Introduction to epidemiological terminology

Zekeriya Aktürk
zekeriya.akturk@gmail.com

Objectives

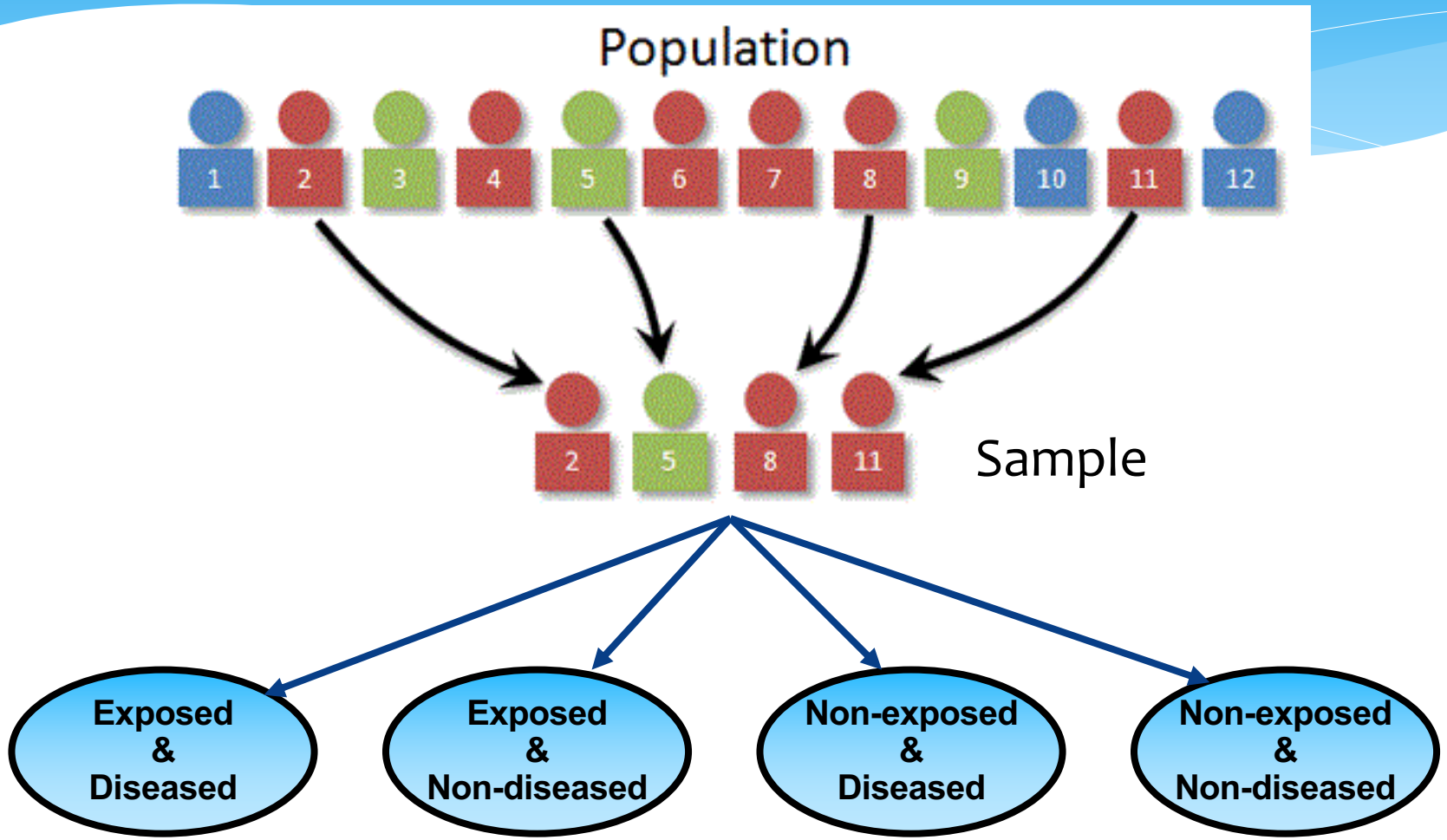
* At the end of this session, the participants are expected to;
  * Explain the basic features of different study designs
  * Explain the different epidemiological terms related with study designs
  * Explain the principles of causality in the context of EBM.

Study Designs

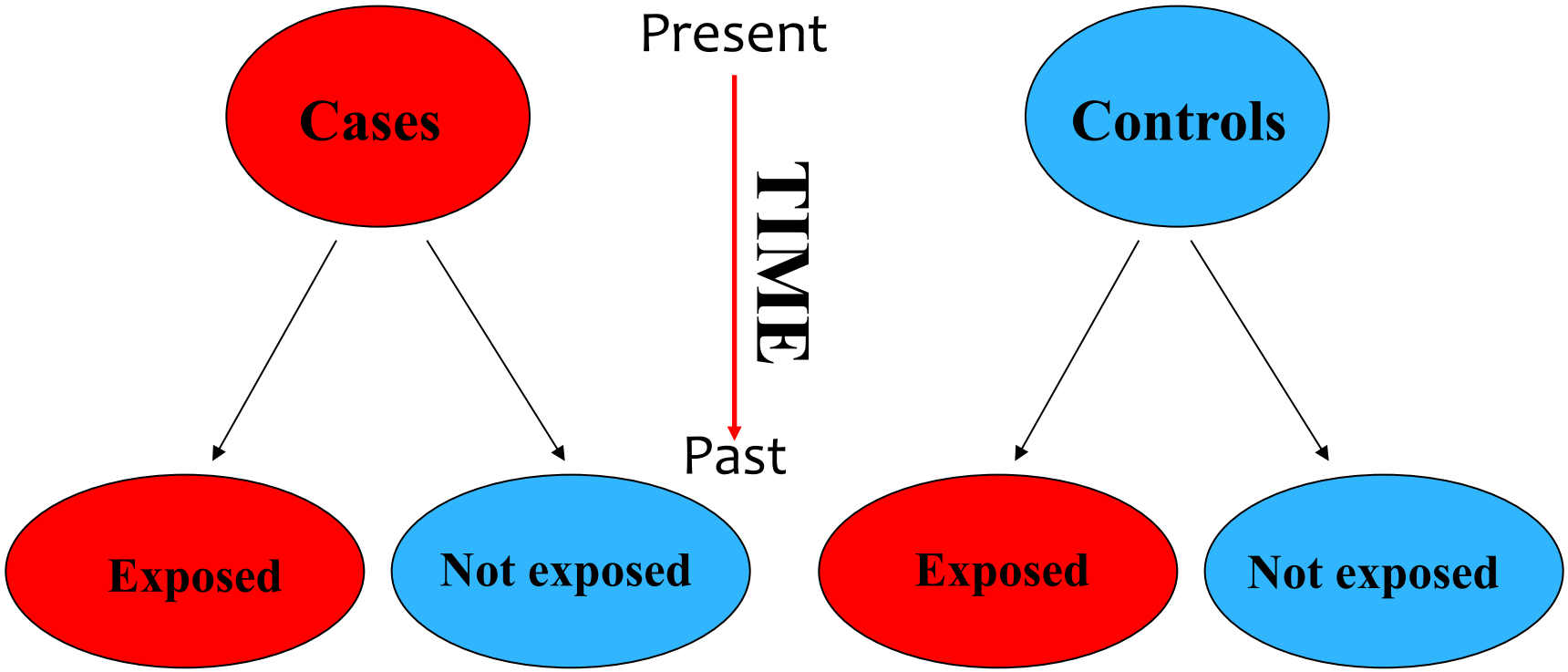
Type of study	Timing	Form	Action in past time	Action in present time (starting point)	Action in future time	Typical uses
Cross-sectional	Cross-sectional	Observational		Collect all information		• Prevalence estimates • Reference ranges and diagnostic tests • Current health status of a group
Repeated cross-sectional	Cross-sectional	Observational		Collect all information ▼ Collect all information ▼ Collect all information		• Changes over time
Cohort (Chapter 15)	Longitudinal (prospective)	Observational		Define cohort and assess risk factors → follow	Observe outcomes	• Prognosis and natural history (what will happen to someone with disease) • Aetiology
Case–control (Chapter 16)	Longitudinal (retrospective)	Observational	Assess risk factors ← trace	Define cases and controls (i.e. outcome)		• Aetiology (particularly for rare diseases)
Experiment	Longitudinal (prospective)	Experimental		Apply intervention → follow	Observe outcomes	• Clinical trial to assess therapy (Chapter 14) • Trial to assess preventative measure, e.g. large-scale vaccine trial • Laboratory experiment

Medical Statistics at a Glance, Eight Edition. Aviva Petrie and Caroline Sabin. 2009 John Wiley & Sons

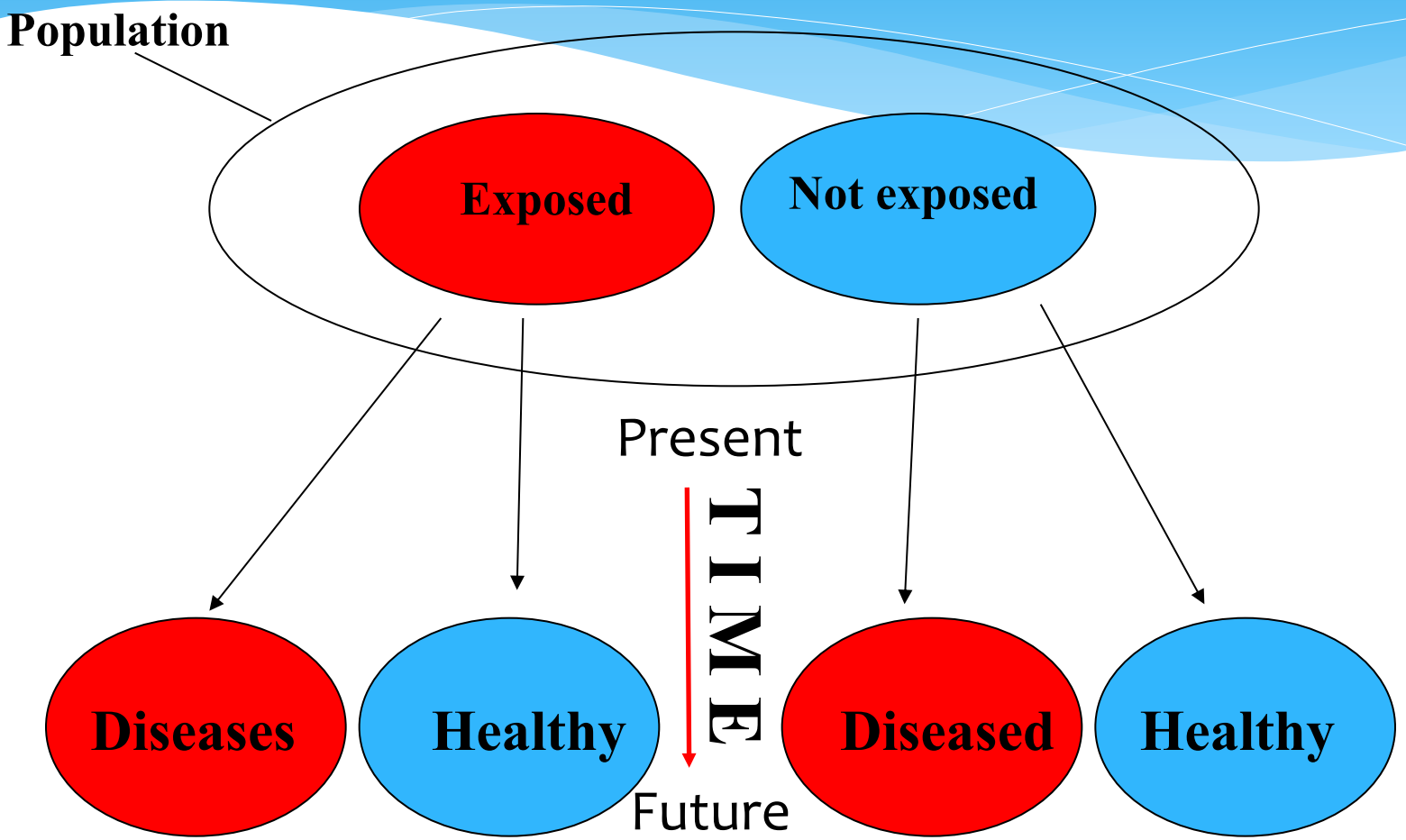
Cross-Sectional Study



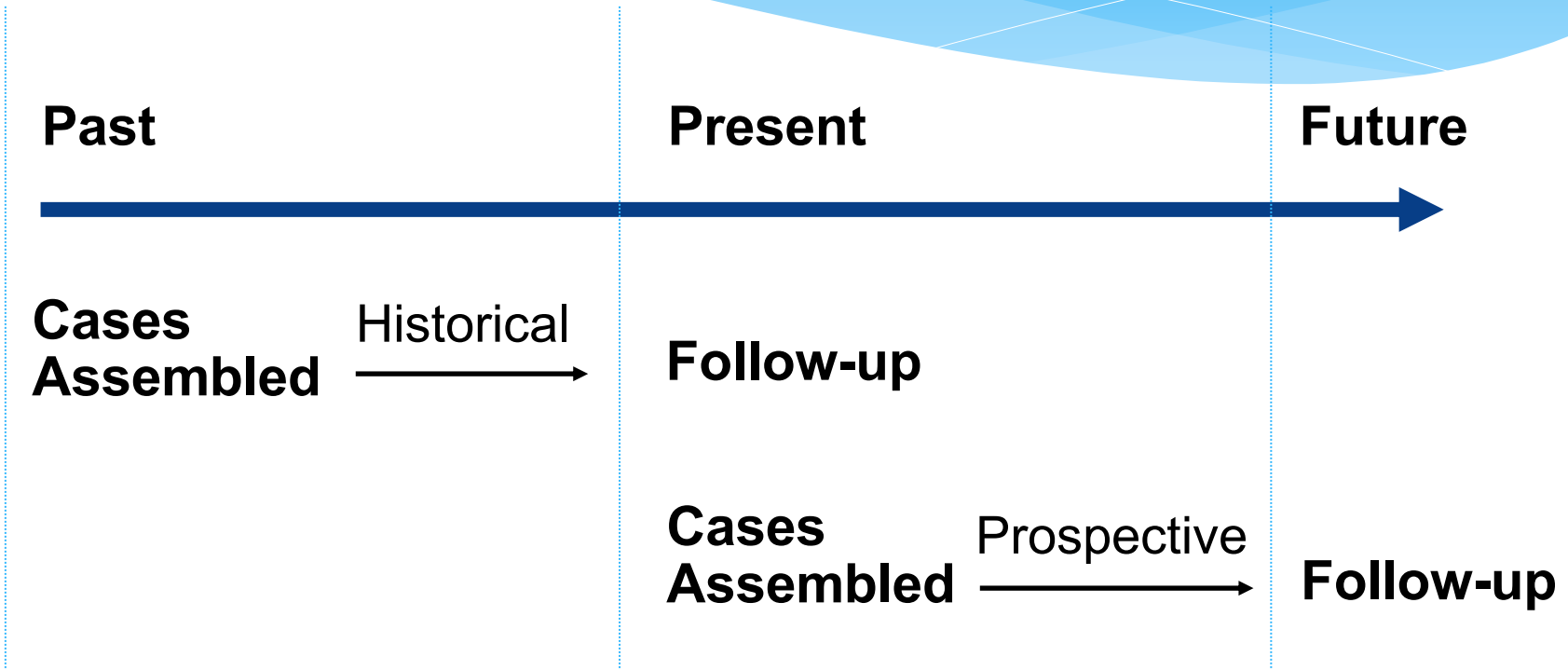
Case-control study



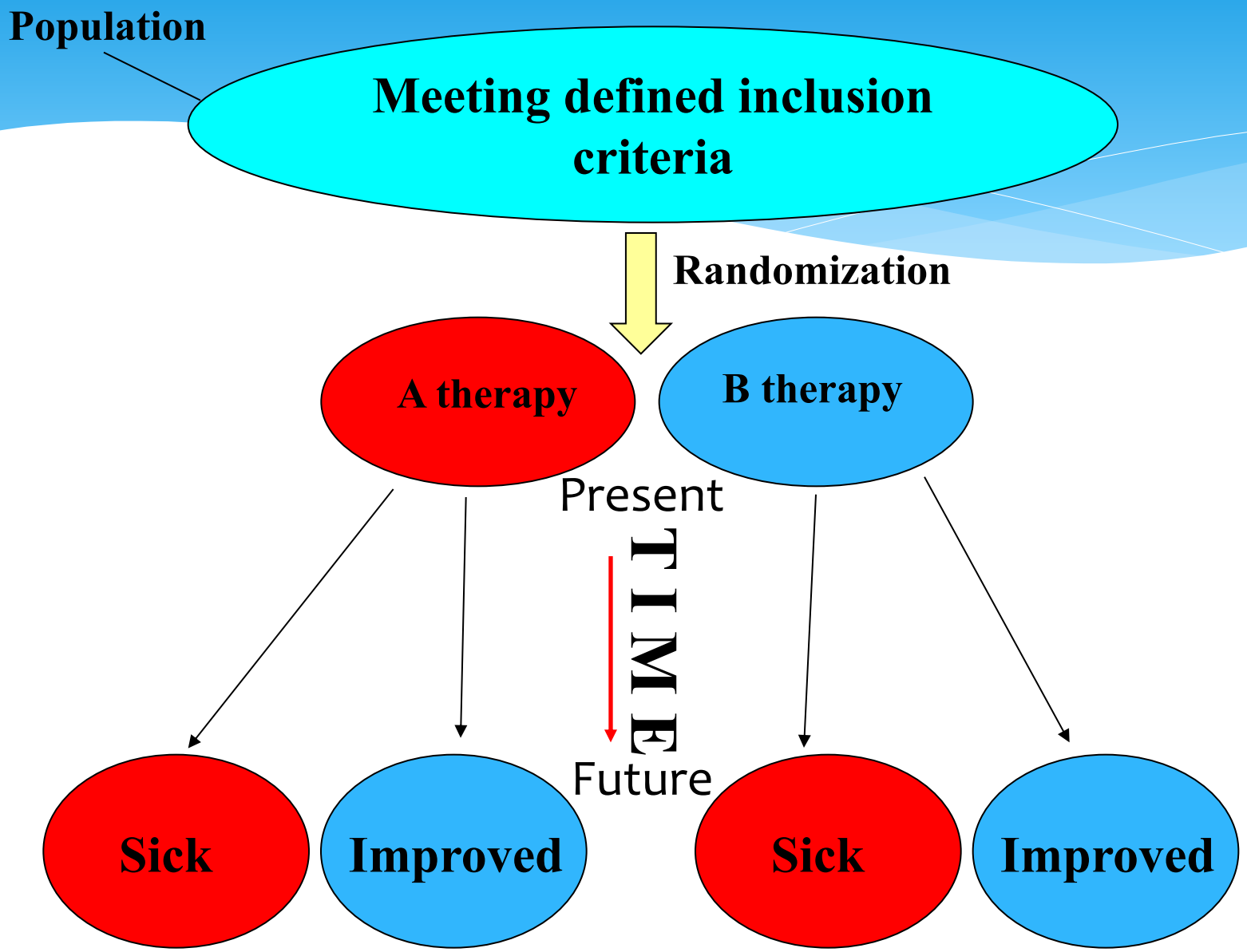
Cohort study



Historical (retrospective) and Concurrent (prospective) Cohort



Randomized controlled study



Incidence

$$\text{Incidence} = \frac{\text{\# of Persons Developing Disease}}{\text{\# of Persons at Risk}}$$

Prevalence

$$\text{Prevalence} = \frac{\text{\# of Persons With Disease}}{\text{\# of Persons at Risk of Having Disease}}$$

Relative Risk

$$\textbf{Relative Risk} = \frac{I_{exposed}}{I_{unexposed}}$$

I: incidence of disease

Risk reduction [edit]

Example of risk reduction

	Experimental group (E)	Control group (C)	Total
Events (E)	EE = 15	CE = 100	115
Non-events (N)	EN = 135	CN = 150	285
Total subjects (S)	ES = EE + EN = 150	CS = CE + CN = 250	400
Event rate (ER)	EER = EE / ES = 0.1, or 10%	CER = CE / CS = 0.4, or 40%	

Equation	Variable	Abbr.	Value
CER - EER	absolute risk reduction	ARR	0.3, or 30%
(CER - EER) / CER	**relative risk reduction**	RRR	0.75, or 75%
1 / (CER – EER)	number needed to treat	NNT	3.33
EER / CER	risk ratio	RR	0.25
(EE / EN) / (CE / CN)	odds ratio	OR	0.167
(CER - EER) / CER	preventable fraction among the unexposed	PFu	0.75

https://en.wikipedia.org/wiki/Relative_risk_reduction

Risk increase [edit]

	Example of risk increase		
	Experimental group (E)	Control group (C)	Total
Events (E)	EE = 75	CE = 100	175
Non-events (N)	EN = 75	CN = 150	225
Total subjects (S)	ES = EE + EN = 150	CS = CE + CN = 250	400
Event rate (ER)	EER = EE / ES = 0.5, or 50%	CER = CE / CS = 0.4, or 40%	

Equation	Variable	Abbr.	Value
EER – CER	absolute risk increase	ARI	0.1, or 10%
(EER – CER) / CER	relative risk increase	RRI	0.25, or 25%
1 / (EER – CER)	number needed to harm	NNH	10
EER / CER	risk ratio	RR	1.25
(EE / EN) / (CE / CN)	odds ratio	OR	1.5
(EER – CER) / EER	attributable fraction among the exposed	AF_e	0.2

Cohort Study

		Disease: Yes	Disease: No	Total
Exposure	Yes	a	b	a + b
	No	c	d	c + d
Total		a + c	b + d	a + b + c + d

Incidence among exposed: a/ (a+b)

Incidence among unexposed: c/ (c+d)

a = # exposed and have the disease
b = # exposed and do not have the disease
c = # not exposed and have the disease
d = # both not exposed and do not have the disease

Cohort Study

Smoking	Lung Cancer: Yes	Lung Cancer: No	Total
Yes	124	9150	9274
No	3	5269	5272
Total	127	14419	14546

Incidence among exposed:
124/9274 = 1.34%
Incidence among unexposed:
3/5272 = 0.06%

124 = # exposed and have the disease
9150 = # exposed and do not have the disease
3 = # not exposed and have the disease
5269 = # both exposed and have the disease

Relative Risk

Relative Risk represents the increased risk of disease among exposed persons as compared with unexposed persons

Incidence of lung cancer:

Smokers 124/9274 Non-smokers 3/5272

RR = (124/9274)/(3/5272) = (13.4/1000)/(0.6/1000) = 23.5

Physicians who smoke are at 23.5 times the risk of developing lung cancer compared to men who don't smoke

Attributable Risk

Relative Risk represents the increased risk of disease among exposed persons as compared with unexposed persons

Attributable Risk is the excess risk of disease in the exposed as compared to the unexposed

Incidence of lung cancer:

Smokers 124/9274 Non-smokers 3/5272

$$AR = \frac{124}{9274} - \frac{3}{5272} = \frac{13.4}{1000} - \frac{0.6}{1000} = \frac{12.8}{1000}$$

For every 1000 men that smoke, there are an additional 12.8 deaths due to lung cancer

Tobacco and Lung Cancer

A.

Cigarette Smoking	Rate of Lung Cancer
Yes	140/100,000/year
No	10/100,000/year

Risk difference (AR):
140 – 10 =130/100,000/yr

Risk ratio (RR): 140/10 = 14.0

B.

Cigarette Smoking	Rate of Ischemic Heart Disease
Yes	669/100,000/year
No	413/100,000/year

Risk difference (AR):
669 – 413 = 256/100,000/yr

Risk ratio (RR): 669/413 = 1.6

Case Control Study

		Disease: Yes	Disease: No	Total
Exposure	Yes	a	b	a + b
	No	c	d	c + d
Total		a + c	b + d	a + b + c + d

a = # exposed and have the disease
b = # exposed and do not have the disease
c = # not exposed and have the disease
d = # both non-exposed and non-diseased

Odds vs. Probability

* Probability = p, the chance of an event
  * Range from 0-1
* Odds = p/(1-p)
  * Range from 0-∞
* Example: probability = 0.5 (flip a coin), odds = 1
* Example: probability = 0.1, odds = 1/9
* Example: probability = 0.975, odds = 39

Odds vs. Probability

Exposure	Cases	Controls
Yes	300	500
No	100	200
	400	700

Odds of exposure for cases:
300/100 = 3.0
Probability of exposure for cases:
300/400 = 0.75

Odds of exposure for controls:
500/200 = 2.5
Probability of exposure for controls:
500/700 = 0.71

OR is the ratio of the two odds:

$$\text{Odds ratio} = \frac{300/100}{500/200} = \frac{(300)\,(200)}{(100)\,(500)} = \frac{300\,/\,500}{100\,/\,200} = 1.2$$

(risk ratio = 0.75/0.71 = 1.05)

Why we calculate the OR?

		Disease: Yes	Disease: No	Total
Exposure	Yes	a	b	a + b
	No	c	d	c + d
	Total	a + c	b + d	a + b + c + d

If the disease is "rare" in the population, then the number of cases in the exposed (a) and non exposed (c) is small. So, the number of exposed persons (a + b) ≅ b, and the number of unexposed persons (c + d) ≅ d.

Example: OR vs. RR

		Disease: Yes	Disease: No	Total
Exposure	Yes	6	99,994	a + b
	No	3	99,997	c + d
Total		a + c	b + d	a + b + c + d

If the disease is "rare" in the population, then the number of cases in the exposed (a) and non exposed (c) is small. So, the number of exposed persons (a + b) ≅ b, and the number of unexposed persons (c + d) ≅ d.

$$RR = \frac{6 / (99{,}994 + 6)}{3 / (99{,}997 + 3)} = 2.0 \cong \frac{6 / 99{,}994}{3 / 99{,}997} = 2.0001 = OR$$

Case Control Study-Example

Data from a case-control study of current oral contraceptive (OC) use and myocardial infarction in premenopausal female nurses

***Data from L.Rosenberg et al., Oral contraceptive use in relation to non-fatal myocardial infarction. *Am. J. Epidemiol*. 111:59, 1980**

		Myocardial infarction		
		Yes	No	Total
Current OC use	Yes	23	304	327
	No	133	2816	2949
	Total	156	3120	3276

Because the cases and controls are selected by the investigator, it's not possible to calculate incidence rates and the RR. However, the RR can be approximated by the odds ratio (OR).

$$OR = \frac{(23)\,(2816)}{(133)\,(304)} = 1.6$$

OC users are 1.6 times as likely (or 60% greater odds) to have had an M.I. as are non-OC users.

Diagnostic Tests

	Disease	
	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Sensitivity = TP/(TP+FN)
  * Pr(Positive test given that disease is truly present)
  * If no FN, sensitivity = 100%
  * High sensitivity means a negative test helps RULE OUT disease (SnNOut mnemonic)

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Specificity = TN/(TN+FP)
  * Pr(Negative test given that disease is truly not present)
  * If no FP, specificity = 100%
  * High specificity means a positive test helps RULE IN disease (SpPIn mnemonic)

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Sensitivity and specificity are inversely related
  * If we make it harder to diagnose a disease (say, use higher level of blood glucose to diagnose diabetes), we will have more FN but more TP: sensitivity decreases, but specificity increases.

Tradeoff in Sensitivity and Specificity in Diagnosing Diabetes

Blood Sugar Level 2 hr after Eating (mg/100 mL)	Sensitivity (%)	Specificity (%)
70	98.6	8.8
80	97.1	25.5
90	94.3	47.6
100	88.6	69.8
110	85.7	84.1
120	71.4	92.5
130	64.3	96.9
140	57.1	99.4
150	50.0	99.6
160	47.1	99.8
170	42.9	100.0
180	38.6	100.0
190	34.3	100.0
200	27.1	100.0

[a] Public Health Service. Diabetes program guide. Publication no. 506. Washington, DC: U.S. Government Printing Office, 1960.

ROC Curve

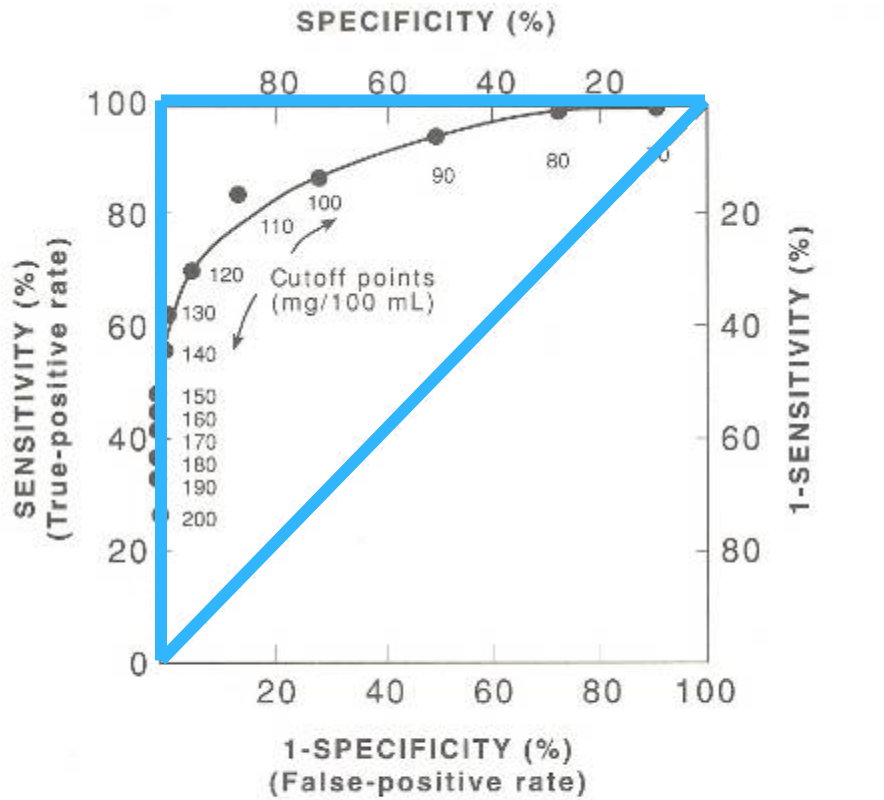


Figure 3.4. A ROC curve. The accuracy of 2-hr postprandial blood sugar as a diagnostic test for diabetes mellitus. (Data from Public Health Service. Diabetes program guide. Publication no. 506. Washington, DC: U.S. Government Printing Office, 1960.)

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Positive predictive value (PPV)
* PPV = TP/(TP+FP)
  * Pr(Disease is present given that test was positive)
  * More clinically relevant – this is what we want to know when treating a patient!!!

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Negative predictive value (NPV)
* NPV = TN/(TN+FN)
  * Pr(Disease is not present given that test was negative)
  * More clinically relevant – this is what we want to know when treating a patient!!!

Diagnostic Tests: The 2x2 Table

	Truly Disease +	Truly Disease -
Test +	TP	FP
Test -	FN	TN

* Prevalence = proportion of the population that actually has the disease
  * **Prevalence has dramatic effect on PPV and NPV**
  * **With low prevalence, PPV will be low even for tests with high sensitivity and specificity**
* (TP+FN)/total number tested
* Accuracy = (TP+TN)/total number tested

Likelihood Ratio

* Converts a pre-test probability to a post-test probability
* Compares the likelihood of a positive result in someone with the disease as compared with someone without the disease (or vice-versa)
* Incorporates both sensitivity and specificity

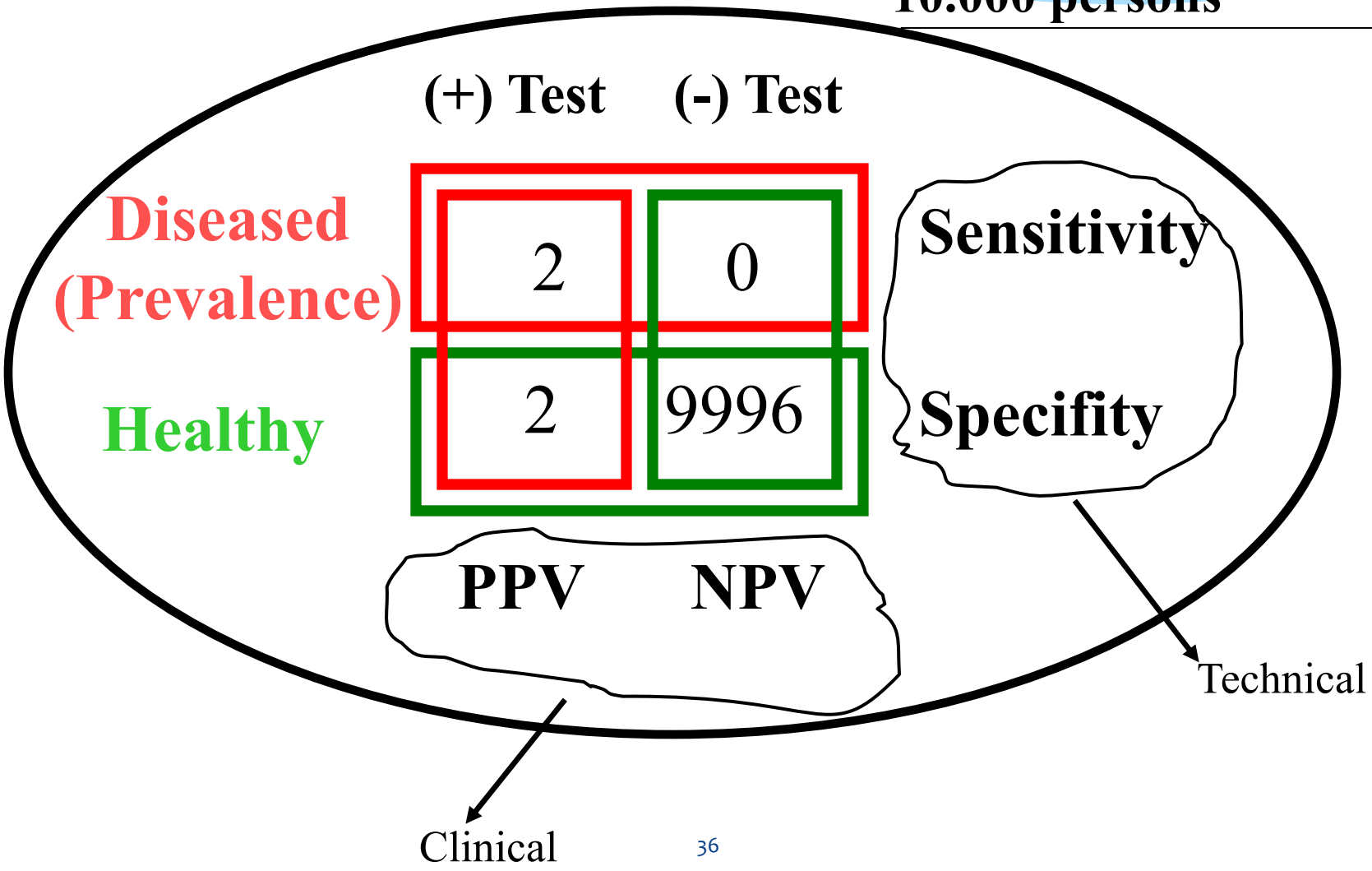
$$LR^{+}=Sensitivity/(1-Specifity)=1/(1-0.9998)=5000$$

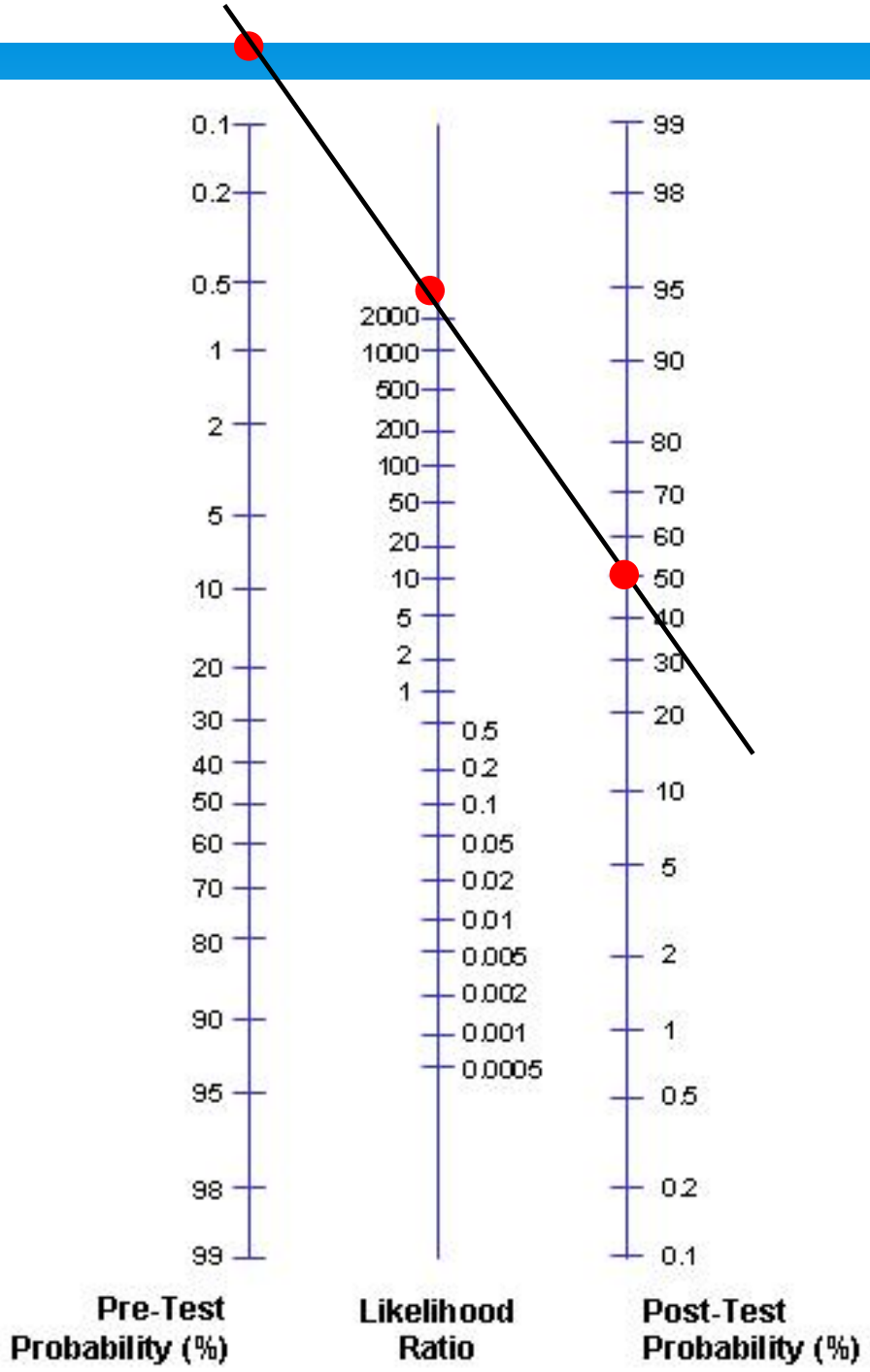
Likelihood Ratio

Qualitative Strength	LR(+)	LR(-)
Excellent	10	0.1
Very good	6	0.2
Fair	2	0.5
Useless	1	1

Mayer D. Essential Evidence-based Medicine. Cambridge University Press. 2004.

Technical vs. Clinical Precision




0.1
0.2
0.5
1
2
5
10
20
30
40
50
60
70
80
90
95
98
99
2000
1000
500
200
100
50
20
10
5
2
1
0.5
0.2
0.1
0.05
0.02
0.01
0.005
0.002
0.001
0.0005
99
98
95
90
80
70
60
50
40
30
20
10
5
2
1
0.5
0.2
0.1
Pre-Test Probability (%)
Likelihood Ratio
Post-Test Probability (%)

Likelihood Ratios and Odds

* Rather than performing the cumbersome calculations, a nomogram can be used (Fagan nomogram).
* It is more important to know how a given likelihood ratio is likely to impact a pre-test probability, something we will discuss further when we analyze studies of diagnostic test performance.

Assessing Causality

1. Strength: how large is the association?
2. Consistency: is exposure observed repeatedly in different environments?
3. Specificity: does one exposure lead to one outcome?
4. Temporality: does exposure precede outcome?
5. Dose-response: does risk increase as exposure increases?
6. Biologic plausibility: consistent with known science?
7. Coherence between epidemiological and laboratory findings.
8. Experiment: It is possible to appeal to experimental evidence.
9. Analogy: is the association similar to established similar ones?
10. Reversibility: does risk decrease after exposure is removed?

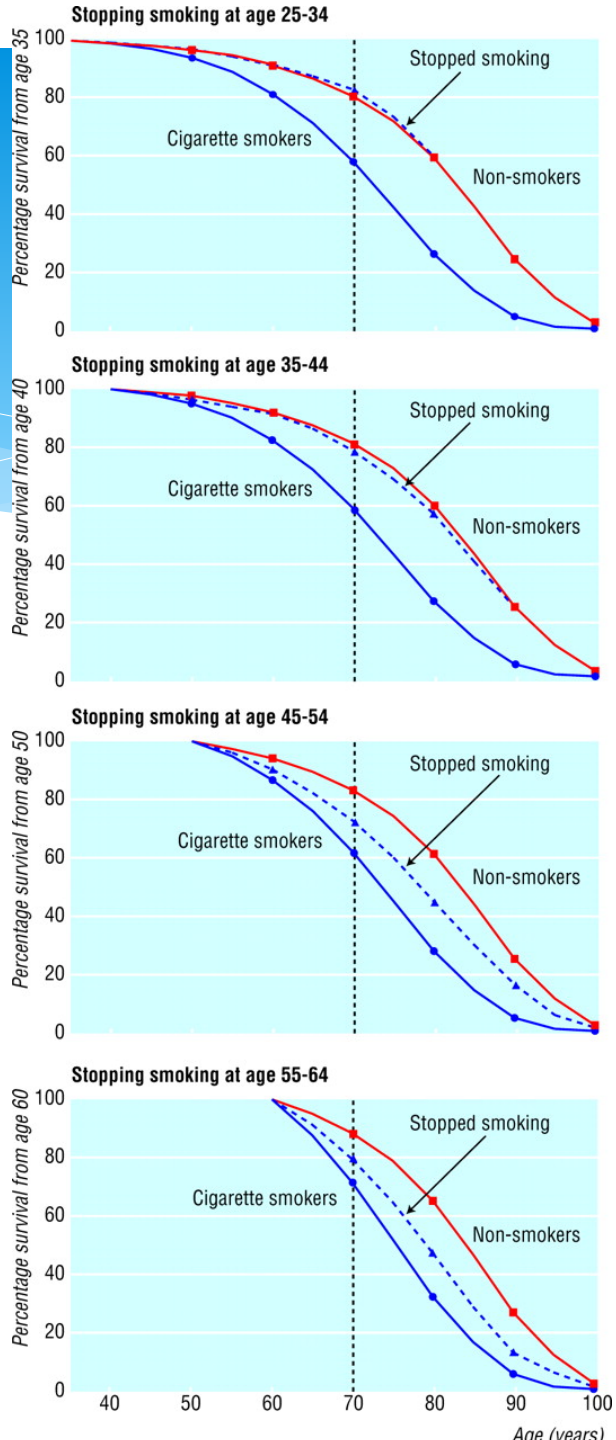
https://en.wikipedia.org/wiki/Bradford_Hill_criteria

Dose-Response Example

* Doll R et al., Mortality in relation to smoking: 20 years' observations on male British doctors. Be Med J. 1976;2:1525-36.
* Lung cancer deaths per million men per year:
  * Non-smokers: 10
  * 1-14 cigarettes per day: 78
  * 15-24 cigarettes per day: 127
  * 25+ cigarettes per day: 251

Reversibility Example

* Mortality rate relative to never-smokers, by years since stopping smoking:
  * 0: 15.8
  * <5: 10.7
  * 5-9: 5.9
  * 10-14: 4.7
  * 15+: 2.0



Doll R et al., Mortality in relation to smoking: 20 years' observations on male British doctors. BMJ. 1976;2:1525-36.

Summary

* What are the different study designs and their features
* Please explain the following terms:
  * Incidence, prevalence
  * Sensitivity, specificity, PPV, NPV
  * absolute risk reduction, RRR, NNT
  * OR, relative risk
* Explain the principles of causality in the context of EBM.