

Introduction to epidemiology



KING'S HEALTH PARTNERS

Pioneering better health for all

Zagreb March 2018, IMforFuture

Frances MK Williams PhD FRCP(E)

Professor of Genomic Epidemiology

Consultant Rheumatologist, Guy's and St Thomas' Hospital

KING'S
College
LONDON

BACKGROUND TO ME

- 1991 BSc in Clinical Science University of London
- 1992 MBBS
- 2002 PhD
- 2004 CCST in Rheumatology (South Thames)
- 2007 Wellcome Trust Fellow in Genetic Epidemiology/Consultant Rheumatologist
- 2017 Professor of Genomic Epidemiology King's College London



IM FOR FUTURE



epi = upon, among
demos = people, district
logos = study

AIMS

- To present and give basic training of **transferable research skills** to ITN students, including **critical analysis**
- To give an overview and basic training of **statistical processes**; enable students to incorporate statistics into their **individual research projects**

TOPICS COVERED

- background to epidemiology
- observation
- study designs
 - case control
 - cohort
 - genetically informative designs
- epidemiological terms
- correlation regression
- sensitivity and specificity, PPPV, NPPV

WHAT IS MEANT BY RESEARCH?

- Derives generalisable new knowledge
- Generates and tests hypotheses as well as studies that aim to test them
- Addresses clearly defined questions, aims and objectives
- Involves collecting data that are additional to those for routine care but may include data collected routinely
- May involve treatments, samples or investigations additional to routine care.

ALL SWANS ARE WHITE



TESTING A RESEARCH IDEA

- A null hypothesis
- A statistical test
- A decision threshold

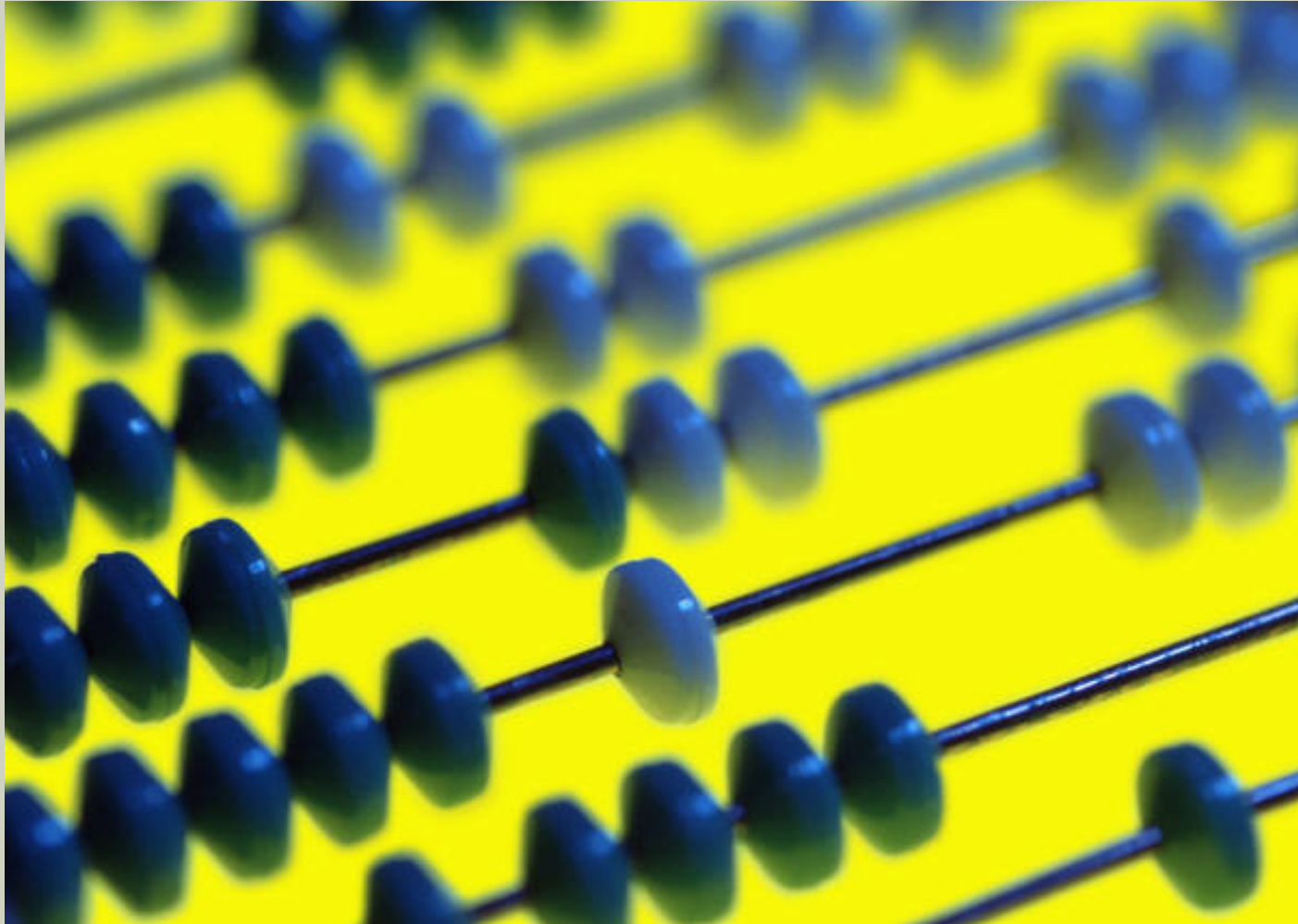
EPIDEMIOLOGY

- The study of patterns of disease in populations
- Distinct from interventional studies, clinical trials

OBSERVATION

- All epidemiology enquiry begins with quantifying the occurrence of illness

COUNTING



OBSERVATIONAL

- Prevalence
 - Existing cases in the population
 - Proportion (%)
- Incidence
 - New cases over time
 - Rate (time^{-1})

OBSERVATIONAL (2)

- Prevalence
 - Point
 - Period
- Incidence
 - New cases occurring/time
 - Lifetime incidence
 - Attack rate
 - Case fatality rate

OBSERVATIONAL (3)

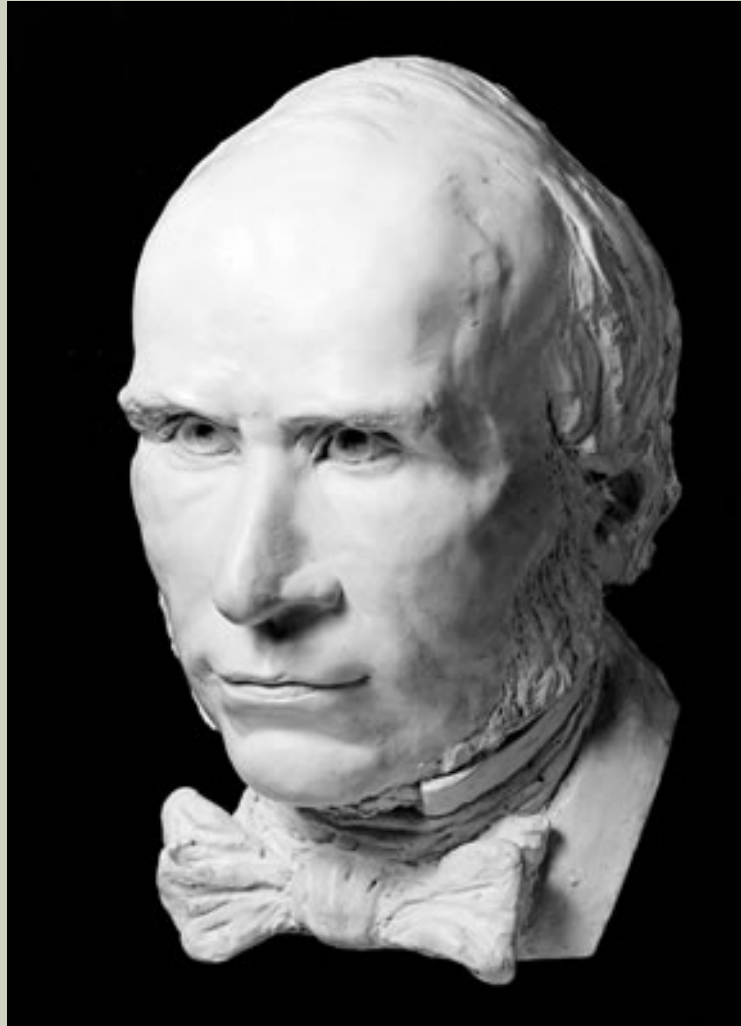
- Prevalence

*The prevalence of rheumatoid arthritis
in the UK is 1%*

- Incidence

*The incidence of rheumatoid arthritis
in the UK is 1/1000/year*

JOHN SNOW (1813-1858)



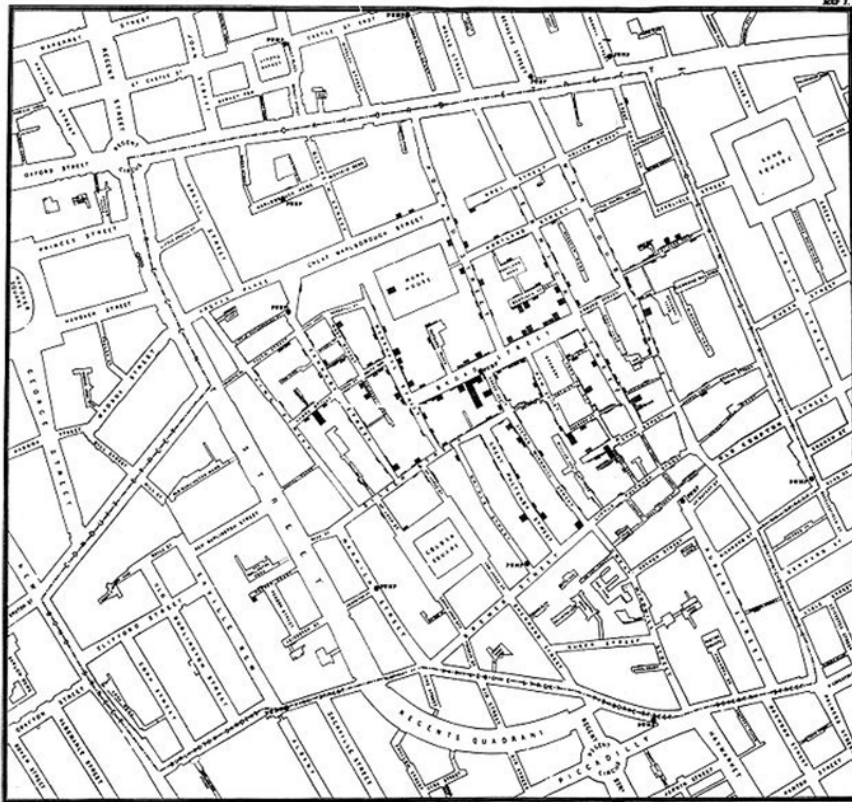
LONDON CHOLERA OUTBREAK 1854

Group of districts with water supplied by	Water supply of individual houses	Population 1851	Deaths from cholera	Cholera death-rate per 1000 population
Southwark and Vauxhall Company	Southwark and Vauxhall Company	167,654**	738	4.4
Lambeth Company	Lambeth Company	19,133**	4	0.2
Both companies	Southwark and Vauxhall Company	98,862	419	4.2
	Lambeth Company	154,615	80	0.5
Rest of London		1,921,972	1,422	0.7

* From Snow, 1855

** Overestimated by a small amount, since this figure includes population with no water supply

FROM OBSERVATION TO ACTION

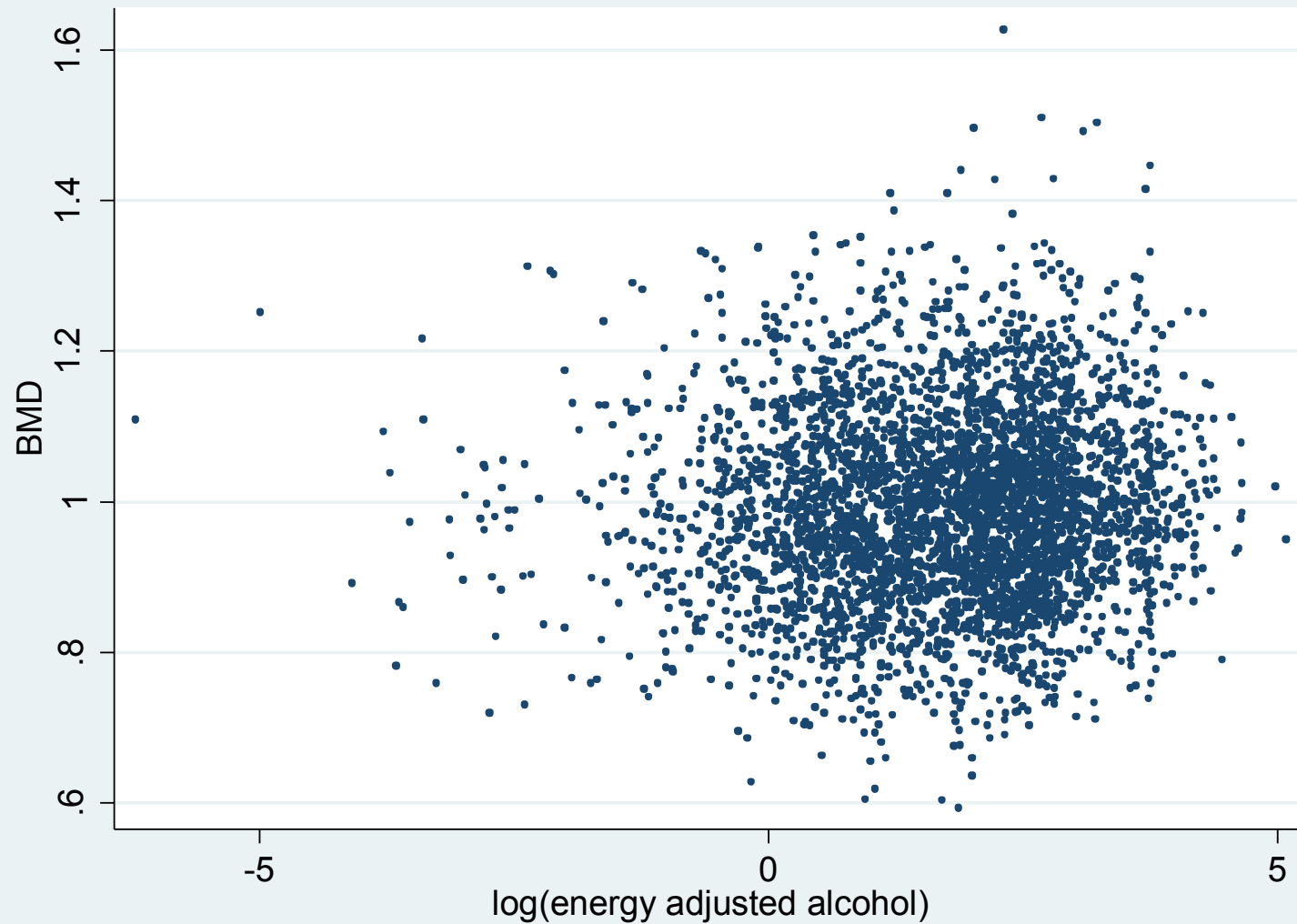


- Removing the handle of the Broad Street pump

WE MUST DEDUCE

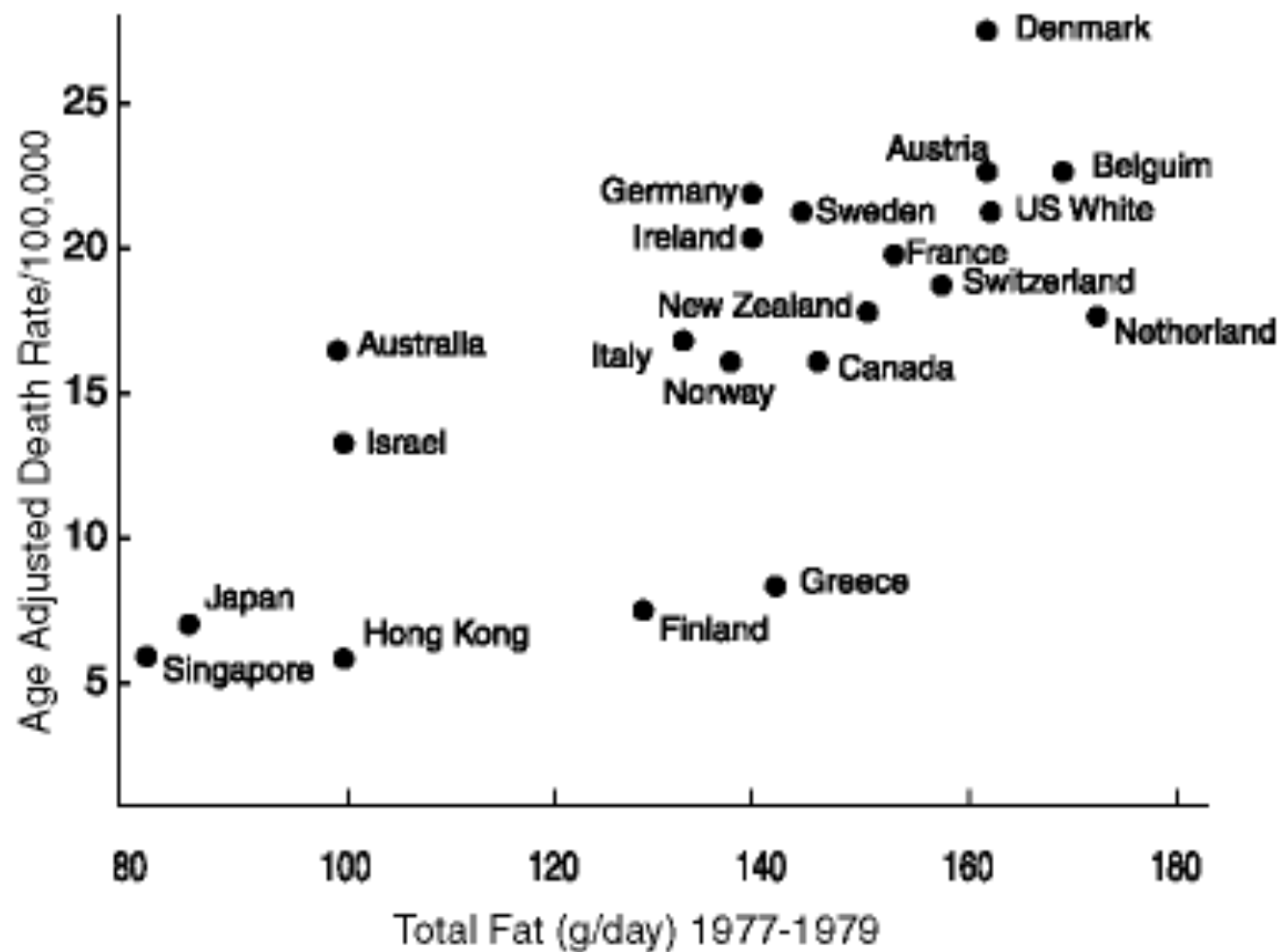


WHY SHE WAS CROSSING THAT ROAD



Correlation

BMD is associated with alcohol intake in 5,000 individuals



OBSERVATIONS

- Correlation

BMD is associated with alcohol intake in 5,000 individuals

- Ecologic study

Average fibre intake is negatively correlated with annual incidence of bowel cancer in different countries

**NOTE: CORRELATION DOES
NOT EQUAL CAUSATION**



OBSERVATIONS IN MEDICINE

- Case reports

Diffuse small bowel thickening in AIDS patient - a case report.

- Case series

Silicone breast implantation-induced scleroderma: description of four patients and a critical review of the literature



Hypothesis generating

CASE CONTROL STUDY

BRITISH MEDICAL JOURNAL

LONDON SATURDAY SEPTEMBER 30 1950

SMOKING AND CARCINOMA OF THE LUNG

PRELIMINARY REPORT

BY

RICHARD DOLL, M.D., M.R.C.P.

Member of the Statistical Research Unit of the Medical Research Council

AND

A. BRADFORD HILL, Ph.D., D.Sc.

Professor of Medical Statistics, London School of Hygiene and Tropical Medicine; Honorary Director of the Statistical Research Unit of the Medical Research Council

In England and Wales the phenomenal increase in the number of deaths attributed to cancer of the lung provides one of the most striking changes in the pattern of mortality recorded by the Registrar-General. For example, in the quarter of a century between 1922 and 1947 the annual number of deaths recorded increased from 612 to 9,287, or roughly fifteenfold. This remarkable increase is, of course, out of all proportion to the increase of population—both in total and, particularly, in its older age groups.

whole explanation, although no one would deny that it may well have been contributory. As a corollary, it is right and proper to seek for other causes.

Possible Causes of the Increase

Two main causes have from time to time been put forward: (1) a general atmospheric pollution from the exhaust fumes of cars, from the surface dust of tarred roads, and from gas-works, industrial plants, and coal fires; and

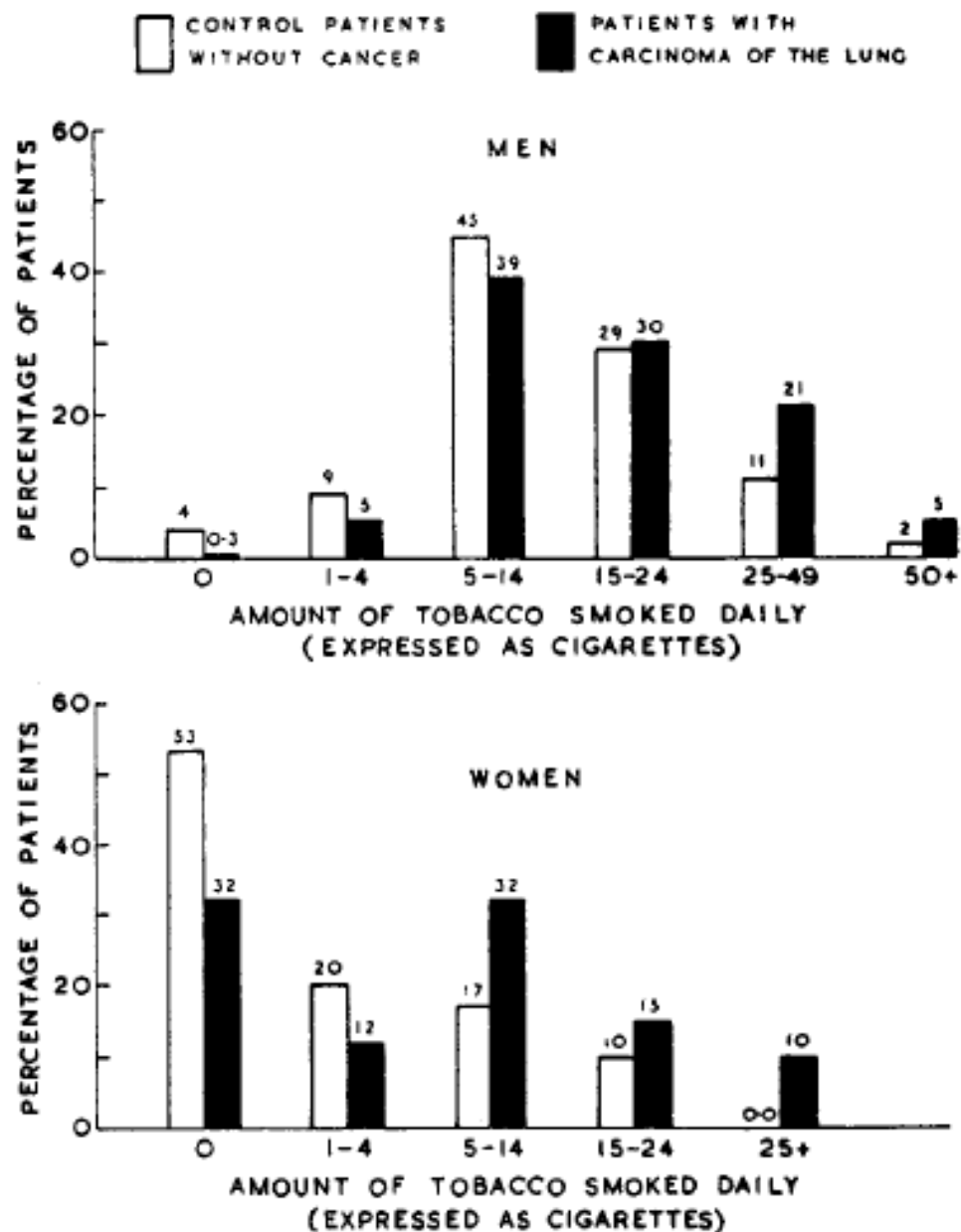


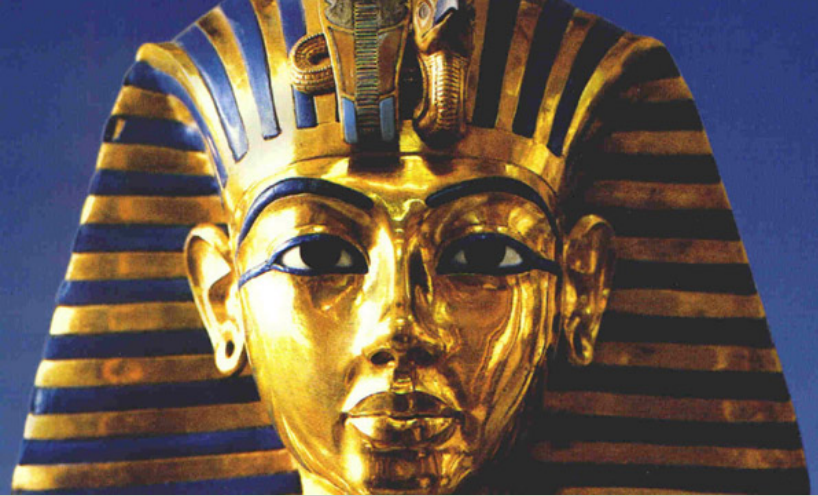
FIG. 1.—Percentage of patients smoking different amounts of tobacco daily.

CASE CONTROL STUDY

- Select cases (with disease)
- Select controls (with no disease)
- Compare exposures
- Higher level of exposure in the disease groups implies an association between exposure and disease

COHORT STUDY

- Select a cohort with no disease
- This will include a group with a certain exposure
- Observe over time (retrospective, prospective or mixed)
- Compare the incidence of disease in the exposed with the unexposed
- Higher incidence of disease in the exposed groups implies an association between exposure and disease



The mummy's curse: historical cohort study

Mark R Nelson

Department of
Epidemiology and
Preventive
Medicine, Monash
University, Alfred
Hospital, Prahran
3181, Australia
Mark R Nelson
NHMRC fellow

mark.nelson@med.
monash.edu.au

BMJ 2002;325:1482-4

"Can you see anything?" It was all I could do to get out the words, "Yes, wonderful things."

Howard Carter¹

Abstract

Objective To examine survival of individuals exposed to the "mummy's curse" reputedly associated with the opening of the tomb of Tutankhamen in Luxor, Egypt, between February 1923 and November 1926.

Design Retrospective cohort study.

Participants 44 Westerners identified by Howard Carter as present in Egypt at the specified dates. 25 of

Introduction

The death in 1923 of George Herbert (Lord Carnarvon), the financier of the expedition that unearthed the tomb of Tutankhamen, unleashed a sensation in the international newspapers. He had developed erysipelas at the site of a mosquito bite, which resulted in septicaemia and pneumonia. The speculation was that his death was due to a "mummy's curse." The press reports of the time had the death of every man and his dog being associated with the curse, no matter how obscure the connection. This was literally the case for Lord Carnarvon as his three legged canine was said to have bayed at the very time his mas-

Group comparison of characteristics of people with data on mortality according to exposure to mummy's curse. Figures are means (SD) unless stated otherwise

	Exposed (n=25)	Unexposed (n=11)	P value
No (%) of men	24 (96%)	7 (64%)	<0.001
Age at classification (years)	49.3 (11.0)	44.1 (9.1)	0.25
Age at death (years)	70.0 (12.4)	75.0 (13.0)	0.87
Survival (years)	20.8 (15.2)	28.9 (13.6)	0.95

FRAMINGHAM HEART STUDY

- 5,209 respondents of a random sample of 2/3 of the adult population of Framingham, Massachusetts, 30 to 62 years of age by household, in 1948.
- Exam 30 for the Original Cohort began in May of 2008 and ended in February of 2010

CASE CONTROL

	Lung cancer	Well	
Smoke	80	20	
Non smoker	20	80	
	100	100	

COHORT

	Lung cancer	Well	
Smoke	80	20	100
Non smoker	20	80	100

CASE CONTROL STUDIES

STRENGTHS

- Quick and inexpensive
- Suited to diseases with long latent periods
- Optimal for rare disease
- Multiple risk factors

WEAKNESSES

- Inefficient for rare exposures
- Cannot compute incidence
- Temporal relationship hard to establish
- Prone to bias

COHORT STUDIES

STRENGTHS

- Suited to rare exposures
- Multiple outcomes
- Temporal relationship
- Prospective, minimises bias
- Measures incidence

WEAKNESSES

- Inefficient for rare disease
- Expensive, time consuming, results may be too late
- Retrospective needs full available records
- Losses to follow up affect validity

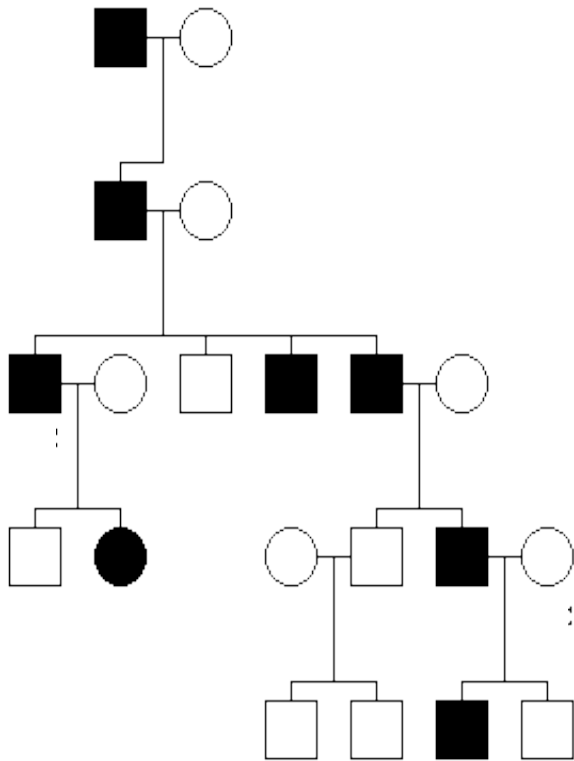
NESTED CASE CONTROL DESIGN

- Case control within a cohort
 - Subset of controls from the cohort are compared to the incident cases
- Useful when exposure is difficult or expensive to obtain
- Useful when outcome is rare
- Controls are in a matched risk set providing greater potential efficiency

GENETICALLY INFORMATIVE DESIGNS



CLUSTERING IN A FAMILY



- Colour blindness
- Being a lawyer

SEPARATING GENES AND ENVIRONMENT

- Migrant
- Admixture and inbreeding
- Familial clustering
- Classical twin study
- Twin / family adoption studies

MIGRANT STUDIES

- Are rates of disease similar to country of origin or host?

Example : Gastric cancer

- 1st generation Japanese migrants to the US show rates similar to Japan;
- 2nd generation migrant rates are similar to that in the US

INBREEDING AND ADMIXTURE STUDIES

- Inbreeding increases homozygosity
- Admixture increases heterozygosity

Example

US Blacks vs African Blacks with Type 1 DM :

- higher frequency of disease
- higher degree of white admixture

TWIN STUDIES

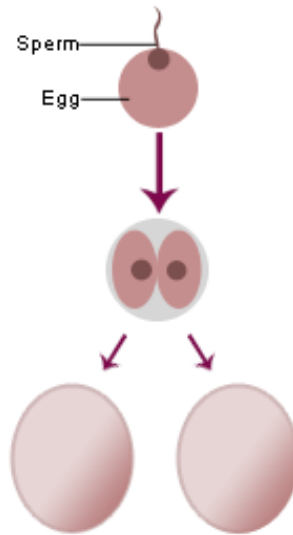
- Excess occurrence in MZ compared with DZ twins can be attributed to genetic factors
- MZ occurrence indicates maximum risk conferred by genetic factors
- assumes MZ and DZ pairs share the environment to the same extent



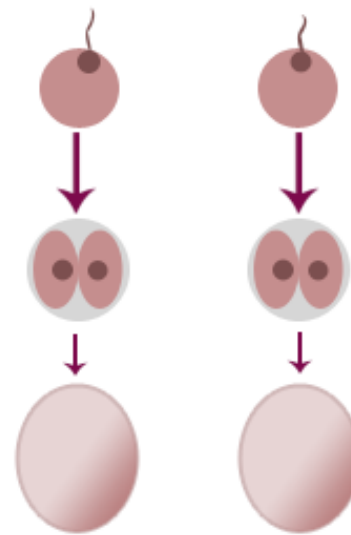


TWINS: in perspective

a) Identical (Monozygotic) Twins



b) Fraternal (Dizygotic) Twins

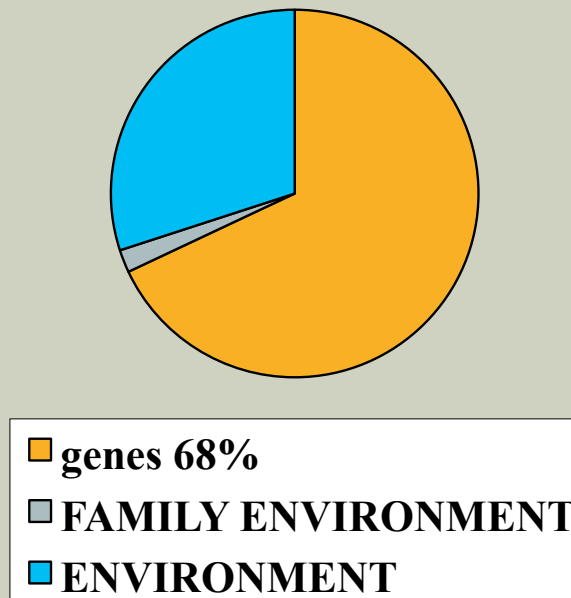


Mechanisms of twinning

TWINS: in perspective

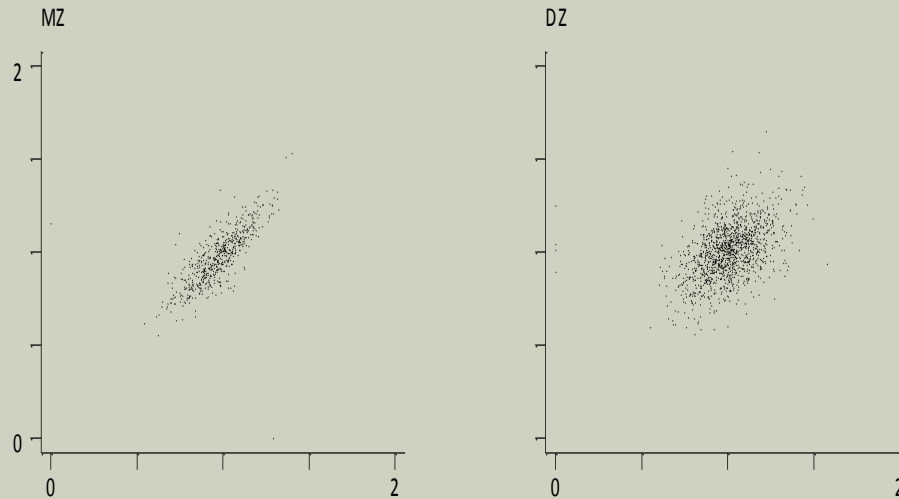
HERITABILITY ESTIMATES

“The proportion of variation between population samples explained by genetic differences.”



THE CLASSICAL TWIN DESIGN

- *Example: Bone mineral density at the spine*



$R_{MZ} = 0.726$

$R_{DZ} = 0.440$

GENETIC ASSOCIATION STUDY

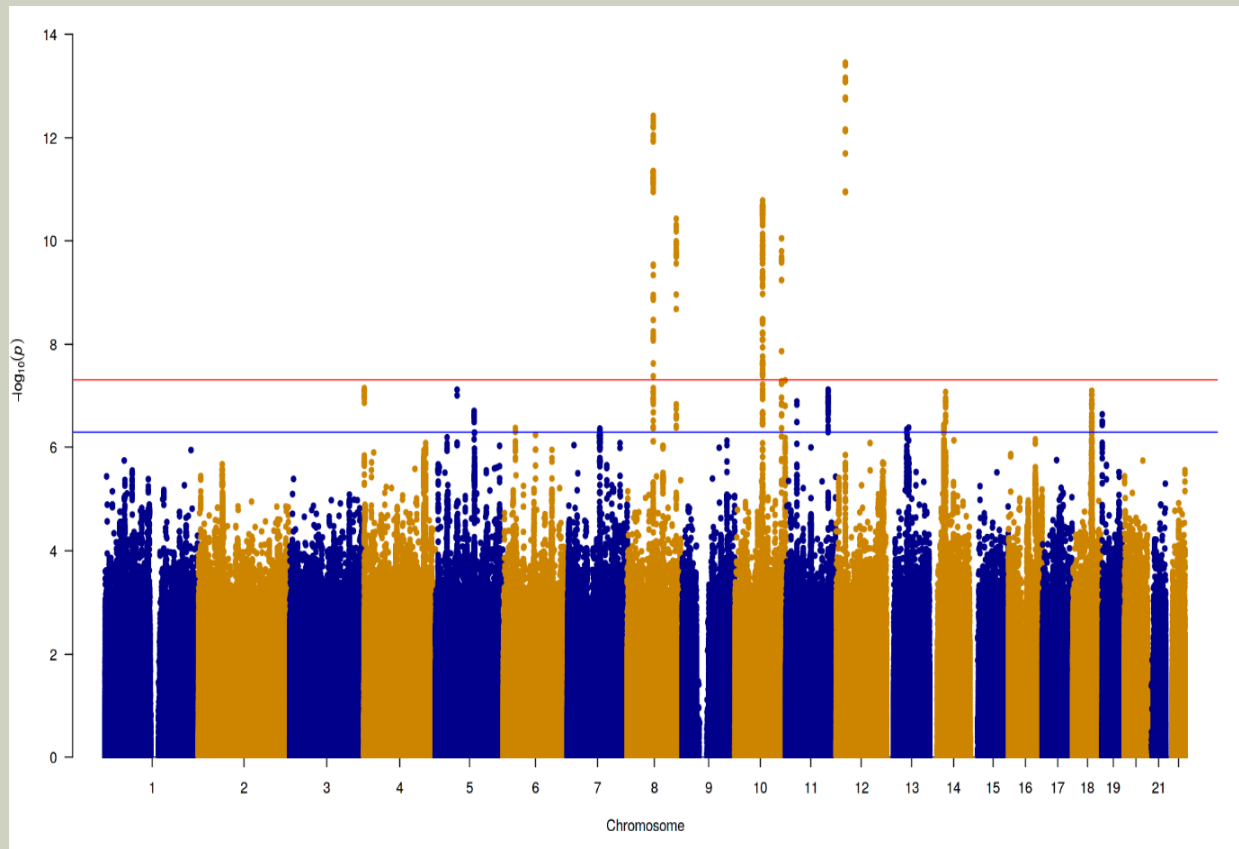
- An association study tests whether presence of a specific genetic variant contributes to disease susceptibility
- Compare the frequency of SNP genotypes (or alleles) in cases and controls
- Population association between allele and disease

STUDY DESIGN

- Cases
 - phenotype definition, clinical criteria
 - ascertainment
 - geographic + ethnic background
- Controls
 - random population samples or screened unaffected individuals
 - geographic + ethnic background

GENOME-WIDE ASSOCIATION (GWAS)

- SNPs distributed throughout the genome
- 100,000 – 5,000,000
- Hypothesis-free
- Corrections for multiple comparisons



BACK PAIN GWAS – UK BIOBANK

SOME TERMS IN EPIDEMIOLOGY

- Proportion
- Probability
- Risk
- Odds
- Rate

PROPORTION

- Of 120 patients attend who attend a rheumatology outpatients, 30 have fibromyalgia.
- The proportion with fibromyalgia is $30/120=0.25$ or 25%
- The probability of having fibromyalgia is 0.25
- The risk of having fibromyalgia is 0.25

ODDS

- Of 120 patients attend who attend a rheumatology outpatients, 30 have fibromyalgia.
- The odds of having fibromyalgia is 30:90 or one to three
- NOTE: numerator not contained in denominator

RATE

- ***time*** in the denominator

OBSERVATIONAL

- Prevalence is a proportion

*The prevalence of rheumatoid arthritis
in the UK is 1%*

- Incidence is a rate – having time denominator

*The incidence of rheumatoid arthritis
in the UK is 1/1000/year*



Would Diana have seen Kate as an ally or rival?



PAGES 22-23

DON'T DRINK MORE THAN 3 GLASSES OF WINE A WEEK

Oxford study claims slashing the official alcohol limit would save 4,500 lives a year

BRITONS should drink no more than three small glasses of wine a week, scientists have said.

More than 4,500 lives could be saved annually by changing official advice on 'safe' levels of alcohol intake, they believe.

It would mean recommended consumption - for men and women - would be cut to half a unit a day, the equivalent of just a couple of gulps of beer. The new advice flies in the face

By **Jenny Hope**
Medical Correspondent

of previous studies, which have shown that drinking alcohol in moderation reduces the risk of dying from heart disease.

But the researchers, from Oxford University, say this benefit is far outweighed by the harm to health caused by regular drinking.

Cutting consumption could stem the epidemic of alcohol-related chronic diseases set to cause 210,000 deaths during the next 30 years. Currently, the

Department of Health says 'safe' drinking levels are three to four units a day for men, or two to three for women.

A small glass of wine contains 1.3 units, while a pint of beer contains at least two units.

But the new study says the ideal intake to prevent chronic disease is five grams a day - around half a unit. This is less than half a small glass of wine and just a quarter of a pint of beer.

The findings could influence the first review of drinking advice for 15 years.

Turn to Page 2



My Twitter ordeal, by Rebecca Adlington

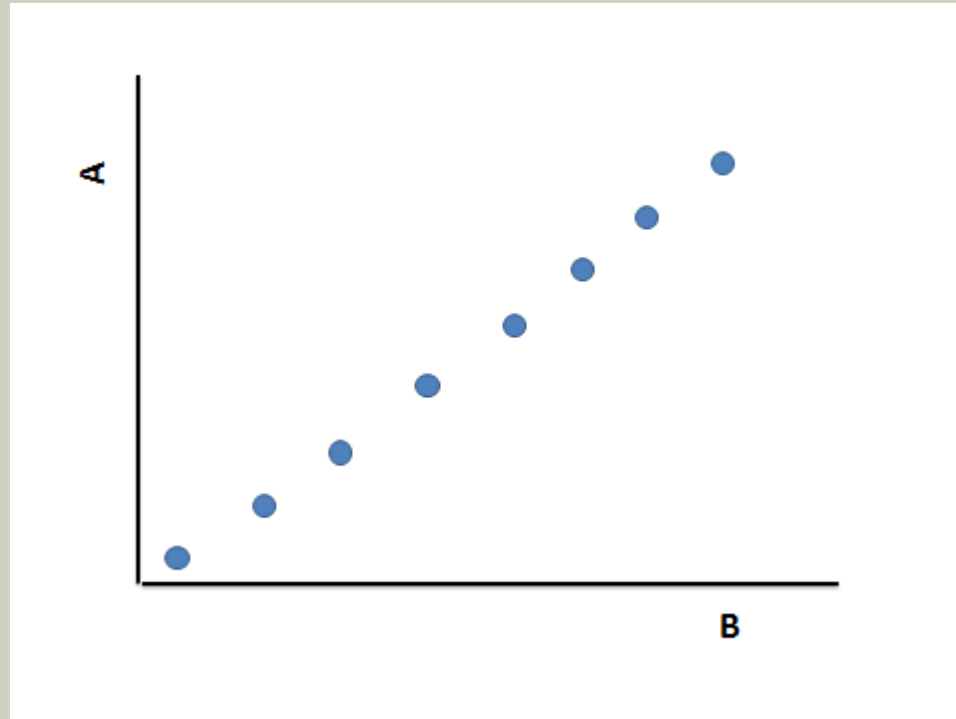
PAGES 24-25

Rebecca: Taunted over her looks

CORRELATION

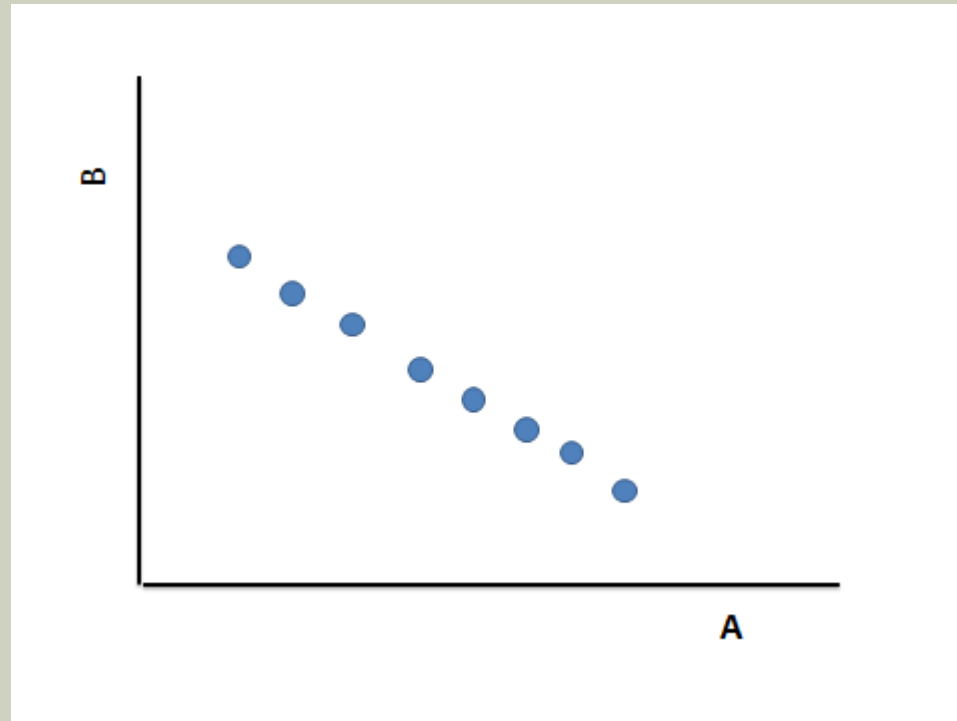
- Continuous data
- Dimensionless
- Range -1 to 1

CORRELATION

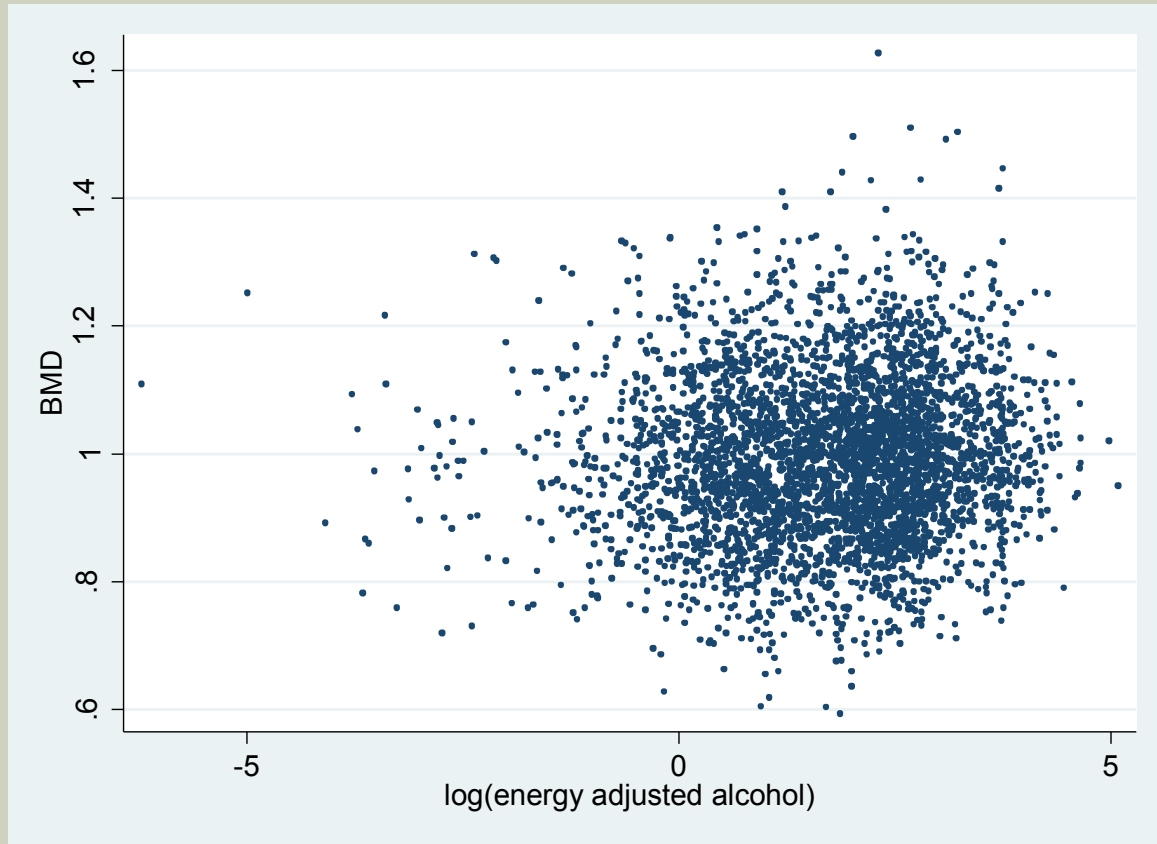


- Perfect linear correlation
- $R=+1$

CORRELATION

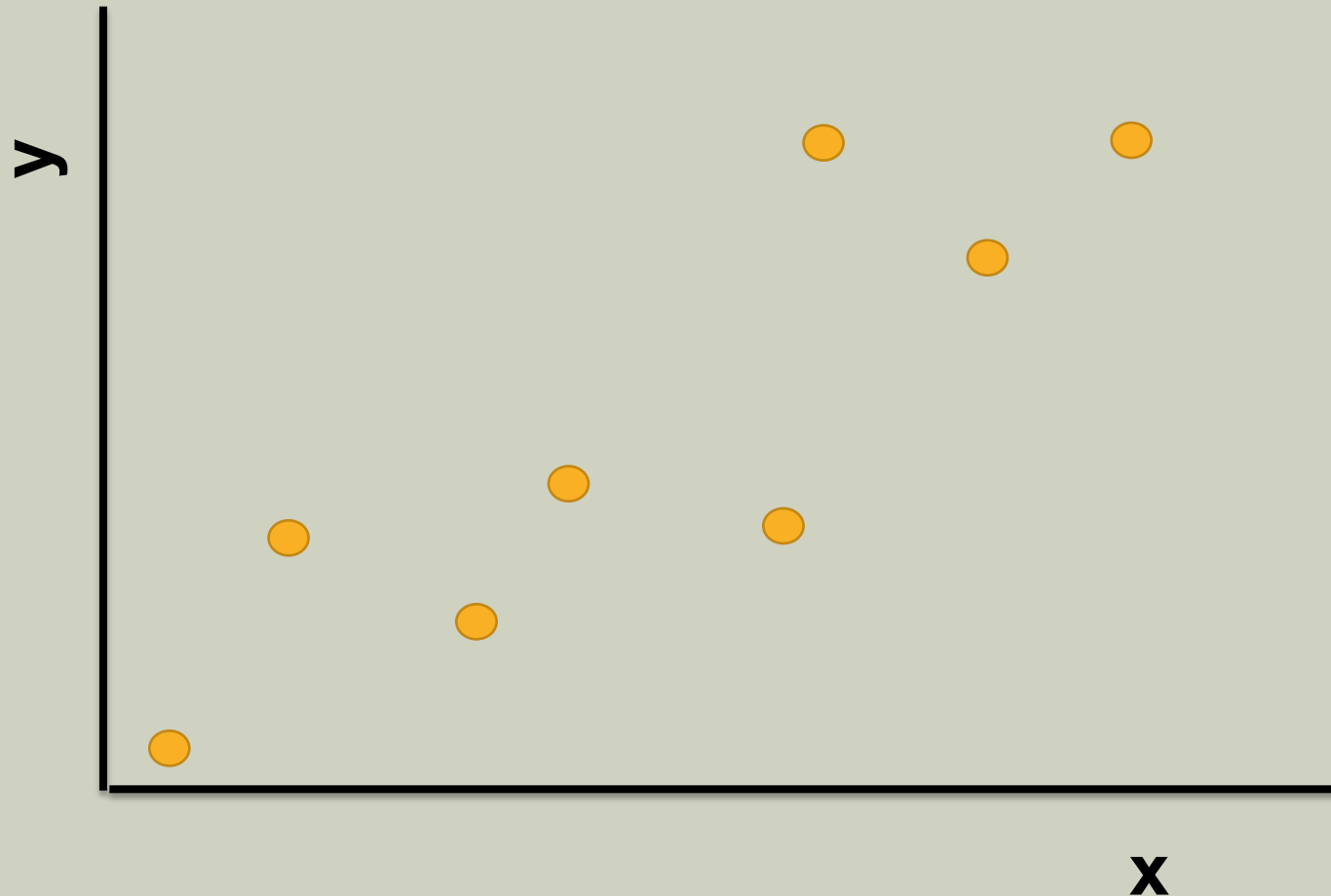


- Perfect linear correlation
- $R = -1$



- Square of correlation is explained variance

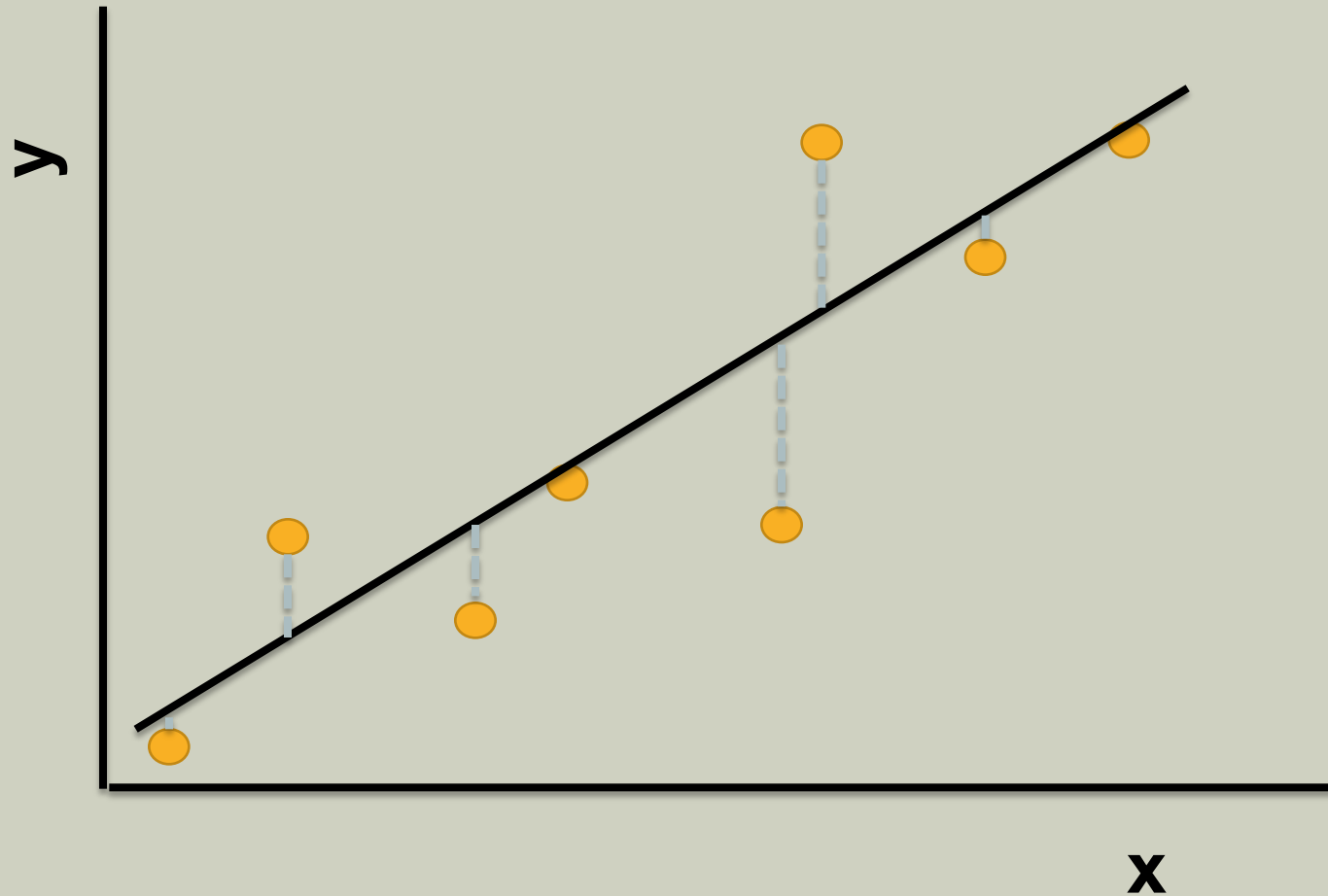
ONE VARIABLE PREDICTING ANOTHER



REGRESSION

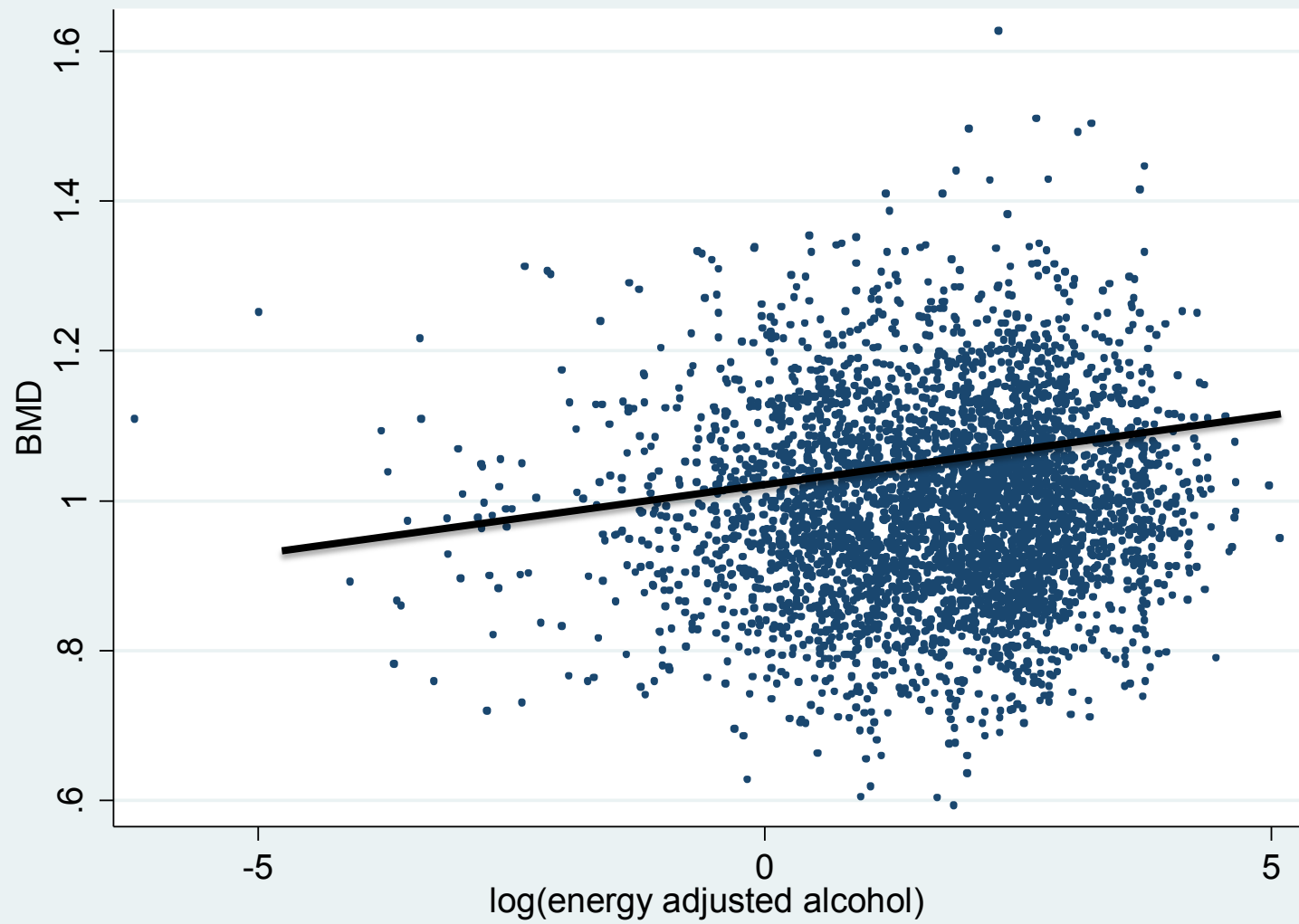
$$y = a + bx$$

REGRESSION

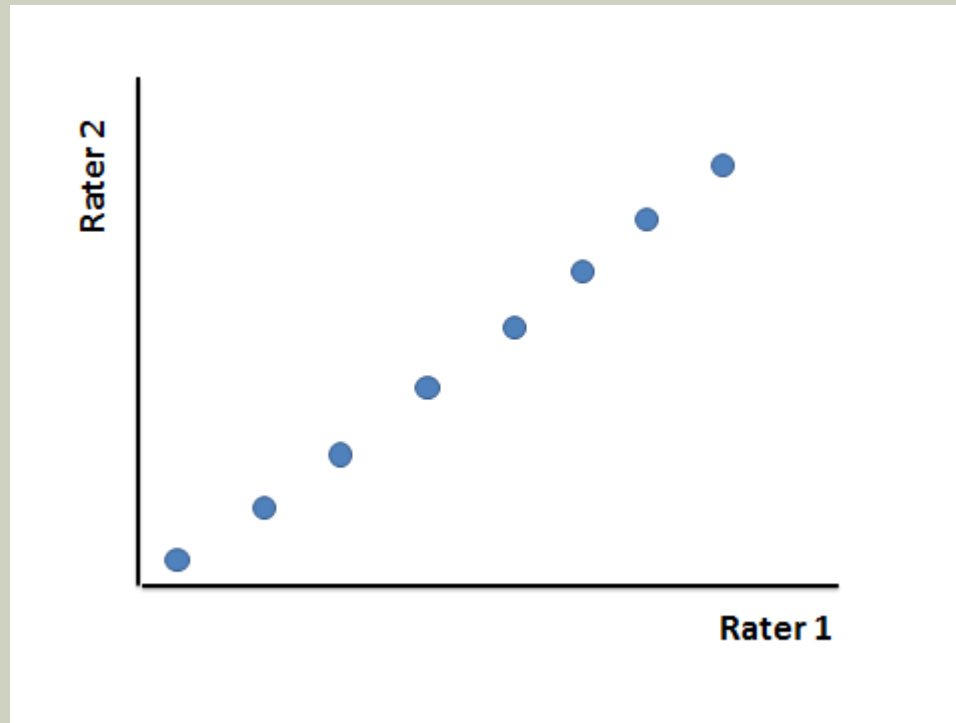


REGRESSION

- $y = a + bx$
- b is the slope of the line
- has dimensions
- Indicates how much variable (outcome) is influenced by another (exposure)

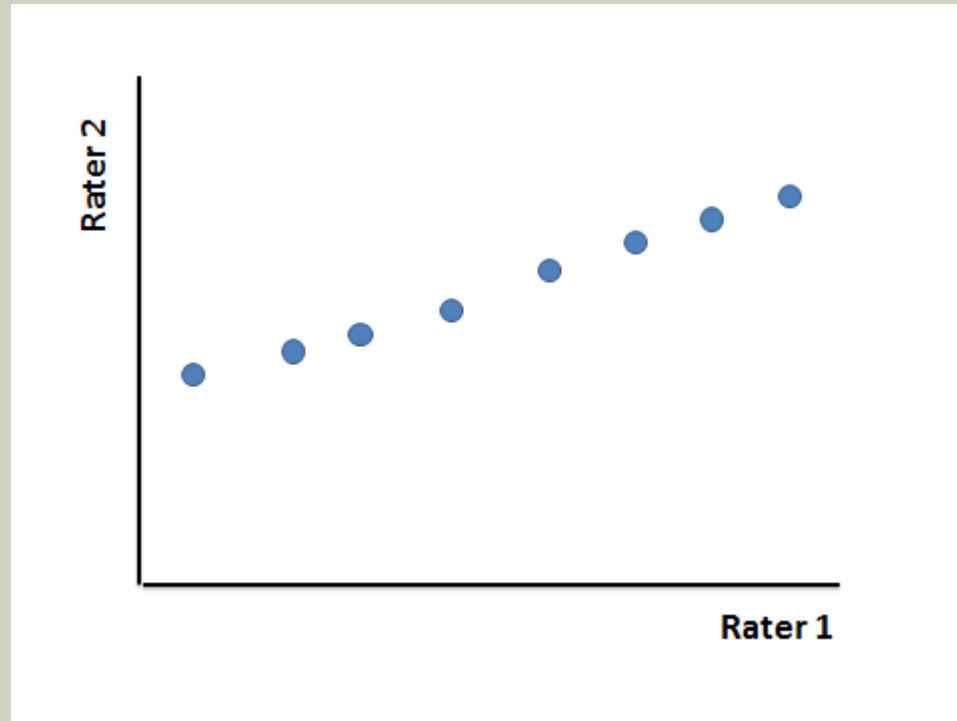


CORRELATION AND AGREEMENT



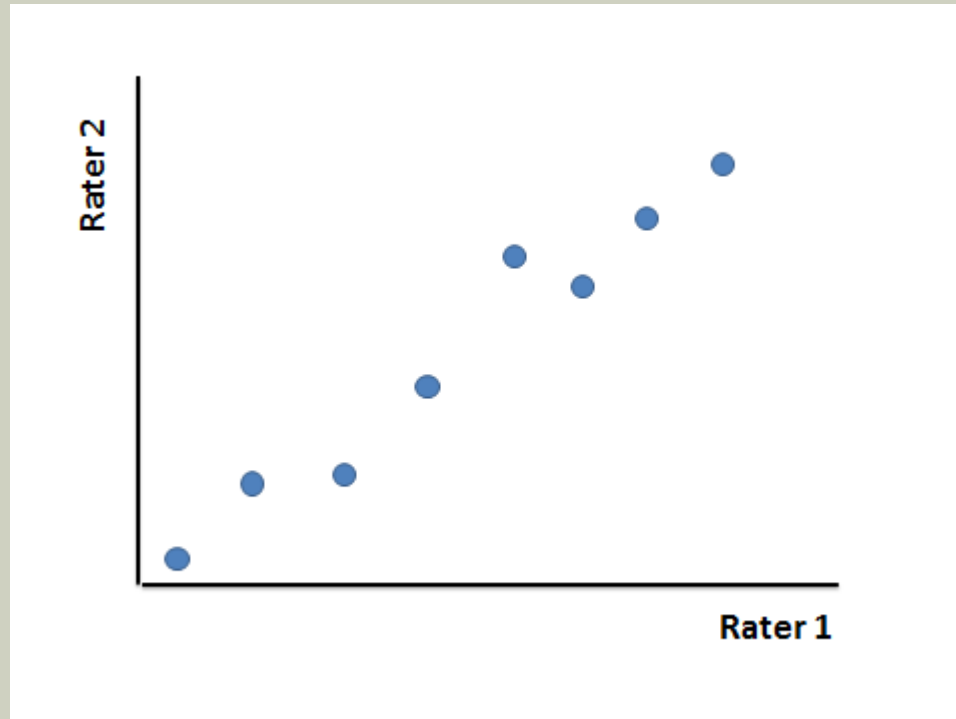
- Perfect linear correlation ($r = +1$)
- Perfect agreement in rater measurements

CORRELATION AND AGREEMENT



- Perfect linear correlation ($r=+1$),
- No agreement in rater's measurements

CORRELATION AND AGREEMENT



- **Intraclass correlation coefficient** is a measure of inter-rater agreement in actual measurements

SENSITIVITY OF A TEST

	D(+)	D(-)	
T(+)	TP	FP	
T(-)	FN	TN	
			N

- Positivity in disease
- $= TP / (TP + FN)$

SPECIFICITY

	D(+)	D(-)	
T(+)	TP	FP	
T(-)	FN	TN	
			N

- Negativity in health
- $= \text{TN} / (\text{FP} + \text{TN})$

POSITIVE PREDICTIVE VALUE

	D(+)	D(-)	
T(+)	TP	FP	
T(-)	FN	TN	
			N

- Risk of disease given a positive test
- $= TP / (TP + FP)$

NEGATIVE PREDICTIVE VALUE

	D(+)	D(-)	
T(+)	TP	FP	
T(-)	FN	TN	
			N

- Risk of no disease given negative test
- $TN/(FN+TN)$

CK AND RISK OF MI

	MI(+)	MI(-)	
CK \geq 80	215	248	453
CK<80	15	1822	1837
	230	2070	2300

- Sensitivity = $215/230$ = 93%
- Specificity = $1822/2070$ = 88%
- PPV = $215/453$ = 46%
- NPV = $1822/1837$ = 99%

SUMMARY

- TOPICS COVERED

- background to epidemiology
- observation
- study designs
 - case control
 - cohort
 - genetically informative designs
- epidemiological terms
- correlation regression
- sensitivity and specificity, PPPV, NPPV

IM FOR FUTURE



ANY QUESTIONS