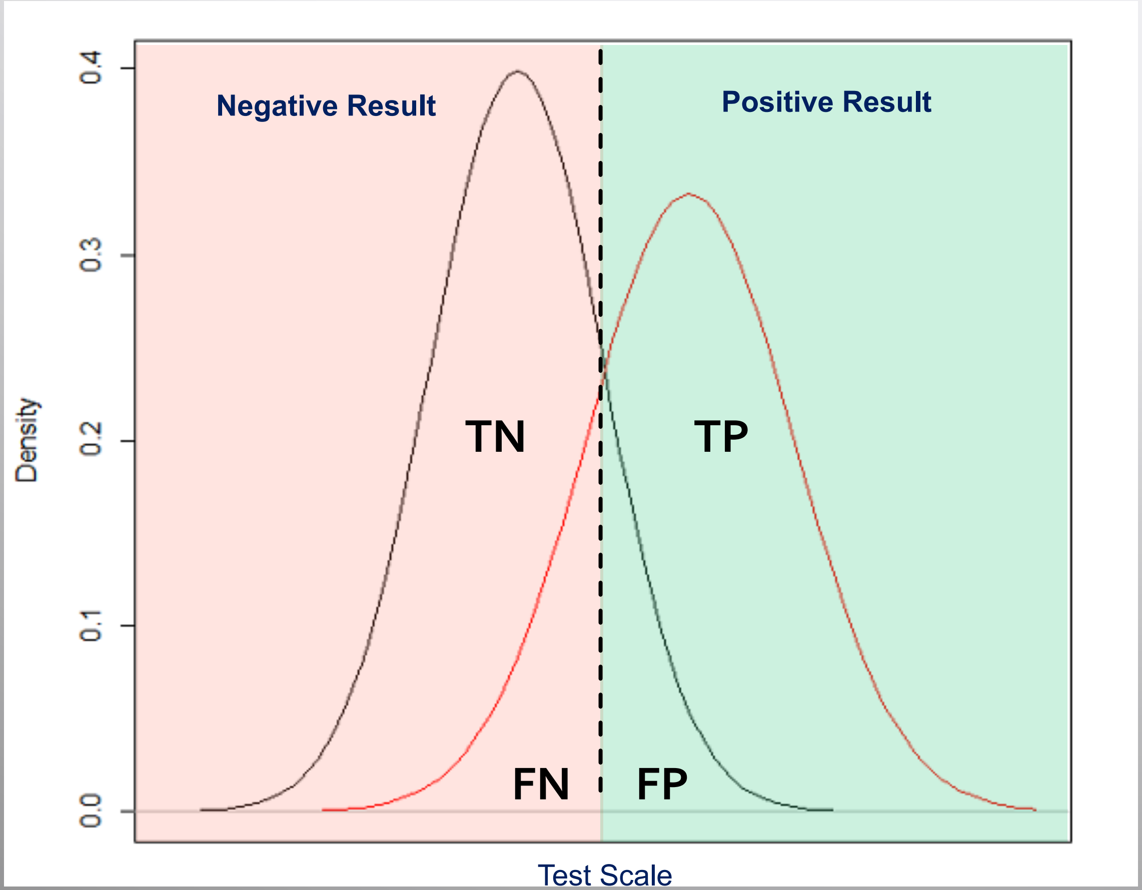
Sensitivity = 97% | Specificity = 30%

| Test \ Disease | + | − |
|---|---|---|
| + | a<br>True positives | b<br>False positives |
| − | c<br>False negatives | d<br>True negatives |

Sensitivity = a/a+c

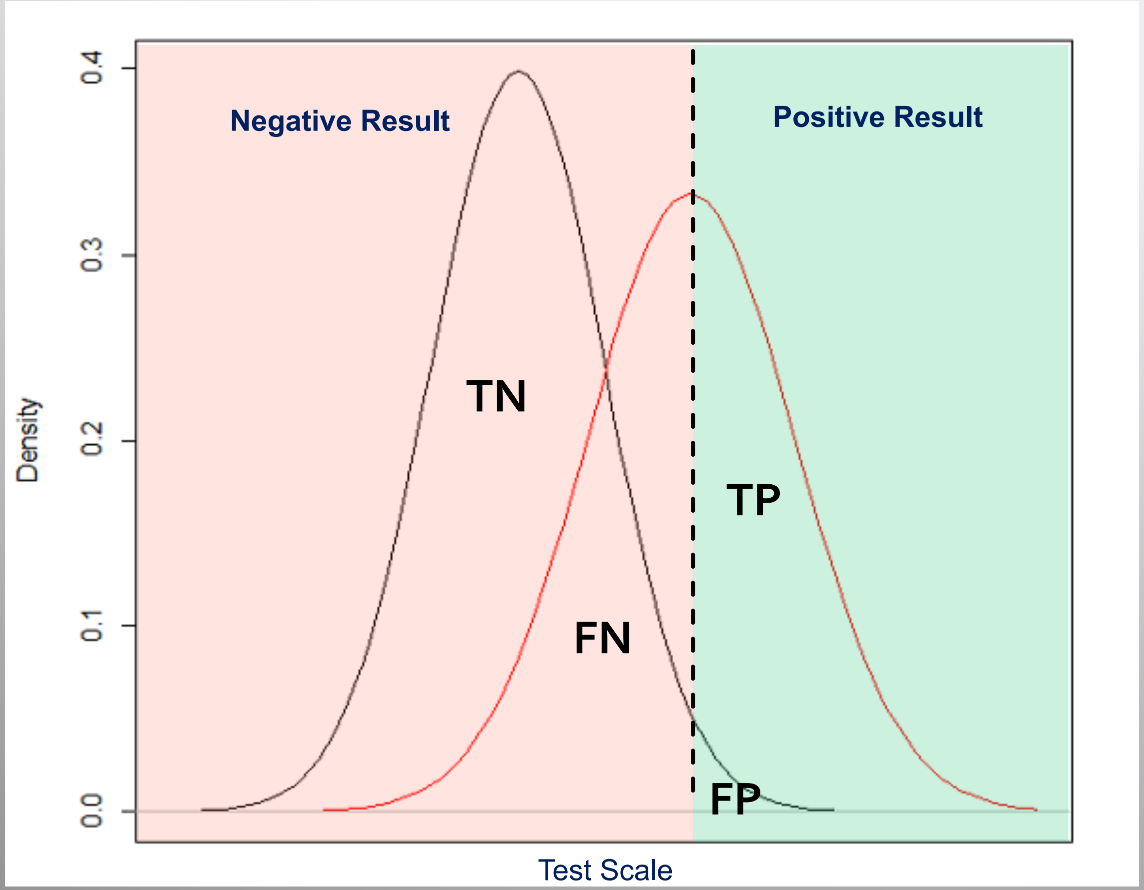
Specificity = d/b+d

# Threshold trade-off



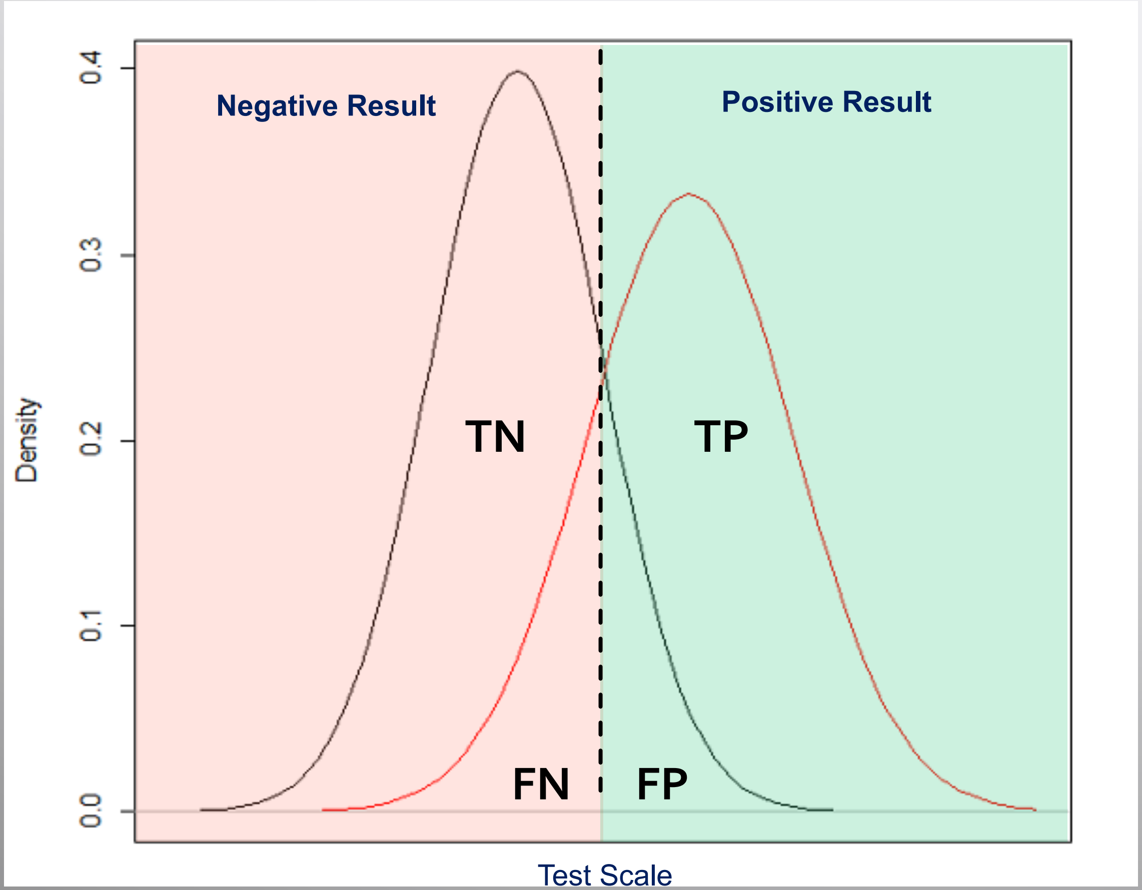
# Threshold trade-off

Move threshold to the right



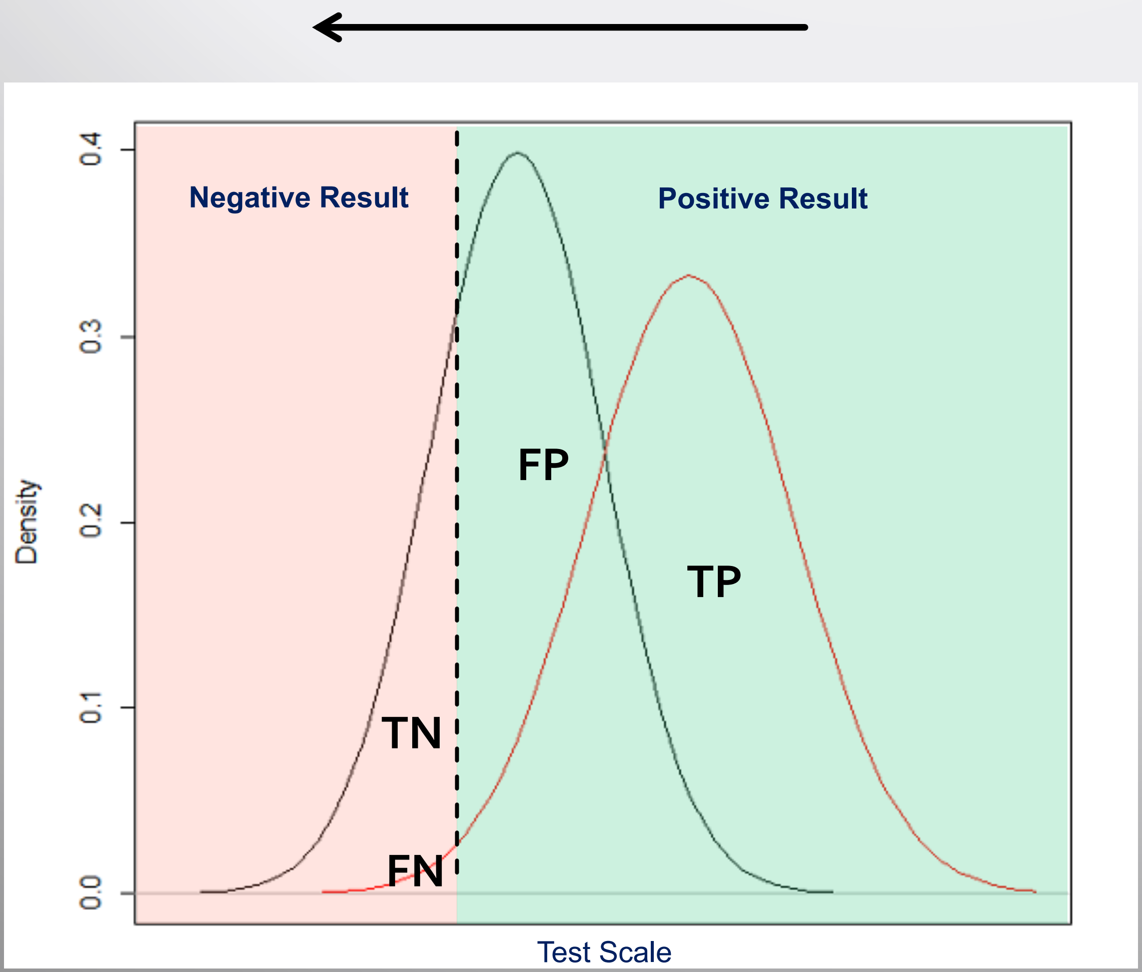

- Reduces number of false-positives
- Therefore higher specificity
- At the cost of reduced sensitivity

# Threshold trade-off



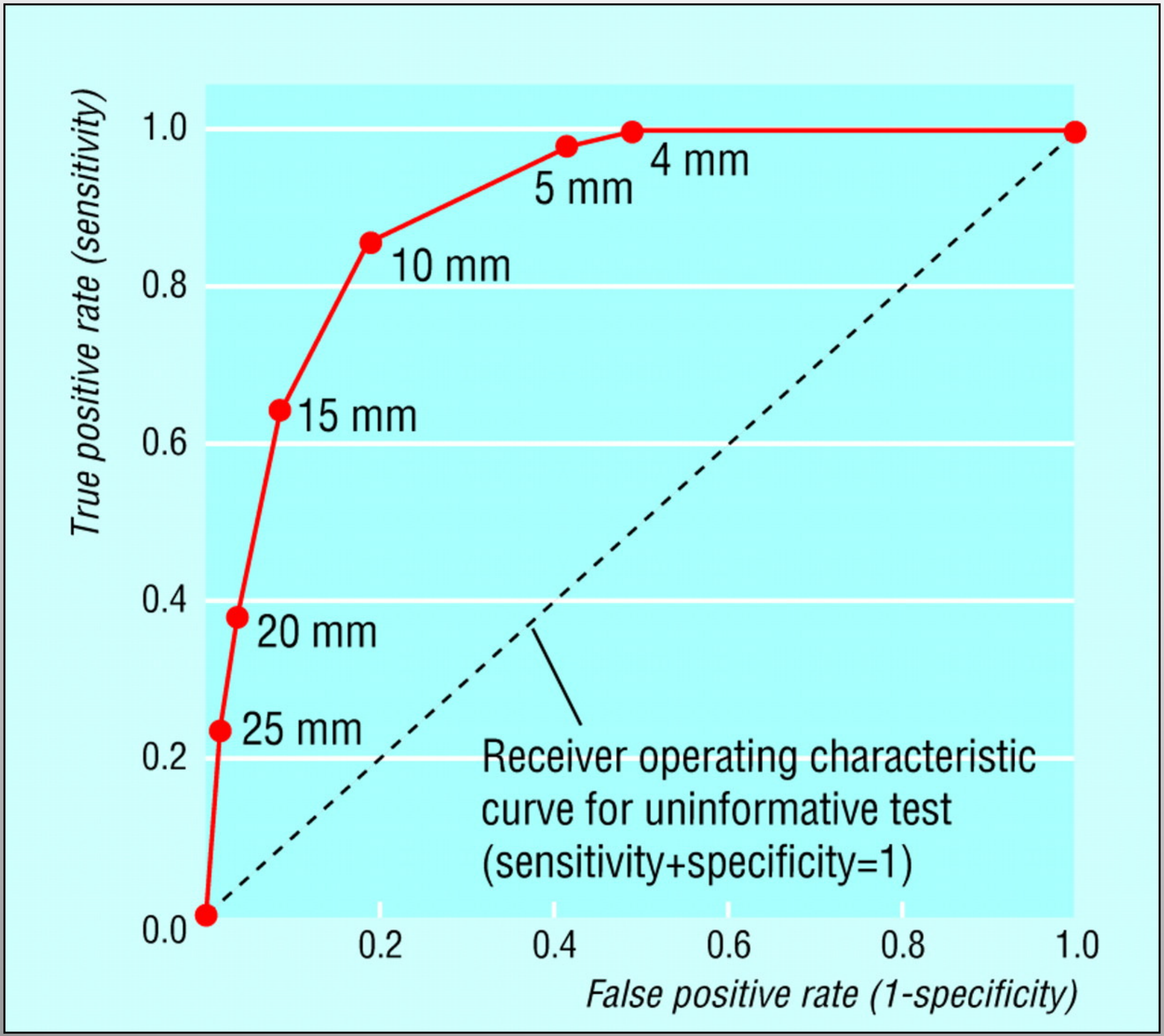
# Threshold trade-off

Move threshold to the left




- Reduces number of false-negatives
- Therefore higher sensitivity
- At the cost of reduced specificity

# ROC Curves




Area under the curve (AUC):

Threshold-independent method for comparing test accuracy head-to-head

*Systematic reviews of evaluations of diagnostic and screening tests. J Deeks. BMJ 2001; 323.*

# Predictive Values

# Positive and Negative Predictive Values

| Test | Disease + | Disease − |
|---|---|---|
| + | a<br>**True positives** | b<br>**False positives** |
| − | c<br>**False negatives** | d<br>**True negatives** |

**PPV = Proportion of people with a <u>positive test</u> who <u>have</u> the disease**

$$PPV = a / a + b$$

$$NPV = d / c + d$$

**NPV = Proportion of people with a <u>negative test</u> who <u>do not</u> have the disease**

# Speed Bump Example

Disease: Appendicitis

Test: Pain over speed bump

| | + | – | |
|---|---|---|---|
| **+** | 33 | 21 | 54 |
| **–** | 1 | 9 | 10 |
| | 34 | 30 | 64 |

**PPV = 33/54 = 61%**

**NPV = 9/10 = 90%**

# Natural Frequencies

Your father went to his doctor and was told that his test for a disease was positive. He is really worried, and comes to ask you for help!

After doing some reading, you find that for men of his age:

The prevalence of the disease is 30%

The test has sensitivity of 50% and specificity of 90%

*"Son, tell me what's the chance I have this disease?"*

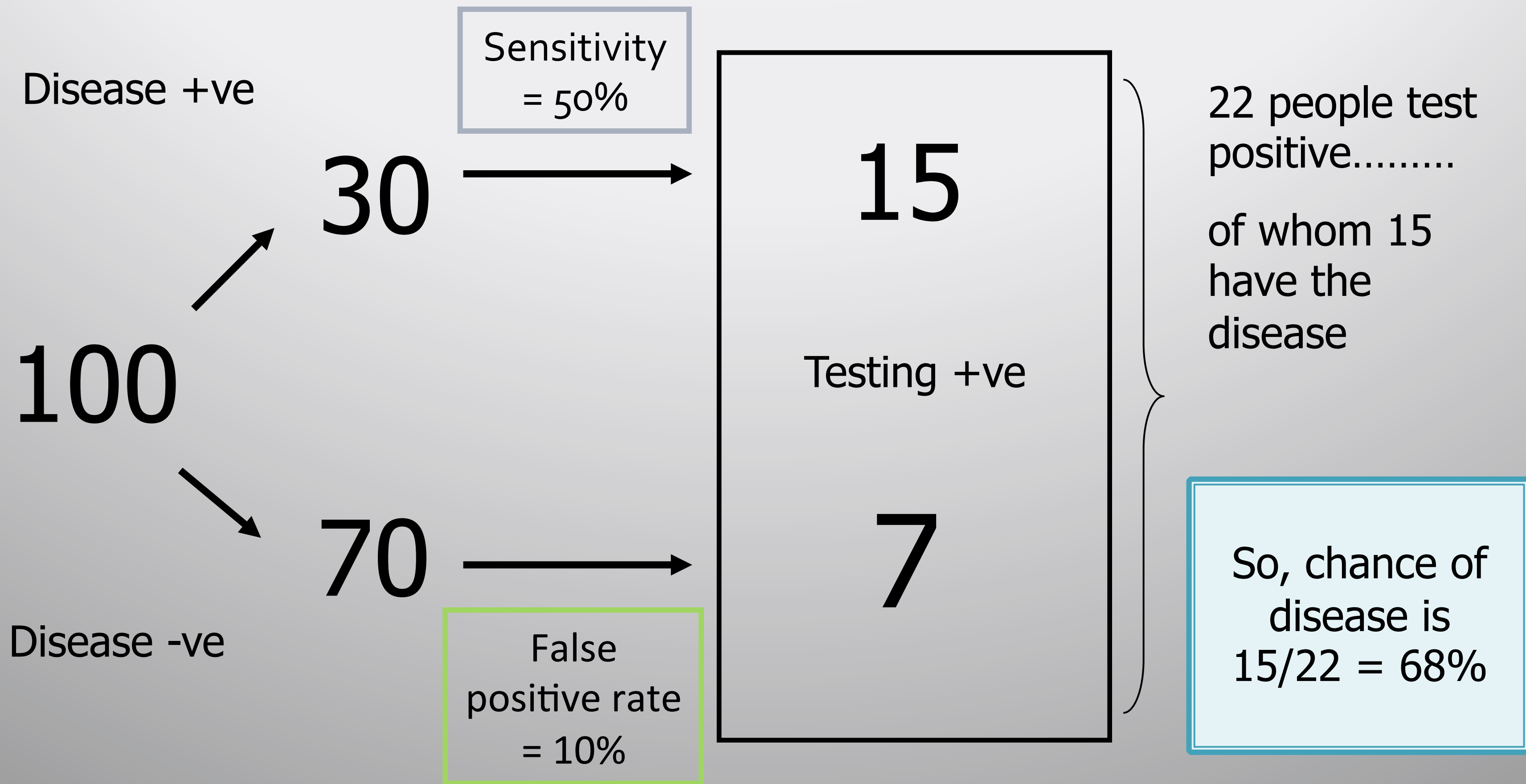
# Predictive Value

Disease has a prevalence of 30%

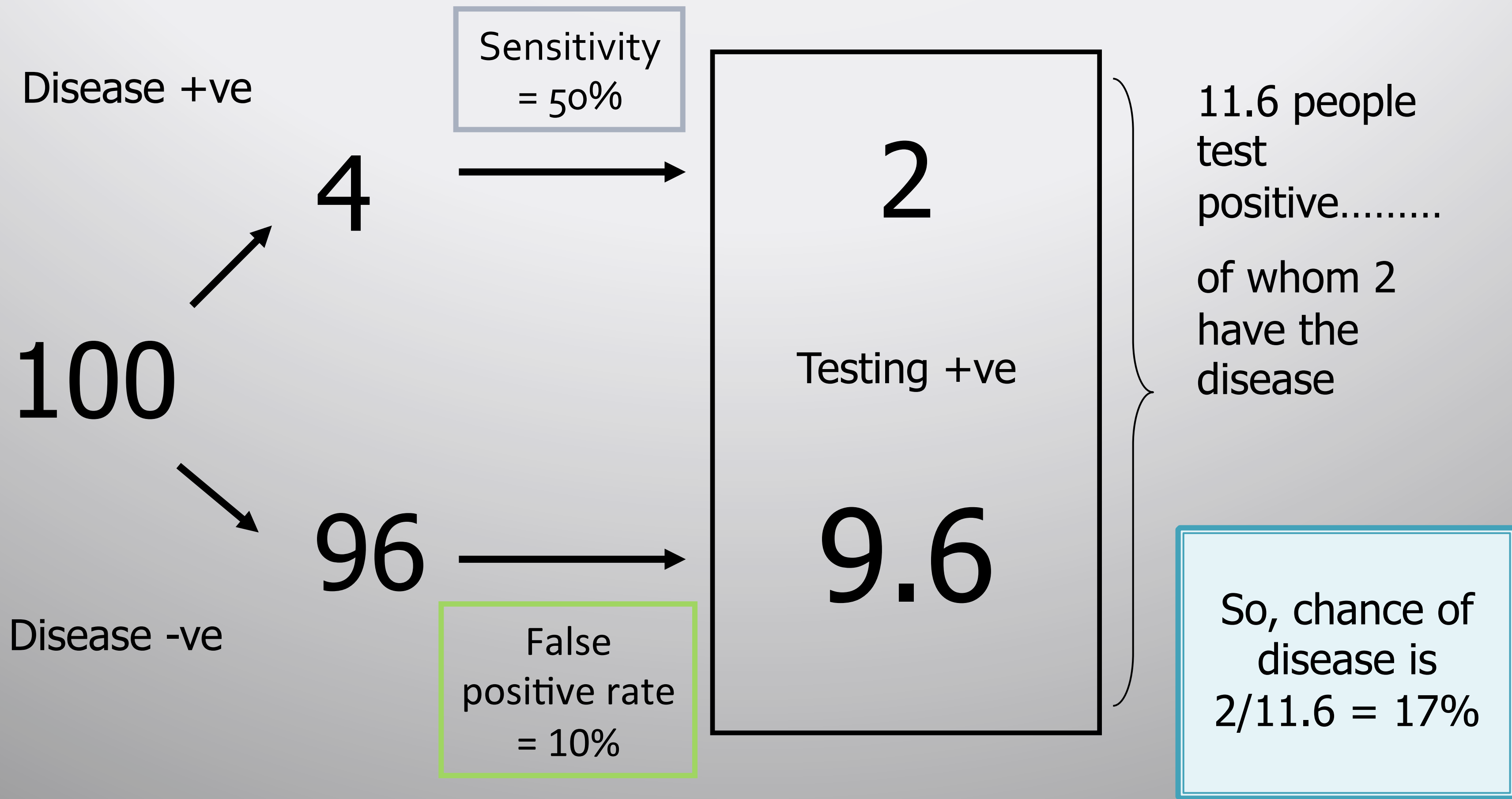
The test has sensitivity of 50% and specificity of 90%

- 100% — Likely
- 50% — Maybe
- 0% — Unlikely

# Prevalence of 30%, sensitivity of 50%, specificity of 90%



# Prevalence of 4%, sensitivity of 50%, specificity of 90%



# Positive and Negative Predictive Value

**NOTE**

- PPV and NPV are not intrinsic to the test – they also depend on the prevalence!

- NPV and PPV should only be used **if the ratio of the number of patients in the disease group and the number of patients in the healthy control group is equivalent to the prevalence** of the disease in the studied population

- Use Likelihood Ratio - does not depend on prevalence

# Likelihood Ratios

# What is a likelihood ratio?

LR = *Test result in those with disease/Test result in those without disease*

Three possible results:

- LR greater than 1 — ↑ With disease
- LR equal to 1 — = No change
- LR less than 1 — ↓ Without disease

# Let's take a LR of 5

LR is greater than one…

…which means it is more likely to occur in those with disease

How much larger than 1? 5 times larger

So, a LR of 5 means that the test result occurs *five times* more often in those with disease than in those without

# Likelihood Ratios

**Positive likelihood ratio (LR+)**

How much more likely is a positive test result to be found in a person with the disease than in a person without it?

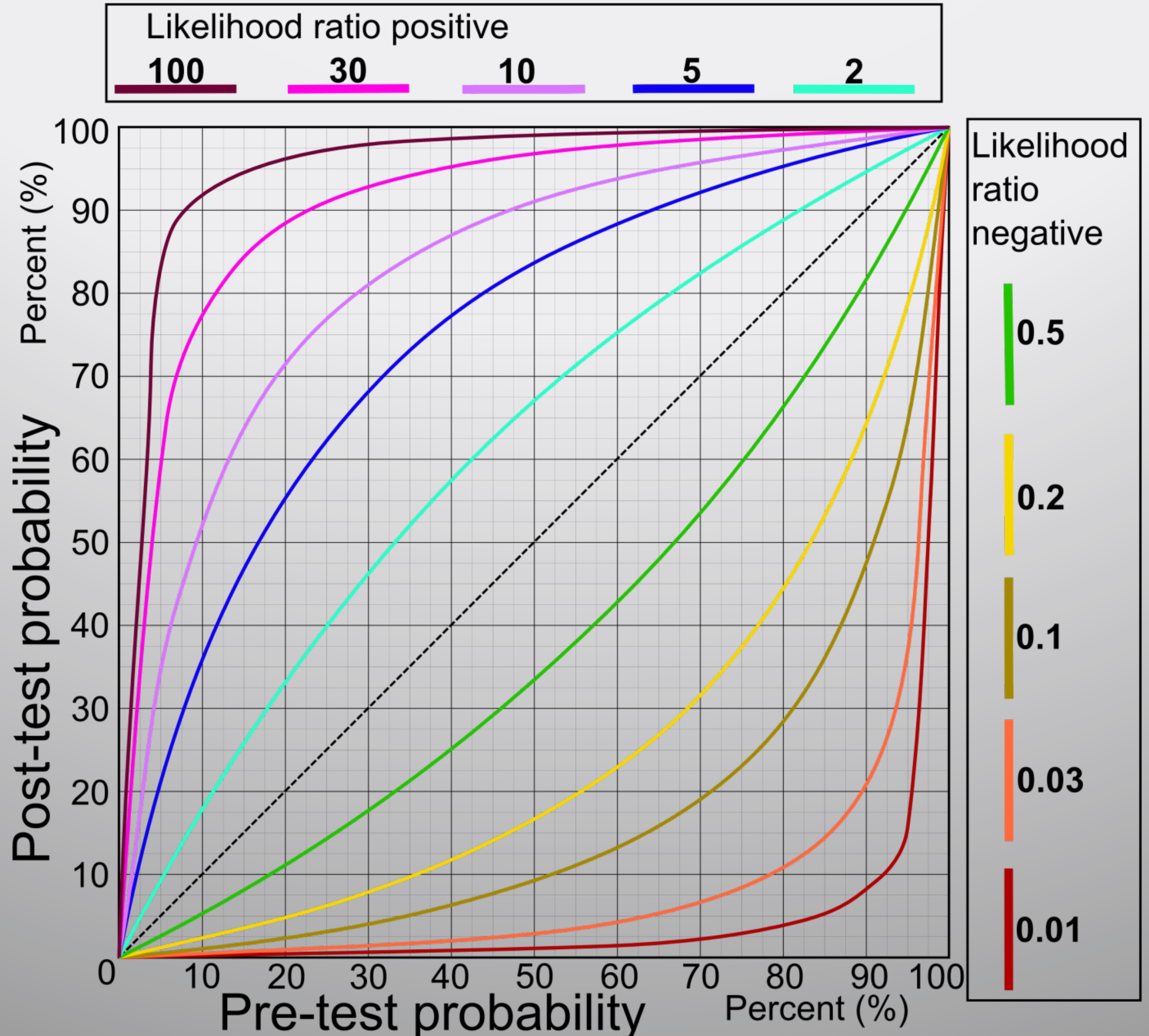
**LR+ = sens/(1-spec)**

**Negative likelihood ratio (LR-)**

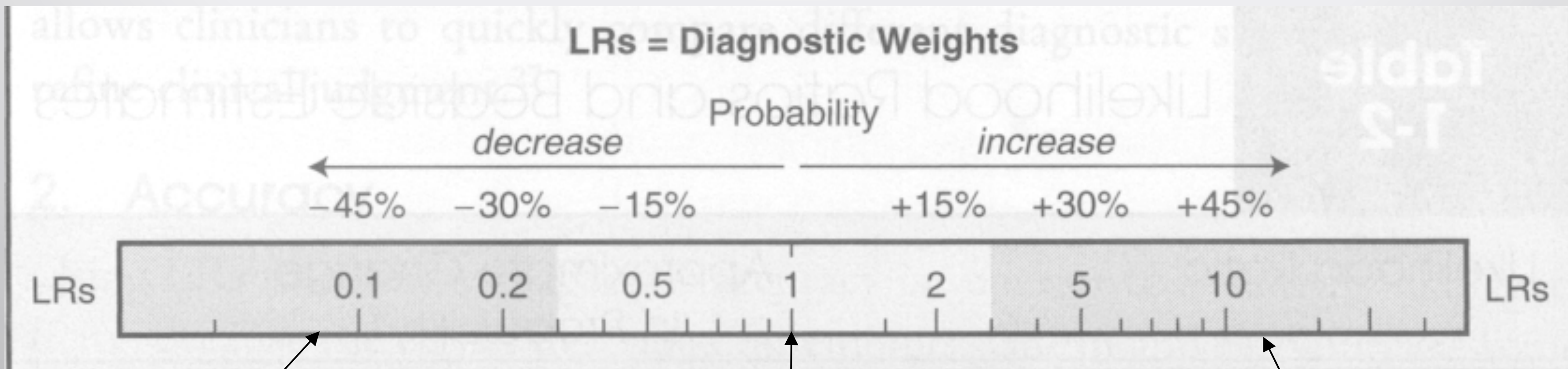
How much more likely is a negative test result to be found in a person without the disease than in a person with it?

**LR- = (1-sens)/(spec)**

# Likelihood ratios directly 'adjust' probability of disease



# Rule of thumb




LR<0.1 = strong negative test result

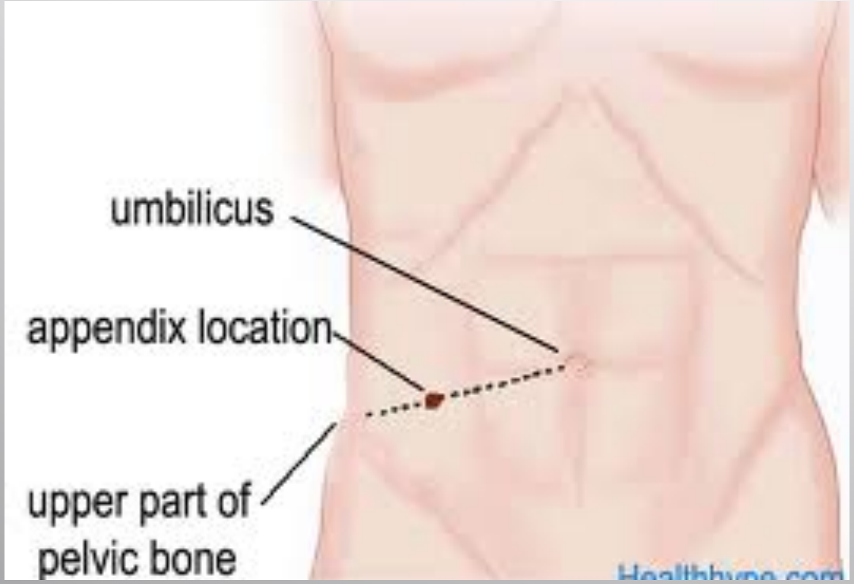
LR=1

No diagnostic value

LR>10 = strong positive test result

# Diagnosis of Appendicitis

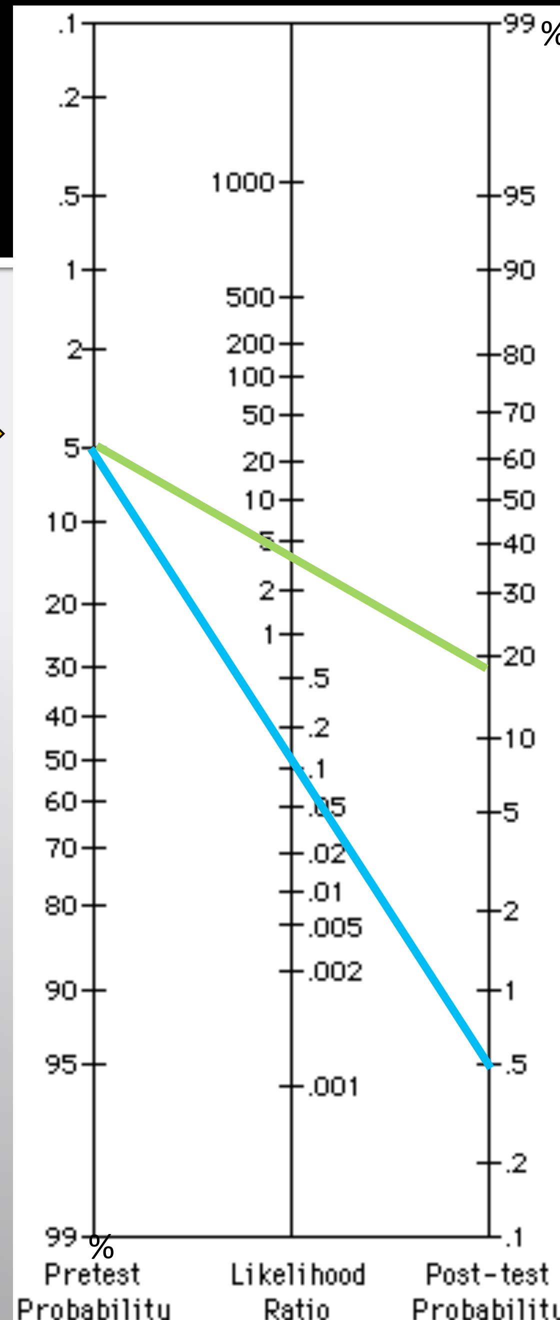
**McBurney's point**
Tenderness right side of abdomen




**Ashdown's sign**
Pain when driving over speed bumps

# Bayesian reasoning

Pre test 5%




Post-test odds for disease after *one* test become pre-test odds for *next* test etc.

Appendicitis:

McBurney tenderness LR+ = 3.4

Post test ~20%

Speed bump test LR- = 0.1

Post test ~0.5%

Post-test odds = Pre-test odds x Likelihood ratio

# Multilevel Likelihood Ratios

| Serum creatinine (micromoles/L) | Stenosis | No stenosis | All | Likelihood ratio (95% CI) |
|---|---|---|---|---|
| ≤ 60 | 1 (1%) | 19 (6%) | 20 (5%) | 0.18 (0.02–1.31) |
| 61–70 | 4 (4%) | 36 (11%) | 40 (9%) | 0.37 (0.14–1.03) |
| 71–80 | 13 (13%) | 67 (20%) | 80 (18%) | 0.65 (0.38–1.13) |
| 81–90 | 12 (12%) | 71 (21%) | 83 (19%) | 0.57 (0.32–1.01) |
| 91–100 | 17 (17%) | 71 (21%) | 88 (20%) | 0.81 (0.50–1.30) |
| 101–110 | 15 (15%) | 41 (12%) | 56 (13%) | 1.23 (0.71–2.13) |
| 111–120 | 7 (7%) | 10 (3%) | 17 (4%) | 2.33 (0.92–6.04) |
| 121–130 | 9 (9%) | 9 (3%) | 18 (4%) | 3.33 (1.37–8.26) |
| 131–150 | 11 (11%) | 8 (2%) | 19 (4%) | 4.58 (1.92–11.20) |
| > 150 | 11 (11%) | 5 (1%) | 16 (4%) | 7.33 (2.64–20.84) |
| All | 100 (100%) | 337 (100%) | 437 (100%) | |

**Multilevel likelihood ratios for serum creatinine concentration for the diagnosis of renal artery stenosis**

*Knottnerus (2002) The Evidence Base of Clinical Diagnosis*

# Back to Carl…

BBC Mobile

News | Sport | Weather | iPlayer | TV

NEWS HEALTH

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Envir

10 August 2010 Last updated at 22:01

# New brain scan to diagnose autism

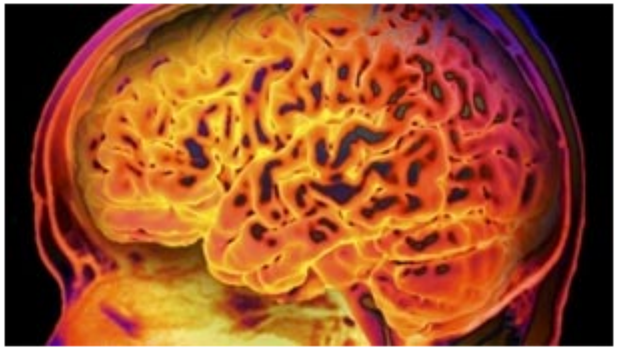
By Jane Hughes
Health correspondent, BBC News

**A brain scan that detects autism in adults could mean much more straightforward diagnosis of the condition, scientists say.**

Experts at King's College London said the scan - tested on 40 people - identified tiny but crucial signs of autism, only detectable by computer.

Current methods of diagnosis can be lengthy and expensive.

But some experts say further research will be needed before the new technique can be widely used.



The computer scan shows up a distinctive pattern associated with autism

The researchers detected autism with over 90% accuracy, the Journal of Neuroscience reports.

The National Autistic Society

Estimated prevalence rate in the UK

The indication from recent studies is that the figures cannot be precisely fixed, but it appears that a prevalence rate of around 1 in 100 is a best estimate a best estimate of the prevalence in children. No prevalence studies have ever been carried out on adults.



**Table 3. Results of SVM classification between ASD and control group using different brain morphometric features in the left and right hemispheres**

| Morphometric feature | Correctly classified (%) | Sensitivity (%) | Specificity (%) | *p* |
|---|---|---|---|---|
| Left hemisphere | | | | |
| All parameters | 85 | 90 | 80 | 0* |
| Cortical thickness | 90 | 90 | 90 | 0* |
| Radial curvature | 72.5 | 65 | 80 | <0.001 |
| Average convexity | 70 | 75 | 65 | <0.004 |
| Metric distortion | 80 | 80 | 80 | 0* |
| Pial area | 77.5 | 70 | 85 | 0* |
| Right hemisphere | | | | |
| All parameters | 65 | 60 | 70 | <0.03 |
| ...ckness | 60 | 65 | 55 | <0.01 |
| ...ature | 52.5 | 50 | 55 | <0.30 |
| ...nvexity | 50 | 40 | 60 | <0.40 |
| ...ortion | 57.5 | 45 | 70 | <0.06 |
| | 45 | 45 | 45 | <0.60 |

...ed ASD cases were considered true positive. **p* values of zero indicate that not a single one of the ...ns provided a better classification.

Neurobiology of Disease

# Describing the Brain in Autism in Five Dimensions—Magnetic Resonance Imaging-Assisted Diagnosis of Autism Spectrum Disorder Using a Multiparameter Classification Approach

Christine Ecker,[1] Andre Marquand,[2] Janaina Mourão-Miranda,[3,4] Patrick Johnston,[1] Eileen M. Daly,[1] Michael J. Brammer,[2] Stefanos Maltezos,[1] Clodagh M. Murphy,[1] Dene Robertson,[1] Steven C. Williams,[3] and Declan G. M. Murphy[1]

[1]Section of Brain Maturation, Department of Psychological Medicine, Institute of Psychiatry, [2]Brain Image Analysis Unit, Department of Biostatistics, Institute of Psychiatry, and [3]Centre for Neuroimaging Sciences, Institute of Psychiatry, King's College, London SE5 8AF, United Kingdom, and [4]Centre for Computational Statistics and Machine Learning, Department of Computer Science, University College London, London WC1E 6BT, United Kingdom

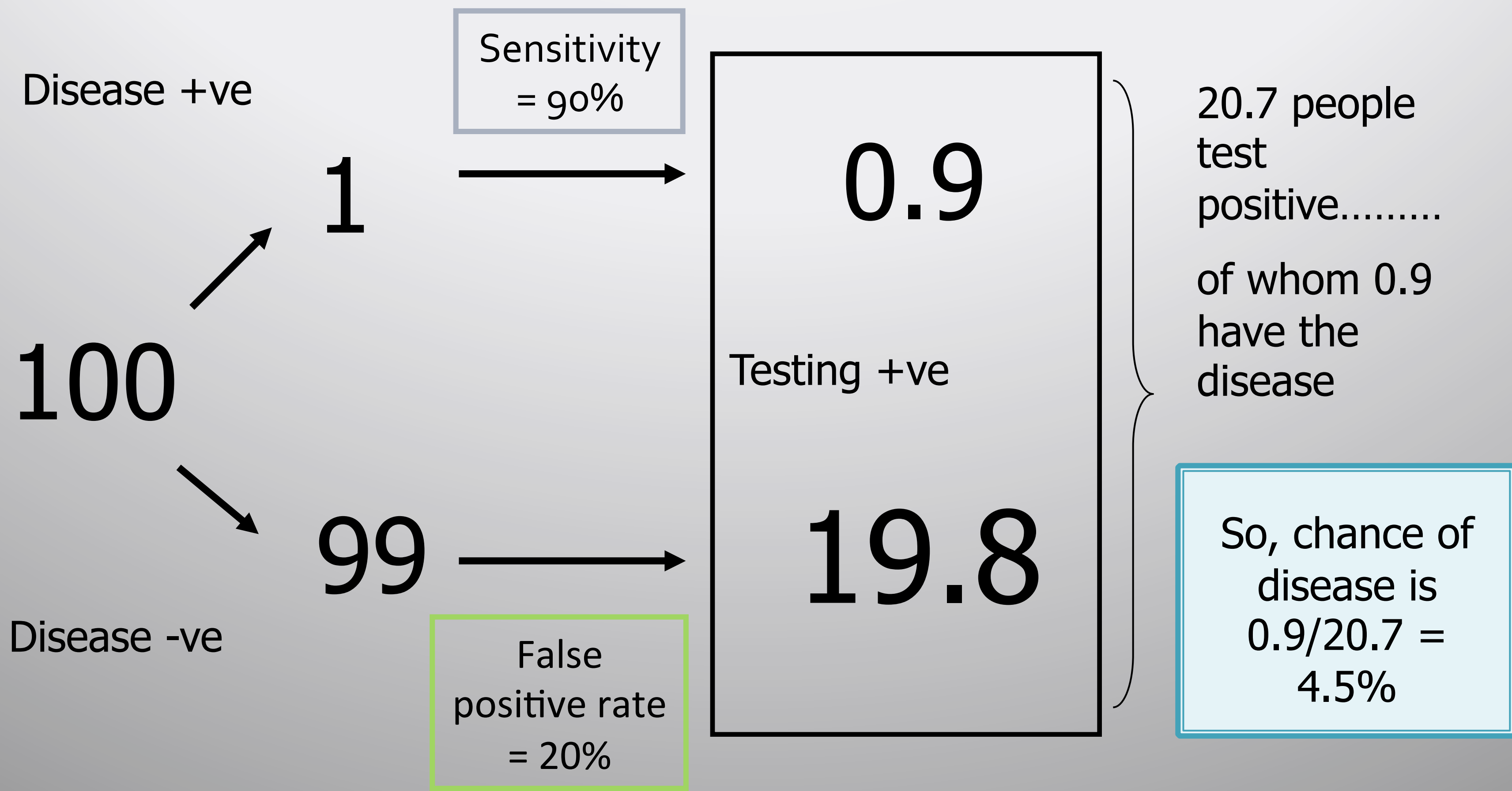
# Natural Frequencies

Autism has a prevalence of 1%.

The test has sensitivity of 90% and specificity of 80%.

Given a positive test, what is the probability the child has autism?

# Prevalence of 1%, sensitivity of 90%, specificity of 80%



CEBM Homepage | About TrustTheEvidence.net

The first place... the last word.


**TRUSTTHEEVIDENCE.NET**

Discover the truth behind the research findings that affect everyday healthcare.



**HONcode Certified**

HON CODE CERTIFIED 09/2010

This site complies with the HONcode standard for trustworthy health information: verify here.

**Bloggers**

**Carl Heneghan**

Director of the CEBM, G
clinical lecturer at the
University of Oxford.

# autism and brain scan test: the real predictive value

Carl Heneghan

Posted 11th August 2010 @ 05:27pm

**Navigator**

Search TrustTheEvidence.net: Search

Twitter TrustTheEvidence.net

What has happened is the sensitivity has been taken for the positive predictive value, which is what you want to know: if I have a positive test do I have the disease?

Sensitivity: The proportion of people with disease who have a positive test. Positive predictive value (+PV): The proportion of people with a positive test who have disease.

So, for a prevalence of 1% the actual positive predictive value is 4.5%. That is about 5 in every 100 with a positive test would have autism. Even at a prevalence of 2%, only 8.5% would be correctly identified.

Suddenly, not that great a test. This has to be one of the worst examples of misinterpreting diagnostic test results in the media I've ever seen.

NOTES&THEORIES

DISPATCHES FROM THE SCIENCE DESK

Previous | Blog home | Next

# Why autism can't be diagnosed with brain scans

Using brain scans to detect autism would be a huge expensive waste of money, says Carl Heneghan



Posted by
Carl Heneghan Thursday
12 August 2010
15.29 BST
guardian.co.uk

The BBC, the Guardian and Reuters this week widely reported British researchers published in the Journal of Neuroscience have developed a brain scan which can detect autism in adults with 90% accuracy.

Dr Christine Ecker, the lead author, showed her imaging technique was able to detect which people in her group had autism. "If we get a new case, we will also hopefully be 90% accurate," she said.

Pretty simple then, you turn up, have the test, and you have a 90% chance of finding out whether you have autism.

Well, you couldn't be any further from the truth.

A larger | smaller

Science
Medical research ·

# Beyond Accuracy

Are the results valid?

- Appropriate spectrum of patients?
- Does everyone get the gold standard?
- Is there an independent, blind or objective comparison with the gold standard?

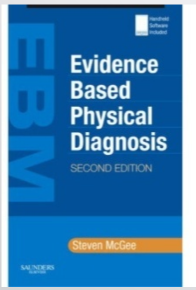
What are the results?

- Sensitivity, specificity
- Likelihood ratios
- Positive and Negative Predictive Values

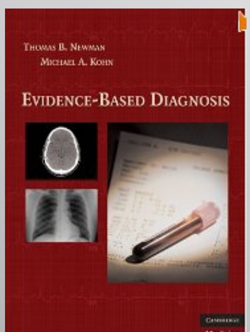
Will they help me look after my patients?

- Can I do the test in my setting?
- Do results apply to the mix of patients I see?
- Will the result change my management?
- Costs to patient/health service?

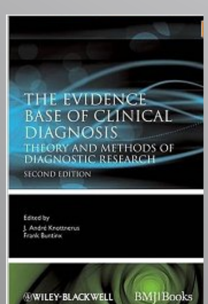
# Useful books on diagnostics



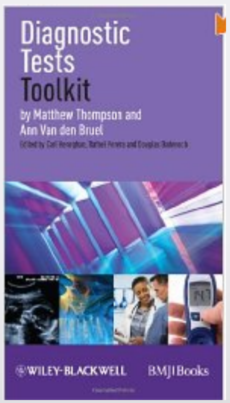

Evidence based
Physical Diagnosis.
Steven McGee.
Saunders



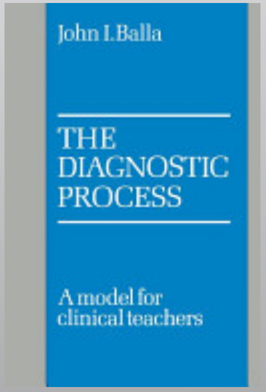

Evidence-based
Diagnosis.
Newman & Kohn.
Cambridge Univ. Press




Evidence base of Clinical
Diagnosis.
Knottnerus & Buntinx.
Wiley-Blackwell




Diagnostic Tests Toolkit.
Thompson & Van den Bruel.
Wiley-Blackwell.




The Diagnostic Process.
John Balla.
Cambridge Univ. Press

# Useful Journal Articles on Diagnostics

- Bossuyt. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making 2009
- Heneghan et al. Diagnostic strategies used in primary care. BMJ 2009
- Ferrante di Ruffano. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. BMJ 2012
- Mallett et al. Interpreting diagnostic accuracy studies for patient care. BMJ 2012
- Bossuyt et al. STARD initiative. Ann Int Med 2003
- Lord et al. Using prinicIples of RCT design to guide test evaluation. Med Decis Making 2009
- Rutjes et al. Evidence of bias and variation in diagnostic accuracy studies. CMAJ 2006
- Lijmer et al. Proposals for phased evaluation of medical tests. Med Decis Making 2009
- Whiting et al. QUADAS-2: revised tool for quality assessment of diagnostic accuracy studies. Ann Int Med 2011

# A bit of practice…

Disease:

| Test: | | + | – |
|---|---|---|---|
| | + | 45 | 40 |
| | – | 5 | 60 |

# What if we adjust the prevalence to 10%?