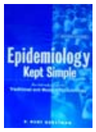
Epidemiology Kept Simple



Chapter 4

Screening for Disease

Gerstman Chapter 4 1

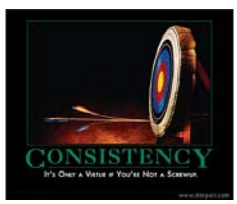
Identifying Cases

- "Case" ≡ someone who truly has the condition we are looking for
- "Diagnostic test" ≡ any method used to detect a cases (not just medical tests)

Gerstman Chapter 4 2

Reproducibility & Validity

- Reproducibility and validity are *different* aspects of accuracy
- **Reproducibility** ≡ agreement upon repetition, i.e., consistency
- **Validity** ≡ ability to discriminate accurately



Gerstman Chapter 4 3

§4.2 Reproducibility

- Two independent raters classify each patient as either positive or negative
- Cross-tabulate results

	Rater B		
Rater A	**+**	**−**	**Total**
+	*a*	*b*	p_1
−	*c*	*d*	q_1
Total	p_2	q_2	*N*

We quantify reproducibility with this **kappa (κ)** statistic

$$\kappa = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1}$$

Gerstman Link to Formula sheet 4

Kappa Interpretation

κ quantifies agreement *above chance*

- $\kappa = 1$ → perfect agreement
- $0.7 < \kappa < 1$ → excellent agreement
- $0.3 < \kappa < 0.7$ → fair agreement
- $\kappa < 0.3$ → poor agreement
- $\kappa \approx 0$ → random agreement
- $\kappa = -1$ → perfect disagreement

Gerstman Chapter 4 5

Example: Kappa

To what extent are these results reproducible?

	Rater B		
Rater A	D+	D−	Total
D+	20	4	24
D−	5	71	76
Total	25	75	100

$$\kappa = \frac{2(ad - bc)}{p_1 q_2 + p_2 q_1} = \frac{2[(20)(71) - (4)(5)]}{[(24)(75) + (25)(76)]} = 0.76$$

∴ Excellent agreement

Gerstman Chapter 4 6

§4.3 Validity

- Compare test results to **gold standard**
- Each patient is classified as either true positive (TP), true negative (TN), false positive (FP), or false negative (FN)
- Crosstab results

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

Gerstman Chapter 4 7

Sensitivity

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

Sensitivity (SEN) ≡ proportion of cases that test positive

$$\text{SEN} = \frac{\text{TP}}{\text{those w/disease}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Gerstman Chapter 4 8

Specificity

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

Specificity (SPEC) ≡ proportion of noncases that test negative

$$\text{SPEC} = \frac{\text{TN}}{\text{those w/out disease}} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

Gerstman Chapter 4 9

Predictive Value Positive

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

Predictive value positive (PVP) ≡ proportion of positive tests that are actually cases

$$\text{PVP} = \frac{\text{TP}}{\text{those who test positive}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Gerstman Chapter 4 10

Predictive Value Negative

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

Predictive value negative (PVN) ≡ proportion of negative tests that are actually non-cases

$$\text{PVN} = \frac{\text{TN}}{\text{those who test negative}} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

Gerstman Chapter 4 11

Prevalence

Test	D+	D−	Total
T+	TP	FP	TP+FP
T−	FN	TN	FN+TN
Total	TP+FN	FP+TN	*N*

- **[True] prevalence** = (TP + FN) / N
- **Apparent prevalence** = (TP + FP) / N

Gerstman Chapter 4 12

Conditional Probabilities

- Pr(A|B) ≡ "the probability of A *given* B", e.g., Pr(T+|D+) ≡ "probability test positive *given* disease positive"
- SEN = Pr(T+|D+)
- SPEC ≡ Pr(T−|D−)
- PVP = Pr(D+|T+)
- PVN= Pr(D−|T−)

Gerstman Chapter 4 13

Example: Low Prevalence Population

Conditions: *N* = 1,000,000; Prevalence = .001

	D+	D−	Total
T+			
T−			
Total	1000		1,000,000

Prevalence = (those with disease) / N

Therefore:

(Those with disease) = Prevalence × N

= .001× 1,000,000 = 1000

Gerstman Chapter 4 14

Example: Low Prevalence Population

Number of non-cases, i.e., TN + FP

	D+	D−	Total
T+			
T−			
Total	1000	999,000	1,000,000

1,000,000 – 1,000 = 999,000

Gerstman Chapter 4 15

Example: Low Prevalence Population

Assume test SENsitivity = .99, i.e.,
Test will pick up 99% of those with disease

	D+	D−	Total
T+	990		
T−			
Total	1000		

TP = SEN × (TP + FN)
= 0.99 × 1000
= 990

Gerstman Chapter 4 16

Example: Low Prevalence Population

It follows that:

	D+	D−	Total
T+	990		
T−	10		
Total	1000		

FN = 1000 – 990 = 10

Gerstman Chapter 4 17

Example: Low Prevalence Population

Suppose test SPECificity = .99
i.e., it will correctly identify 99% of the noncases

	D+	D−	Total
T+			
T−		989,010	
Total		999,000	

TN = SPEC × (TN + FP)
= 0.99 × 999,000
= 989,010

Gerstman Chapter 4 18

Example: Low Prevalence Population

It follows that:

	D+	D−	Total
T+		9,990	
T−		989,010	
Total		999,000	

FPs = 999,000 – 989,010 = 9,900

Gerstman Chapter 4 19

Example: Low Prevalence Population

It follows that the Predictive Value Positive is :

	D+	D−	Total
T+	990	9,990	10,980
T−	10	989,010	989,020
Total	1000	999,000	1,000,000

PVP = TP / (TP + FP) = 990 / 10,980 = 0.090
Strikingly low PVP!

Gerstman Chapter 4 20

Example: Low Prevalence Population

It follows that the Predictive Value Negative is:

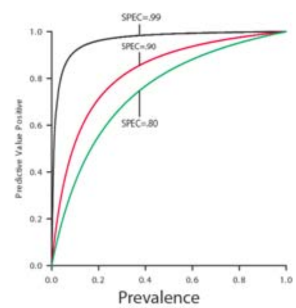
	D+	D−	Total
T+	990	9,990	10,980
T−	10	989,010	989,020
Total	1000	999,000	1,000,000

PVN = TN / (TN + FP) = 989010 / 999000 = 0.99

Gerstman Chapter 4 21

PVP and Prevalence

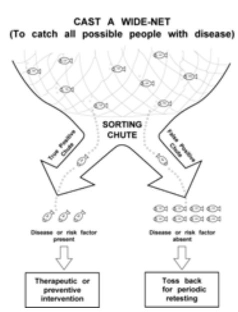
- PVP a function of
  - PREValence
  - SENsitivity
  - Specificity
- Figure shows relation between PVP, PREV, & SPEC (test SEN = constant .99)



Gerstman Chapter 4 24

Screening Strategy

- First stage ⇒ high SENS (few cases missed)
- Second stage ⇒ high SPEC (sort out false positives from true positives)



Gerstman Chapter 4 25

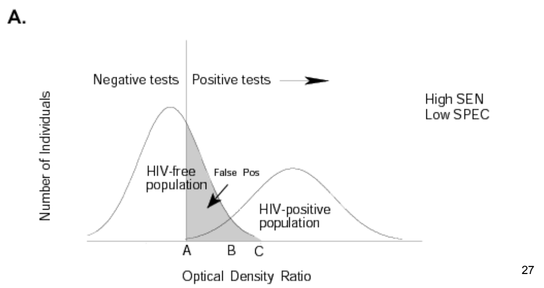
Cutoff Point Concept

- Sensitivity and specificity are influenced by they cutoff point used to determine positive results
- Example: Immunofluorescence HIV optical density ratio
- At what point do we say optical density is sufficiently high to say the test is positive?

Gerstman Chapter 4 26

Low Cutoff

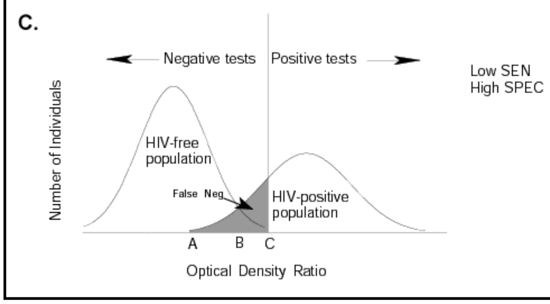
High sensitivity and low specificity



27

High Cutoff

Low sensitivity and high specificity



Intermediate Cutoff

moderate sensitivity & moderate specificity