Clinical Epidemiology

Lecture notes 2015 - Part I

Lecture 1
Diagnostic Process
Interobserver agreement

Diagnostic Process

Simplified assumptions

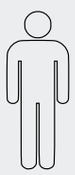
Patient

physical exam
history
laboratory data

benefits from treatment

does not benefit from treatment

has disease



does not have disease

D+ probability (D+) D-

Dichotomous test

positive test

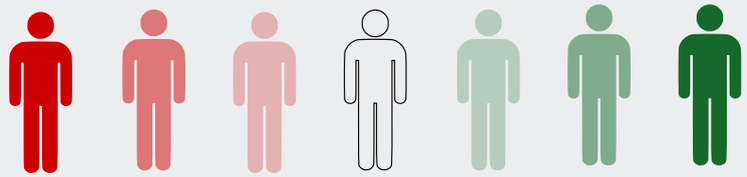
negative test

Reality

Patient

physical exam
history
laboratory data

has disease



does not have disease

spectrum of disease

prior probability might
affect test characteristics

Multilevel and continuous test

more than two test results

continuous test can be dichotomized
with arbitrary cut-point
BUT information is lost

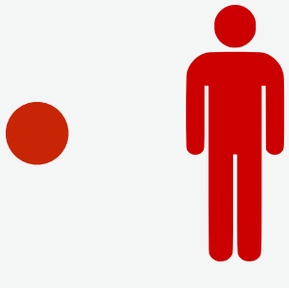
Evaluating diagnostic tests

Reliability
repetitions of the test
give the same result

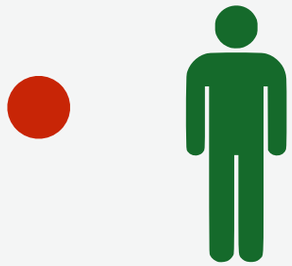
reliable

not reliable

Accuracy
test gives
the right answer

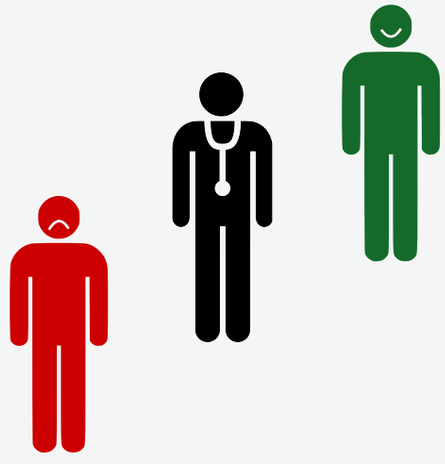


accurate

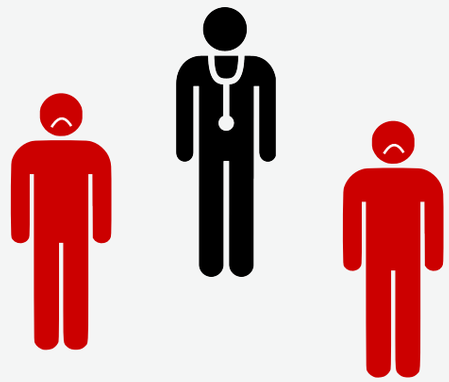


inaccurate

Usefulness
right answer improves
outcome by favorably
affecting decision



useful



useless

Value
expected improvement
in health outcomes
justifies
the risk and costs

Inter-observer reliability (KAPPA)

Sensitivity
Specificity
ROC curve

Treating threshold

Evaluating diagnostic tests

inter-observer agreement

Observer 1

Observer 2

	result 1	result 2	result 3	Total
result 1	a	b	c	**R1**
result 2	d	e	f	**R2**
result 3	g	h	i	**R3**
Total	**C1**	**C2**	**C3**	**N**

add along the diagonal

observed agreement (%) = concordance = (a+e+i)/N

observed agreement can be misleading: if two observers both know an abnormality is uncommon, they can have nearly perfect agreement just by never or rarely saying that it is present

assess for unbalanced disagreement by looking for asymmetry across the diagonal

i.e. observer 1 is the more concerned radiologist and is more likely to diagnose abnormalities on CXR's

Kappa is not particularly useful in assessing systematic disagreement

this is based on the marginals

expected agreement (%) = $(R_1 \times C_1/N + R_2 \times C_2/N + R_3 \times R_3/N)/N$

the amount of agreement beyond what would be expected from the marginals (row and column totals)

$$\text{Kappa} = \frac{\text{observed agreement} - \text{expected agreement}}{100\% - \text{expected agreement}}$$

weighted Kappa

for **ordinal** variable with more than 2 values, gives partial credit for 2 observers being close

i.e. partial credit for b, d, f and h

Evaluating diagnostic tests

inter-observer agreement: example

two doctors are estimating the size of a lesion

L	M
M	S
M	M
L	M
L	L
S	S
L	M
M	M
L	L
L	M

sorted

S	S
M	S
M	M
M	M
L	M
L	M
L	M
L	M
L	L
L	L

there is highly unbalanced disagreement

	S	M	L	Total
S	1	1	0	**2**
M	0	2	4	**6**
L	0	0	2	**2**
Total	**1**	**3**	**6**	**10**

observed agreement = (1+2+2)/10 = 0.5

expected agreement = (2 x 1/10) + (6 x 3/10) + (2 x 6/10) = 0.32

Kappa = (0.5-0.32)/(1-0.32) = 0.26

Lecture 2
Dichotomous Tests

Diagnostic Process

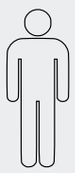
Simplified assumptions

Patient

physical exam
history
laboratory data

benefits from treatment

has disease



does not benefit from treatment

does not have disease

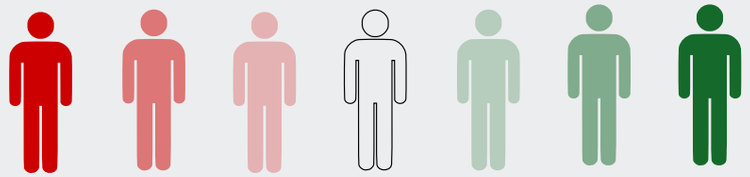
D+ probability (D+) D-

Reality

Patient

physical exam
history
laboratory data

has disease



does not have disease

spectrum of disease

prior probability might
affect test characteristics

Dichotomous test

positive test

negative test

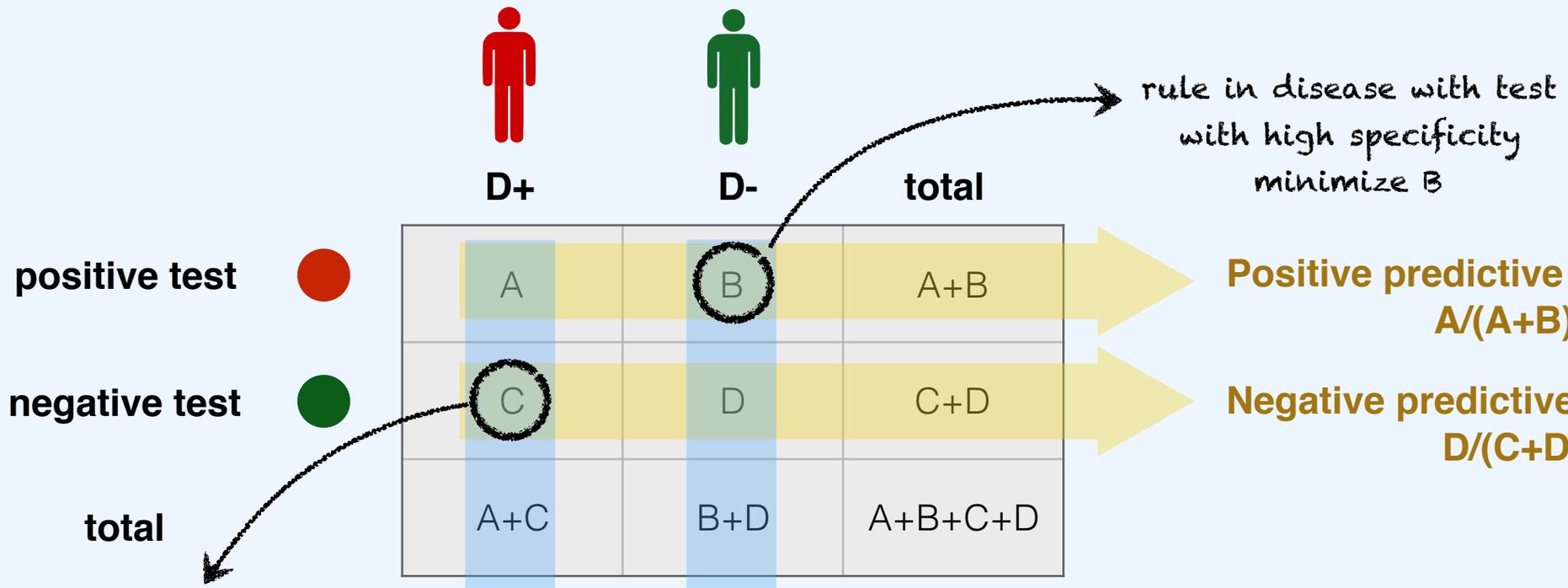
Multilevel and continuous test

more than two test results

continuous test can be dichotomized
with arbitrary cut-point
BUT information is lost

Dichotomous test

Cross-sectional sampling



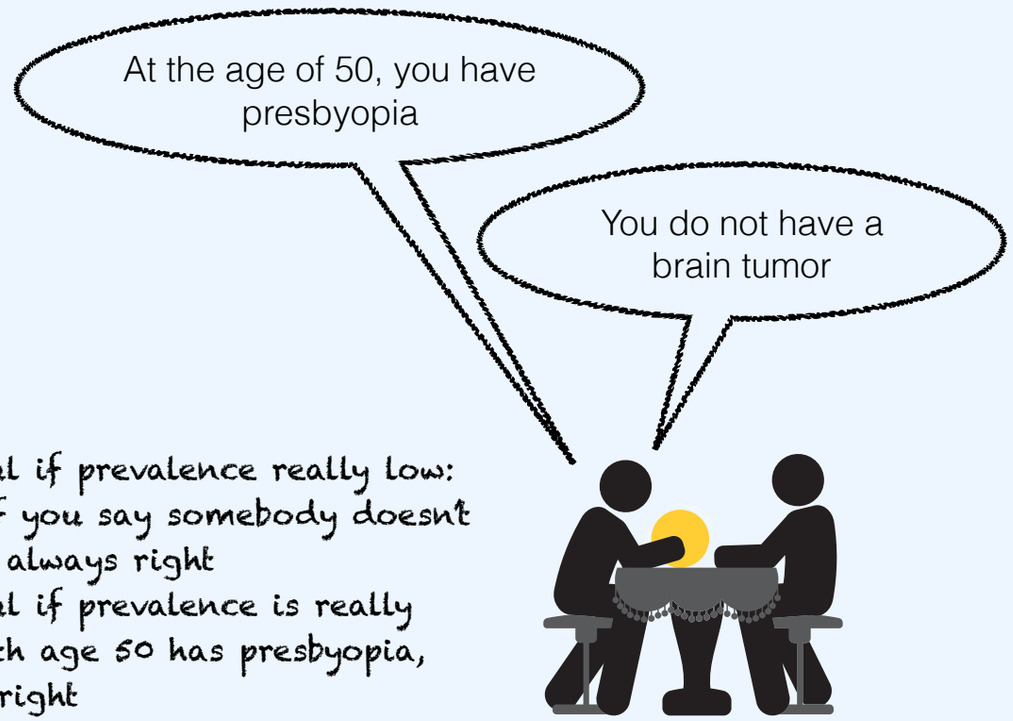
Prevalence = (A+C)/(A+B+C+D)

how many patients have disease in a population

Accuracy = (A+D)/(A+B+C+D)

how many times is test right

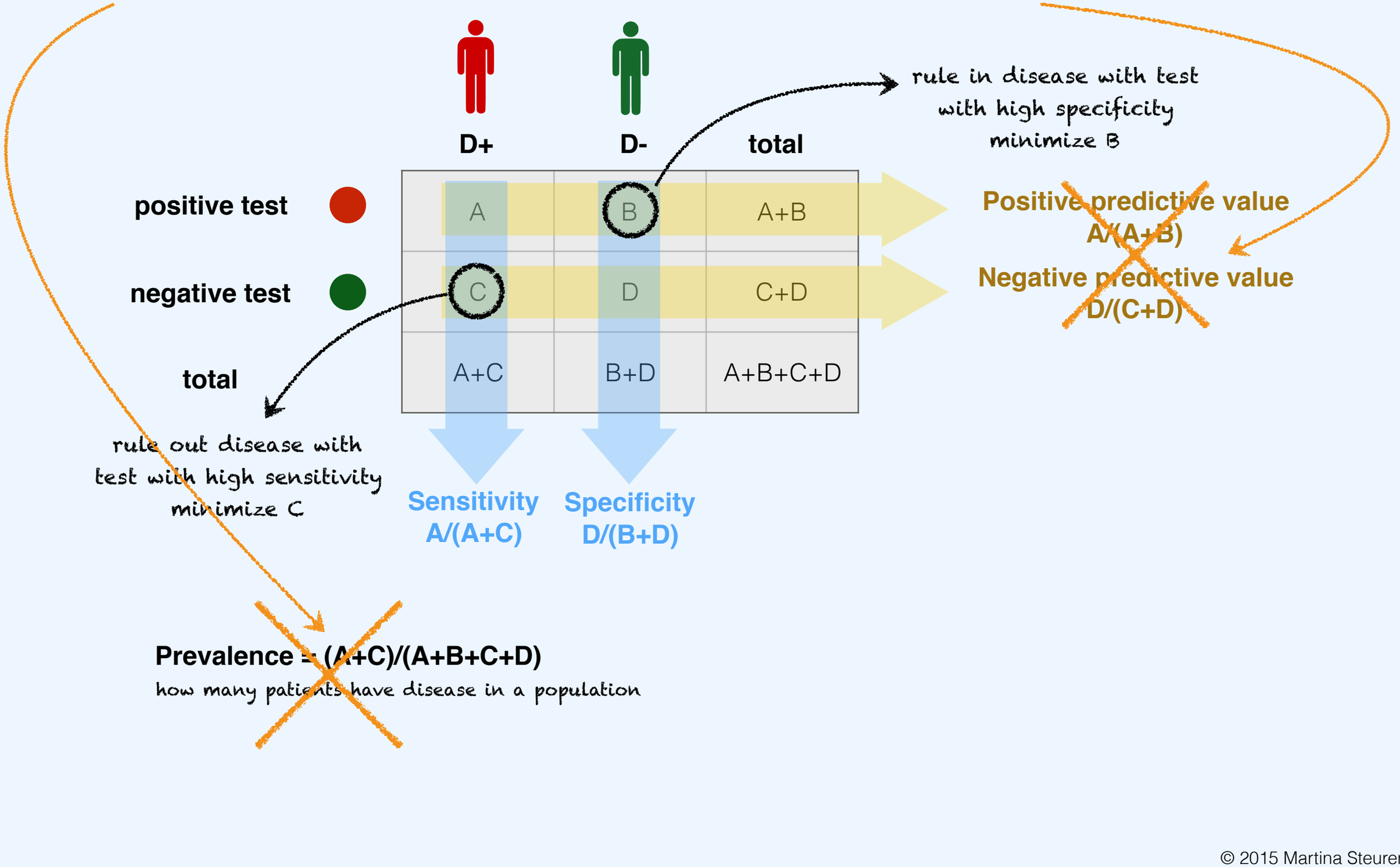
the fortune teller secret:
1. accuracy is not very useful if prevalence really low: screening for brain tumor, if you say somebody doesn't have it, then you are almost always right
2. accuracy is not very useful if prevalence is really high: if you say a person with age 50 has presbyopia, then you are almost always right



Dichotomous test

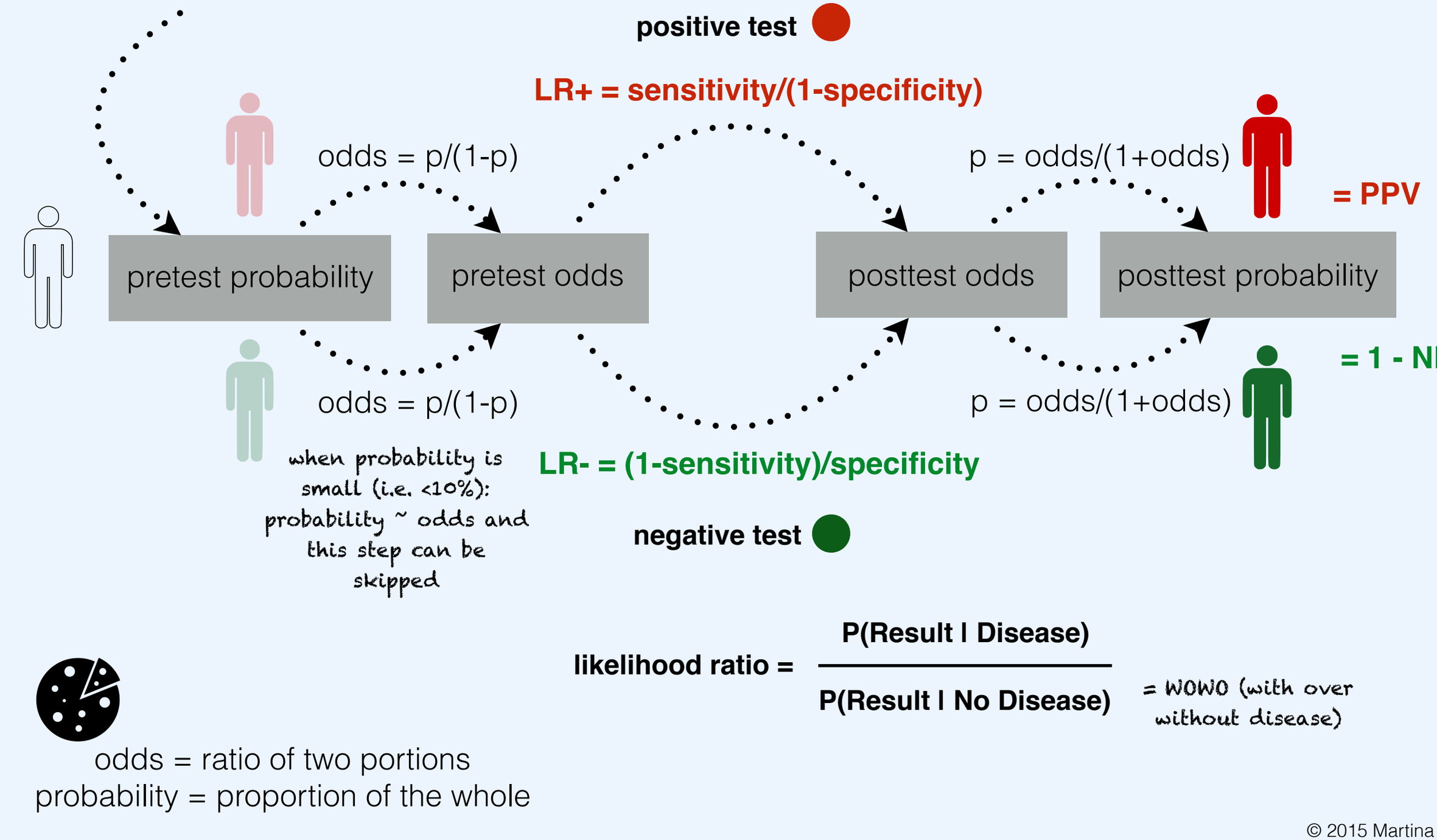
Case-control sampling

Ratio between people with and without disease is artificially set by investigator



Prior and posterior probabilities

prevalence = pretest probability for screening test
prevalence and patient presentation = pretest probability in diagnostic test



$$\text{likelihood ratio} = \frac{P(\text{Result} \mid \text{Disease})}{P(\text{Result} \mid \text{No Disease})}$$

= WOWO (with over without disease)

odds = ratio of two portions
probability = proportion of the whole

Treatment and Testing thresholds

in the green area test is not useful because prior probability is so low that even after a positive test result your posterior probability of disease would still be below treatment threshold

treatment threshold (PTT) = lowest probability of disease at which the expected benefits of treatment exceed its expected risk and costs

in the brown area test is not useful because prior probability is already so high that you can just treat (i.e. setting of an epidemic)

cost of treatment x probability of unnecessary treatment (no disease)
C x (1-p)

how bad is it to treat someone who does not have the disease (=C)

net benefit foregone x probability of failure to offer treatment (disease)
B x p

how bad is it not to treat someone who does have the disease (=B)

No treat test threshold:
below this threshold the prior probability is so low that it is impossible to get above the PTT after the test, thus it is not worth testing, just don't treat

in gray area test is useful because prior probability close enough to treatment threshold that test result will impact decision

Treat-test threshold:
above this it is not worth testing because the prior probability is already so high that it is ok just to treat

Lecture 3
Multilevel and continuous test

Diagnostic Process

Simplified assumptions

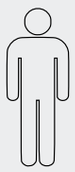
Patient

physical exam
history
laboratory data

benefits from treatment

does not benefit from treatment

has disease



does not have disease

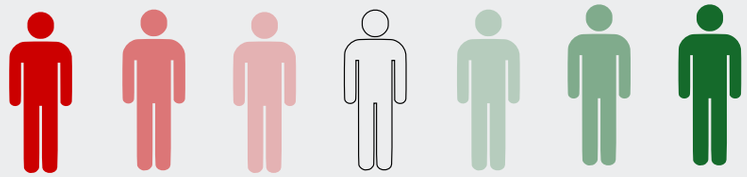
D+ probability (D+) D-

Reality

Patient

physical exam
history
laboratory data

has disease



does not have disease

spectrum of disease

prior probability might affect test characteristics

Dichotomous test

LR+ = sensitivity/(1-specificity)

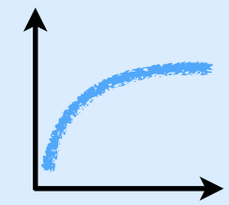
LR- = (1-sensitivity)/specificity

positive test

negative test

continuous test can be dichotomized with arbitrary cut-point
BUT information is lost

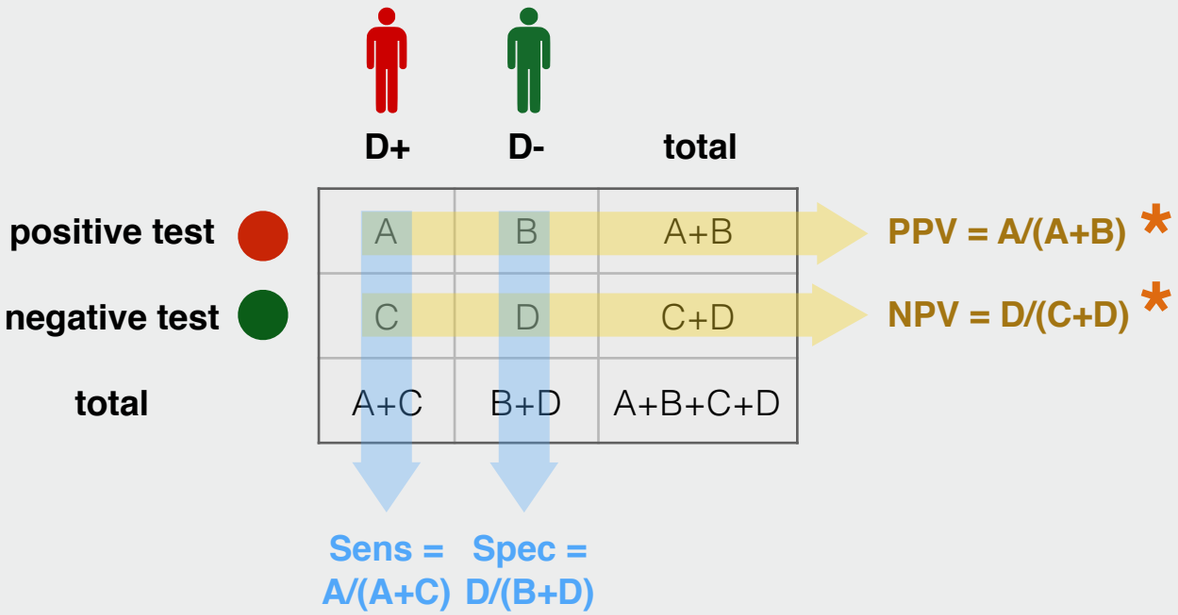
Multilevel and continuous test



more than two test results

caveat:
never calculate LR+ and LR- for test with more than 2 results, in these cases calculate interval LR, LR equals slope of ROC curve in that interval

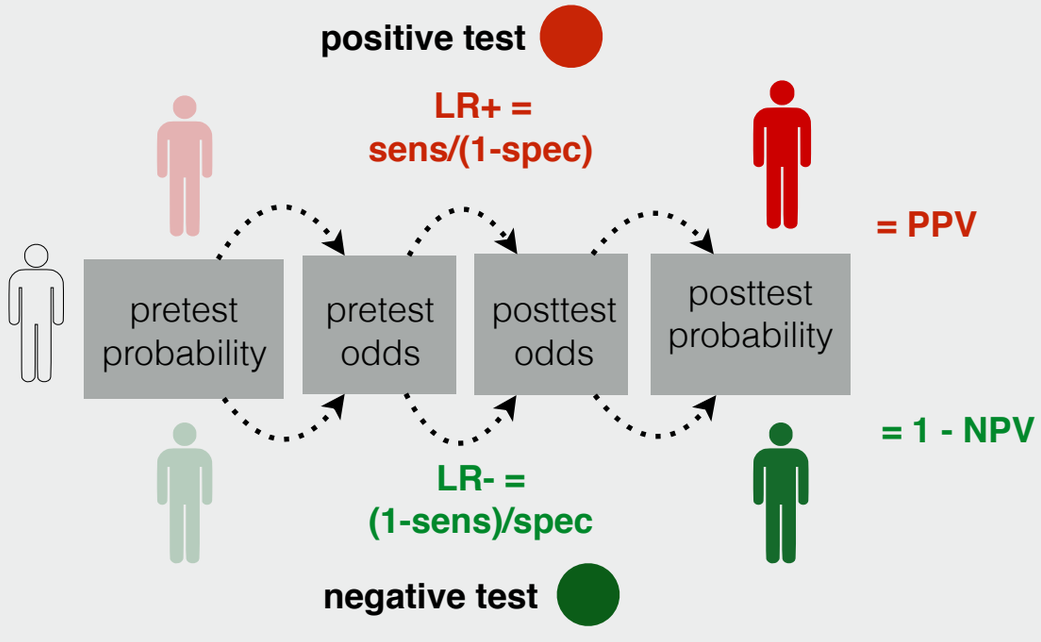
Dichotomous test



Prevalence = (A+C)/(A+B+C+D) *

Accuracy = (A+D)/(A+B+C+D) *

*not useful in case-control studies because prevalence set by investigator with case-control ratio



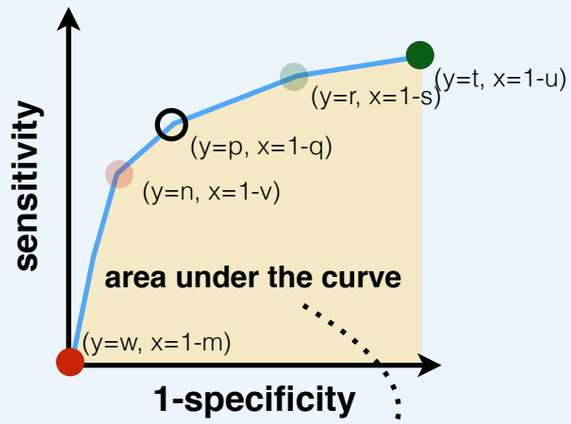
Multilevel and continuous test

ROC table

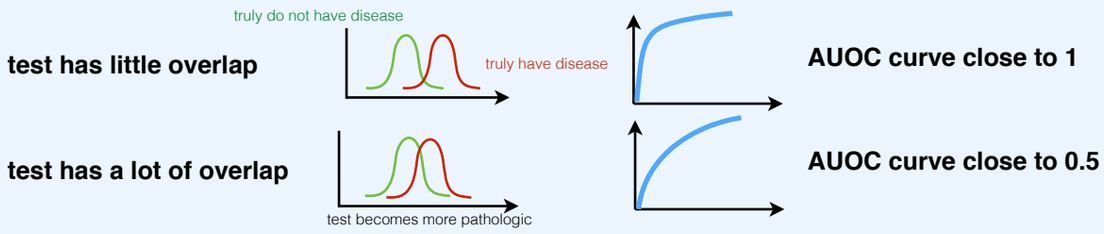
test result	sensitivity	1-specificity
●	w	m
●	n	v
○	p	q
●	r	s
●	t	u

don't report LR in an ROC table, report LR in LR table

ROC curve

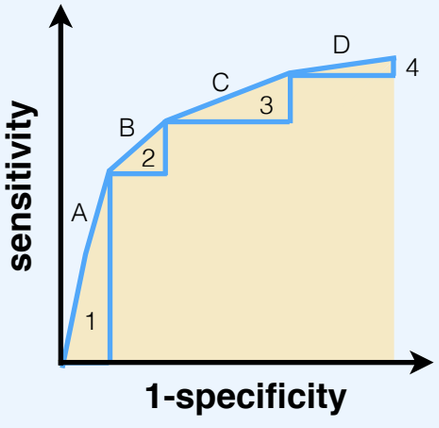


summary measure of discriminatory ability



LR table

test interval	% of D+	% of D-	interval LR
1	e	f	A
2	g	h	B
3	i	j	C
4	k	l	D



Interval LR represent slope of curve in this interval