Epidemiology in Practice

Dr John Chambers
Reader in Cardiovascular Epidemiology
Imperial College London

Overview

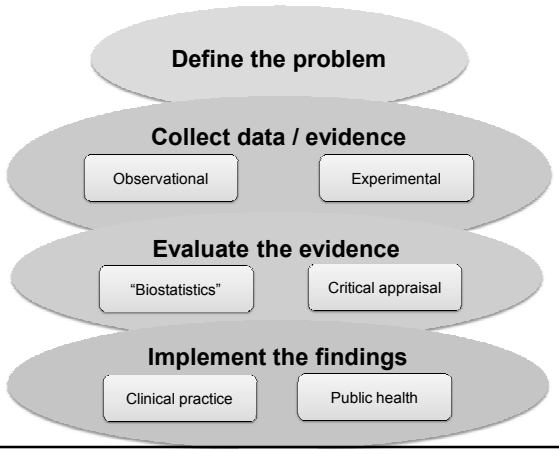
- What is epidemiology?
- What are the core components?
- Why is epidemiology important?

What is Epidemiology?

"The study of the distribution and determinants of health related states or events in specified populations **and** the application of this study to control health problems - to promote, protect and restore health"

Epidemiology = the study of groups of people

What are the core components?



Why is Epidemiology important?

Epidemiology | **Clinical Medicine**

Sir Richard Doll, epidemiologist who found the link between smoking and cancer, CVD etc has "saved millions of lives – the number of lives globally that he helped to save is incalculable" (Maurice Slevin, MD, FRCP, Chairman of CancerBACUP)

One extra doctor, on average, results in a reduction of 14.4 deaths, per 100,000 population (*BMJ, 2005*)

Epidemiology is central to research and delivery in clinical medicine

- Identifying disease patterns
- Insights into disease causation
- Experimental evaluation of interventions
- Identification of high risk people
- Guiding clinical and public health strategy

Course structure

- Patterns of disease
  - Global and National
  - Mortality and Morbidity
- Data and evidence
  - Observational data
  - Experimental data
- Evaluation of data
  - Biostatistics
  - Critical appraisal
- Health promotion / disease prevention
  - Screening
  - Health promotion

Timetable

- Monday 27th February: 15.00-17.00
  - Evaluation 1: Association, causation, appraisal
- Tuesday 28th February: 9.00-12.00
  - National and Global Health
  - Evaluation 2: Biostatistics
- Monday 8th March: 14.00-17.00
  - Small group tutorial
- Friday 12th March: 9.00 -12.00
  - Evidence 1: Routine data
  - Evidence 2: Clinical trials
  - Evidence 3: Case-control and cohort studies
- Wednesday 16th March: 10.00-12.00
  - Epidemiology at work: Screening

Reading

The Epidemiological Approach
Nicholas J Wald

Paperback - **£11.99**

ISBN: 9781853155840
Published: 27/01/2004

Extent: 88 pages

Any problems/questions

john.chambers@ic.ac.uk

Global Health

GLOBAL HEALTH RISKS

Mortality and burden of disease attributable to selected major risks

World Health Organization

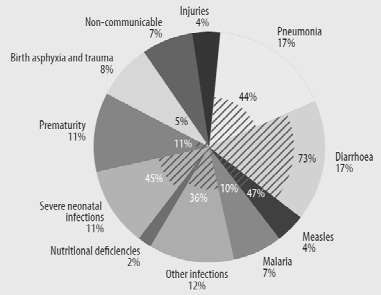
http://www.who.int/healthinfo/global_burden_disease/global_health_risks

Global causes of mortality

2002					2030
Disease or injury	% total deaths	Rank	Rank	% total deaths	Disease or injury
Ischaemic heart disease	12.6%	1	1	13.1%	Ischaemic heart disease
Cerebrovascular disease	9.7%	2	2	10.3%	Cerebrovascular disease
Lower respiratory infections	6.9%	3	3	8.7%	HIV/AIDS
HIV/AIDS	4.8%	4	4	7.9%	Chronic obstructive pulmonary disease
Chronic obstructive pulmonary disease	4.8%	5	5	3.5%	Lower respiratory infections
Perinatal conditions	4.3%	6	6	3.1%	Diabetes mellitus
Diarrhoeal diseases	3.3%	7	7	3.0%	Trachea, bronchus, lung cancers
Tuberculosis	2.7%	8	8	2.8%	Road traffic accidents
Trachea, bronchus, lung cancers	2.2%	9	9	2.4%	Tuberculosis
Road traffic accidents	2.1%	10	10	2.1%	Perinatal conditions
Diabetes mellitus	1.7%	11	11	1.8%	Stomach cancer
Malaria	1.6%	12	12	1.8%	Hypertensive heart disease
Hypertensive heart disease	1.6%	13	13	1.5%	Self-inflicted injuries
Self-inflicted injuries	1.5%	14	14	1.3%	Nephritis and nephrosis
Stomach cancer	1.5%	15	15	1.3%	Liver cancer

Deaths in children

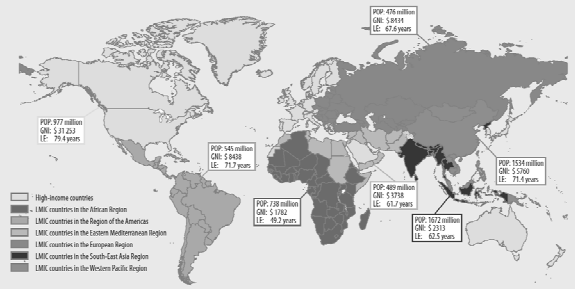
Figure 8: Major causes of death in children under 5 years old with disease-specific contribution of undernutrition, 2004.



Shaded area indicates contribution of undernutrition to each cause of death

Important regional differences

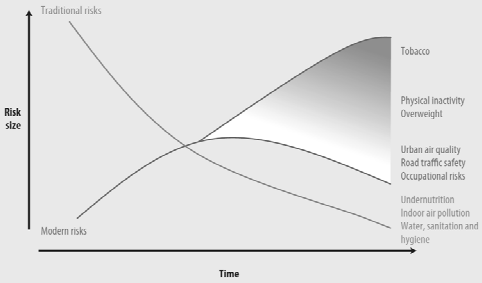
Figure 5: Low- and middle-income countries grouped by WHO region, 2004. Refer to Table A5 (Annex A) for a list of countries and definitions of categories.



POP, population; **GNI,** gross national income per capita (international dollars); **LE,** life expectancy at birth; **LMIC,** low- and middle-income countries.

The Epidemiologic Transition

Figure 2: The risk transition. Over time, major risks to health shift from traditional risks (e.g. inadequate nutrition or unsafe water and sanitation) to modern risks (e.g. overweight and obesity). Modern risks may take different trajectories in different countries, depending on the risk and the context.



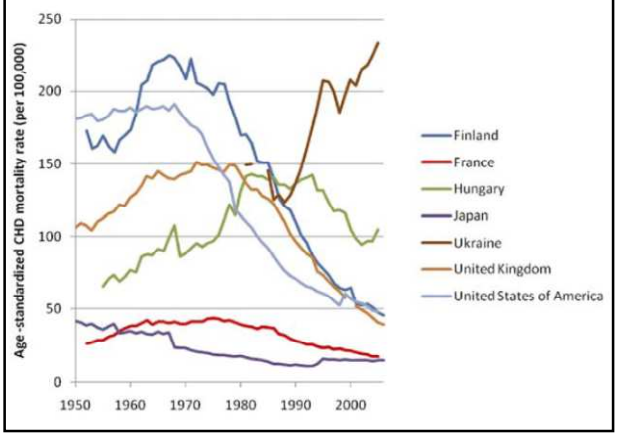
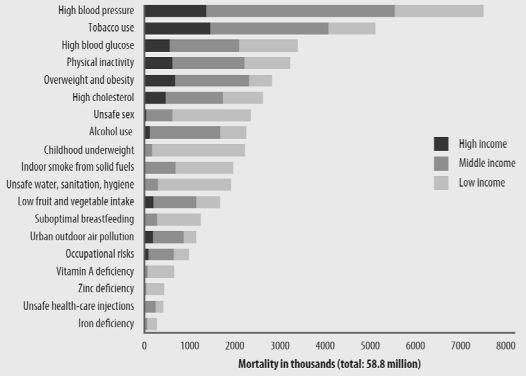


Table 1: Ranking of selected risk factors: 10 leading risk factor causes of death by income group, 2004

	Risk factor	Deaths (millions)	Percentage of total		Risk factor	Deaths (millions)	Percentage of total
	World				*Low-income countries**		
1	High blood pressure	7.5	12.8	1	Childhood underweight	2.0	7.8
2	Tobacco use	5.1	8.7	2	High blood pressure	2.0	7.5
3	High blood glucose	3.4	5.8	3	Unsafe sex	1.7	6.6
4	Physical inactivity	3.2	5.5	4	Unsafe water, sanitation, hygiene	1.6	6.1
5	Overweight and obesity	2.8	4.8	5	High blood glucose	1.3	4.9
6	High cholesterol	2.6	4.5	6	Indoor smoke from solid fuels	1.3	4.8
7	Unsafe sex	2.4	4.0	7	Tobacco use	1.0	3.9
8	Alcohol use	2.3	3.8	8	Physical inactivity	1.0	3.8
9	Childhood underweight	2.2	3.8	9	Suboptimal breastfeeding	1.0	3.7
10	Indoor smoke from solid fuels	2.0	3.3	10	High cholesterol	0.9	3.4
	*Middle-income countries**				*High-income countries**		
1	High blood pressure	4.2	17.2	1	Tobacco use	1.5	17.9
2	Tobacco use	2.6	10.8	2	High blood pressure	1.4	16.8
3	Overweight and obesity	1.6	6.7	3	Overweight and obesity	0.7	8.4
4	Physical inactivity	1.6	6.6	4	Physical inactivity	0.6	7.7
5	Alcohol use	1.6	6.4	5	High blood glucose	0.6	7.0
6	High blood glucose	1.5	6.3	6	High cholesterol	0.5	5.8
7	High cholesterol	1.3	5.2	7	Low fruit and vegetable intake	0.2	2.5
8	Low fruit and vegetable intake	0.9	3.9	8	Urban outdoor air pollution	0.2	2.5
9	Indoor smoke from solid fuels	0.7	2.8	9	Alcohol use	0.1	1.6
10	Urban outdoor air pollution	0.7	2.8	10	Occupational risks	0.1	1.1

* Countries grouped by gross national income per capita – low income (US$ 825 or less), high income (US$ 10 066 or more).

Figure 6: Deaths attributed to 19 leading risk factors, by country income level, 2004.



Worldwide, overweight and obesity cause more deaths than underweight.
The combined burden of these diet-related risks and physical inactivity in low- and middle-income countries is similar to that caused by HIV/AIDS and tuberculosis.

DALYs – considering morbidity

Box 1: Disability-adjusted life years (DALYs)

DALYs are a common currency by which deaths at different ages and disability may be measured. One DALY can be thought of as one lost year of "healthy" life, and the burden of disease can be thought of as a measurement of the gap between current health status and an ideal situation where everyone lives into old age, free of disease and disability.

DALYs for a disease or injury are calculated as the sum of the years of life lost due to premature mortality (YLL) in the population and the years lost due to disability (YLD) for incident cases of the disease or injury. YLL are calculated from the number of deaths at each age multiplied by a global standard life expectancy of the age at which death occurs. YLD for a particular cause in a particular time period are estimated as follows:

YLD = number of incident cases in that period × average duration of the disease × disability weight

The disability weight reflects the severity of the disease on a scale from 0 (perfect health) to 1 (death). The disability weights used for global burden of disease DALY estimates are listed elsewhere *(6)*.

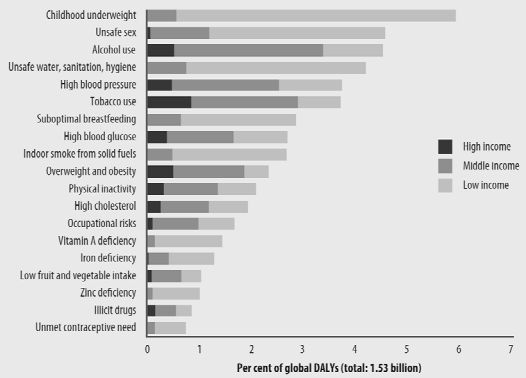
In the standard DALYs in recent WHO reports, calculations of YLD used an additional 3% time discounting and non-uniform age weights that give less weight to years lived at young and older ages *(7)*. Using discounting and age weights, a death in infancy corresponds to 33 DALYs, and deaths at ages 5–20 years to around 36 DALYs.

Table 2: Ranking of selected risk factors: 10 leading risk factor causes of DALYs by income group, 2004

	Risk factor	DALYs (millions)	Percentage of total		Risk factor	DALYs (millions)	Percentage of total
	World				*Low-income countries*[a]		
1	Childhood underweight	91	5.9	1	Childhood underweight	82	9.9
2	Unsafe sex	70	4.6	2	Unsafe water, sanitation, hygiene	53	6.3
3	Alcohol use	69	4.5	3	Unsafe sex	52	6.2
4	Unsafe water, sanitation, hygiene	64	4.2	4	Suboptimal breastfeeding	34	4.1
5	High blood pressure	57	3.7	5	Indoor smoke from solid fuels	33	4.0
6	Tobacco use	57	3.7	6	Vitamin A deficiency	20	2.4
7	Suboptimal breastfeeding	44	2.9	7	High blood pressure	18	2.2
8	High blood glucose	41	2.7	8	Alcohol use	18	2.1
9	Indoor smoke from solid fuels	41	2.7	9	High blood glucose	16	1.9
10	Overweight and obesity	36	2.3	10	Zinc deficiency	14	1.7
	Middle-income countries[a]				*High-income countries*[a]		
1	Alcohol use	44	7.6	1	Tobacco use	13	10.7
2	High blood pressure	31	5.4	2	Alcohol use	8	6.7
3	Tobacco use	31	5.4	3	Overweight and obesity	8	6.5
4	Overweight and obesity	21	3.6	4	High blood pressure	7	6.1
5	High blood glucose	20	3.4	5	High blood glucose	6	4.9
6	Unsafe sex	17	3.0	6	Physical inactivity	5	4.1
7	Physical inactivity	16	2.7	7	High cholesterol	4	3.4
8	High cholesterol	14	2.5	8	Illicit drugs	3	2.1
9	Occupational risks	14	2.3	9	Occupational risks	2	1.5
10	Unsafe water, sanitation, hygiene	11	2.0	10	Low fruit and vegetable intake	2	1.3

[a] Countries grouped by 2004 gross national income per capita – low income (US$ 825 or less), high income (US$ 10 066 or more).

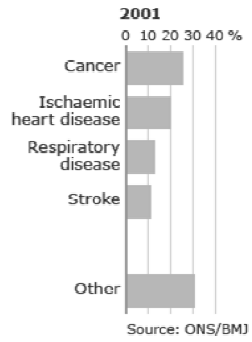
Figure 7: Percentage of disability-adjusted life years (DALYs) attributed to 19 leading risk factors, by country income level, 2004.



Some key Global take homes …

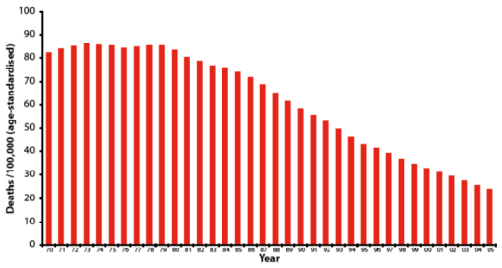
- Worldwide, Africa accounts for 9 out of every 10 child deaths due to malaria, for 9 out of every 10 child deaths due to AIDS, and for half of the world's child deaths due to diarrhoeal disease and pneumonia.
- In low-income countries, the leading cause of death is pneumonia, followed by heart disease, diarrhoea, HIV/AIDS and stroke. In developed or high-income countries, the list is topped by heart disease, followed by stroke, lung cancer, pneumonia and asthma or bronchitis.
- Men between the ages of 15 and 60 years have much higher risks of dying than women in the same age category in every region of the world. This is mainly because of injuries, including violence and conflict, and higher levels of heart disease. The difference is most pronounced in Latin America, the Caribbean, the Middle East and Eastern Europe.
- Depression is the leading cause of years lost due to disability, the burden being 50% higher for females than males. In all income strata, alcohol dependence and problem use is among the 10 leading causes of disability.

UK causes of mortality

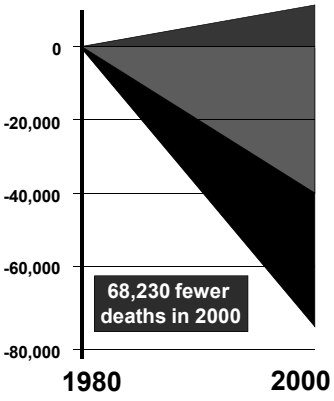


Changing face of CHD

UK CHD mortality rates – BHF 2008



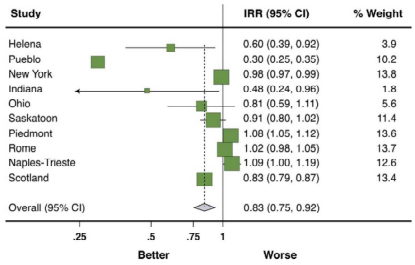
Deaths averted



Risk factors worse	**+13%**
Obesity	+3.5%
Diabetes	+4.8%
Less physical activity	+4.4%
Risk factors better	**-71%**
Smoking	-41%
Cholesterol	-9%
Popul'n BP fall	-9%
Deprivation	-3%
Other factors	-8%
Treatments	**-42%**
AMI treatments	-8%
Secondary prevention	-11%
Heart failure	-12%
Angina: CABG/PCI	-4%
Angina: drugs	-5%
BP treatment	-3%

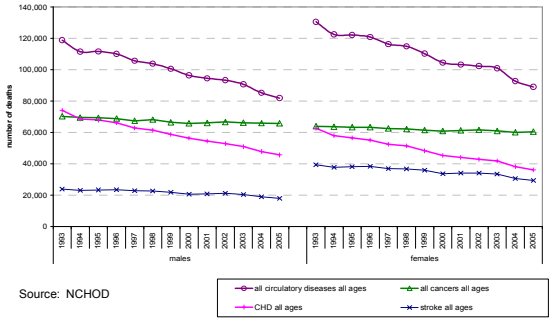
Redrawn from Capewell and colleagues

Effect of public smoking ban on CHD mortality rates



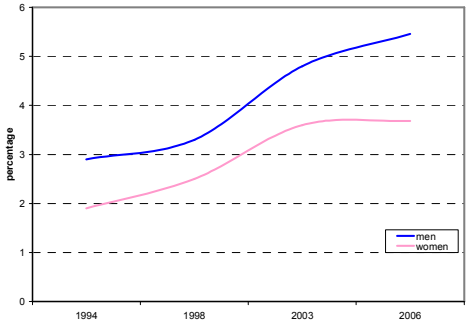
JACC 2009

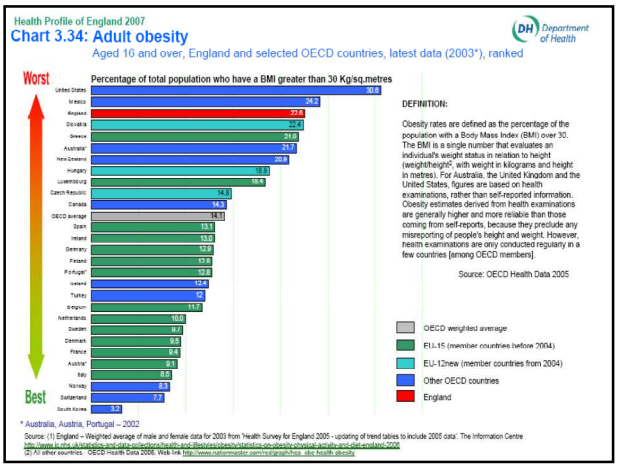
Changes in big UK killers (all ages)

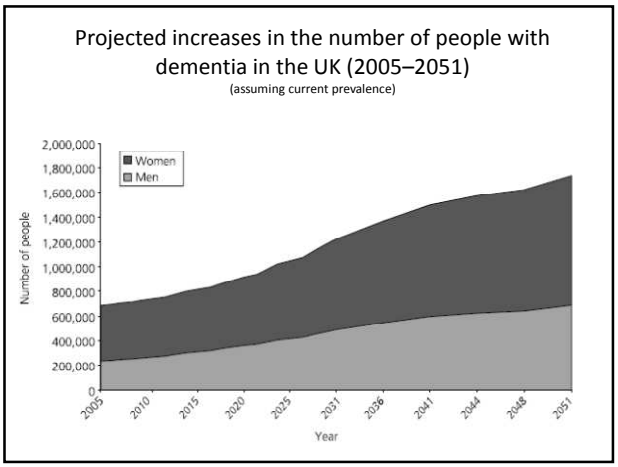


Source: NCHOD

Prevalence of doctor-diagnosed diabetes, 16+
England 1994-2006







WHO: the six major causes of global morbidity / mortality

- Hypertension
- Overweight
- Smoking
- Alcohol consumption
- Underweight
- Sexual behaviour

(Lopez et al., 2006).

BioStatistics

Dr John Chambers

Q1: Data distribution - which of the following statements is correct?

A. The data are likely to be normally distributed
B. The data are negatively skewed
C. At least 50% of patients are admitted for over 3 days
D. Approximately 25% of patients stay over 19 days
E. 25% of patients stay between 2 and 9 days

Q2: Choosing the right test - what is the most appropriate test to determine whether the two variables are related?

A. Pearsons correlation co-efficient
B. Chi-squared test
C. Analysis of variance
D. Students t test
E. Rank sign test

Q3: Choosing the right test - what is the most appropriate statistical test for assessing whether attendances vary significantly by day of week?

A. Correlation co-efficient
B. ANOVA
C. Students t test
D. Cox proportional hazard
E. Chi-squared test

Q4: You read the results of a paper describing the efficacy of treatment A in condition B. Which of the following features indicate that the study is not well designed?

A. Double blind design
B. Prior publication of a methodology paper
C. Placebo controlled
D. Intention to treat analysis
E. Systematic bias between treatment groups

Q5: In a newly developed test, the following statement is incorrect?

A. Sensitivity = The proportion of positives that are correctly identified by the test
B. Specificity = The proportion of negatives that are correctly identified by the test
C. Positive predictive value = The proportion of patients with a positive test result who are correctly diagnosed
D. Negative predictive value = The proportion of patients with a negative test result who are correctly diagnosed
E. Predictive value is not influenced by disease prevalence in the population

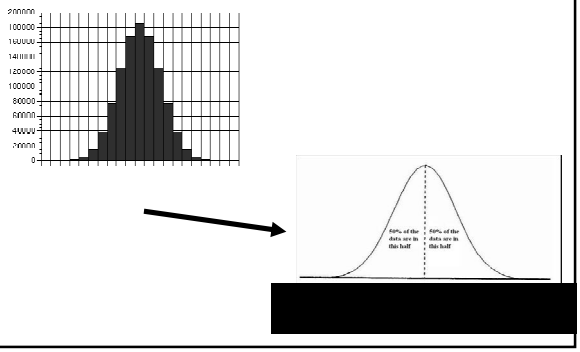
Q6: ROC graphs - which of the following statements is incorrect?

A. ROC graphs plot sensitivity against specificity
B. The optimal cutoff is typically identified from the maximum of the sum of specificity and sensitivity
C. High sensitivity indicates a low false negative rate
D. Sensitivity equals (1 – specificity)
E. High specificity indicates a low false positive rate

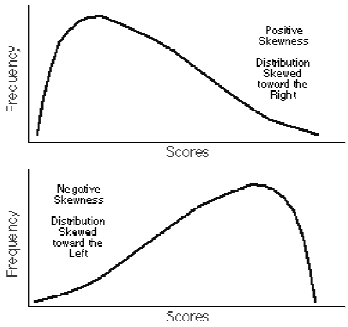
Data types

- What type of data?
  - Categorical
  - Continuous

Distribution of values in a population

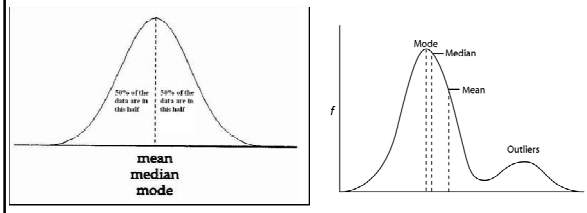


Skewed distributions

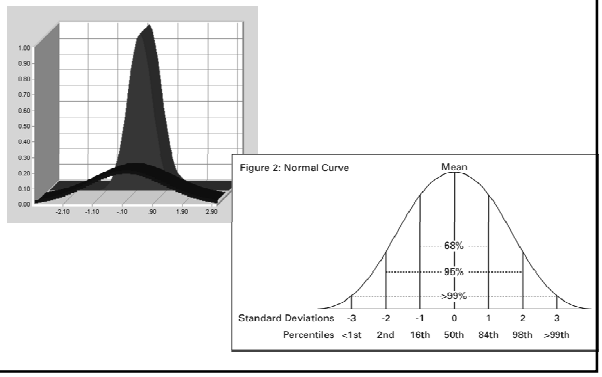


What is the central value?

Mean – average
Median – mid value
Mode – most common value



Standard deviation – a measure of the spread



Total cholesterol – where is the mean?

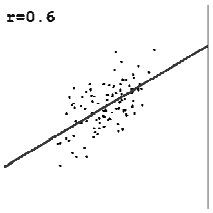
Standard Error of the mean

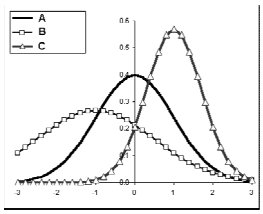
- The spread over which true population mean might be
- SEM = SD / (√ [number of observations])
- More subjects – lower SEM – more accurate

What test?

Continuous data

- Relationship: Correlation coefficient
- Difference 2 groups: Student's t test
- Difference N groups: ANOVA





Categorical data

- Distribution differs between groups: chi-squared

Hair Colour	Eye Colour: Blue	Green	Brown	Black	Total
Blonde	2	1	2	1	6
Red	1	1	2	0	4
Brown	1	0	4	2	7
Black	1	0	2	0	3
Total	5	2	10	3	20

Odds Ratios and Relative Risks

- Relative Risk = risk in prospective studies
- Odds Ratio = risk in cross-sectional studies
- OR = RR for uncommon disorders

Relative risk

Relative Risk

= Risk in Exposed / Risk in Non-exposed

	Disease	No disease	Total	Risk
Exposed	A	B	A+B	A/(A+B)
Not Exposed	C	D	C+D	C/(C+D)

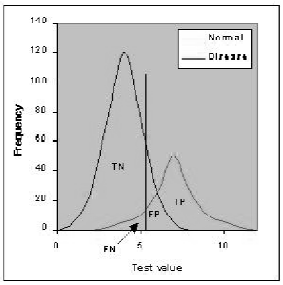
Odds Ratio

$$\text{Odds ratio} = \frac{PG_1 / (1 - PG_1)}{PG_2 / (1 - PG_2)}$$

PG = proportion exposed (1=cases, 2=control)

	Case	Control
Exposed	A	B
Not Exposed	C	D
PG	A/(A+C)	B/(B+D)
1-PG	C/(A+C)	D/(B+D)
PG/(1-PG)	[A/(A+C)]/[C/(A+C) = A/C	[B/(B+D)]/[D/(B+D) = B/D

Evaluation of a test



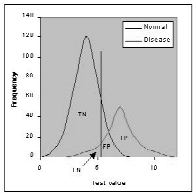
	Disease –ve	Disease +ve
Test – ve	A	B
Test +ve	C	D

- True negative = A
- False negative = B
- False positive = C
- True positive = D

Evaluation of a test 2

- True negative rate
- False negative rate
- False positive rate
- True positive rate
- Sensitivity
- Specificity
- Positive predictive value
- Negative predictive value

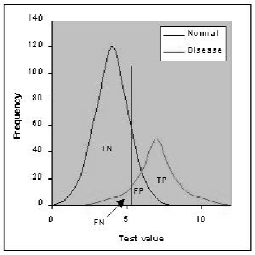
Evaluation of a test 3



	Disease –ve	Disease +ve
Test – ve	A	B
Test +ve	C	D

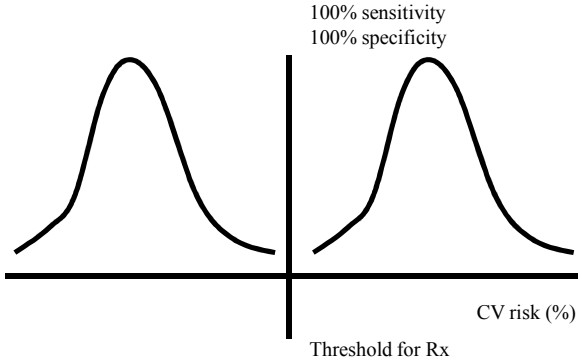
- **Sensitivity**: % of disease pts correctly identified = D / (B+D)
- **Specificity**: % of healthy pts correctly identified = A / (A+C)
- **PPV**: % of +ve test results with disease present = D / (C+D)
- **NPV**: % of -ve test results with no disease present = A / (A+B)

Sensitivity / Specificity change as cut-off changes

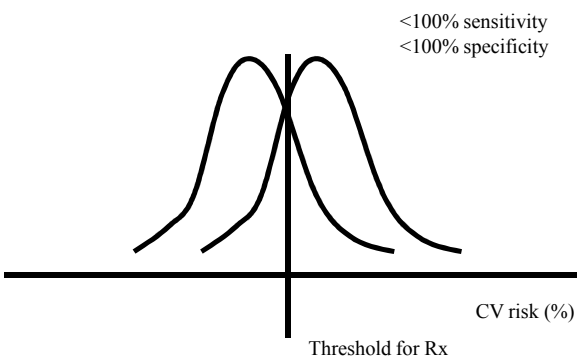


As you reduce threshold, SENSITIVITY INCREASES, but SPECIFICITY DECREASES

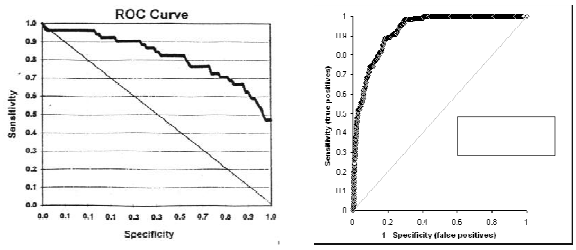
The perfect diagnostic test



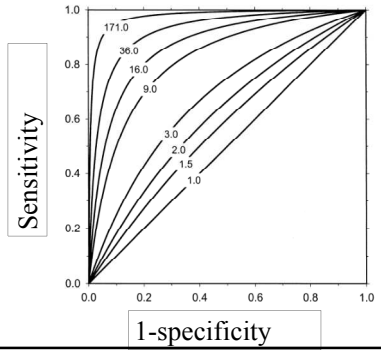
Diagnostic tests



Receiver Operator Characteristic Curves



Disease prediction



Absolute and Relative Risk

- Risk reduction: decrease in risk with a treatment
  - Relative risk reduction: as a percent of risk in control group
  - absolute risk reduction: as an absolute number
- Number needed to treat (NNT): Number of patients you need to treat to prevent one event
- Intention to treat analysis: Analysing results of an RCT based upon the initial randomisation of subjects, irrespective of whether they later dropped out / crossed over

Other terms

- Incidence: number of new cases in a population
- Prevalence: number of existing cases in a population
- Blinding
  - Single blind: Patient blinded
  - Double blind: Researcher and Patient blinded
- Bias: systematic error that leads to results that do not represent the true nature of things
- Variability: variation in test results on same sample / pt
  - Intra-observer
  - Inter-observer

Other terms 2

- Type I error: failure to find a difference where one truly exists
  - Eg small studies
- Type 2 error: finding a difference where none exists
  - Eg by chance (1/20)
  - Eg bias

Study power

- Power: ability of a study to detect a difference between groups
- More subjects / events = greater power
- Greater power =
  - Detect smaller difference
  - Less error
  - More accuracy
  - More significant results