Principles of Assessing Diagnostic Imaging Tests

7/2022

Jerry Jarvik, M.D., M.P.H.

Professor of Radiology, Neurological Surgery and Health Services

Adjunct Professor Orthopedic Surgery & Sports Medicine and Pharmacy

Co-Director, Comparative Effectiveness, Cost and Outcomes Research Center (CECORC)

Director, UW Clinical, Learning, Evidence And Research (CLEAR) Center for Musculoskeletal Conditions

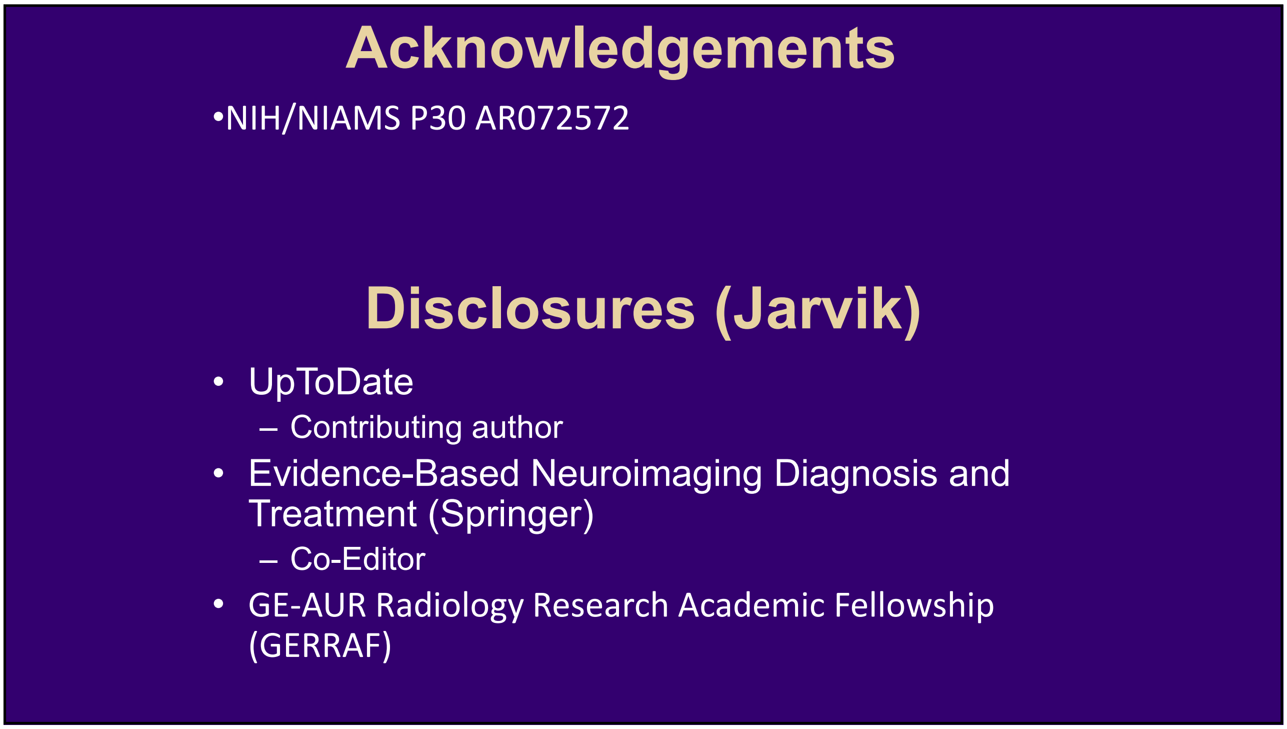
1

U.S. Department of Health and Human Services

Supported by the

NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases

2



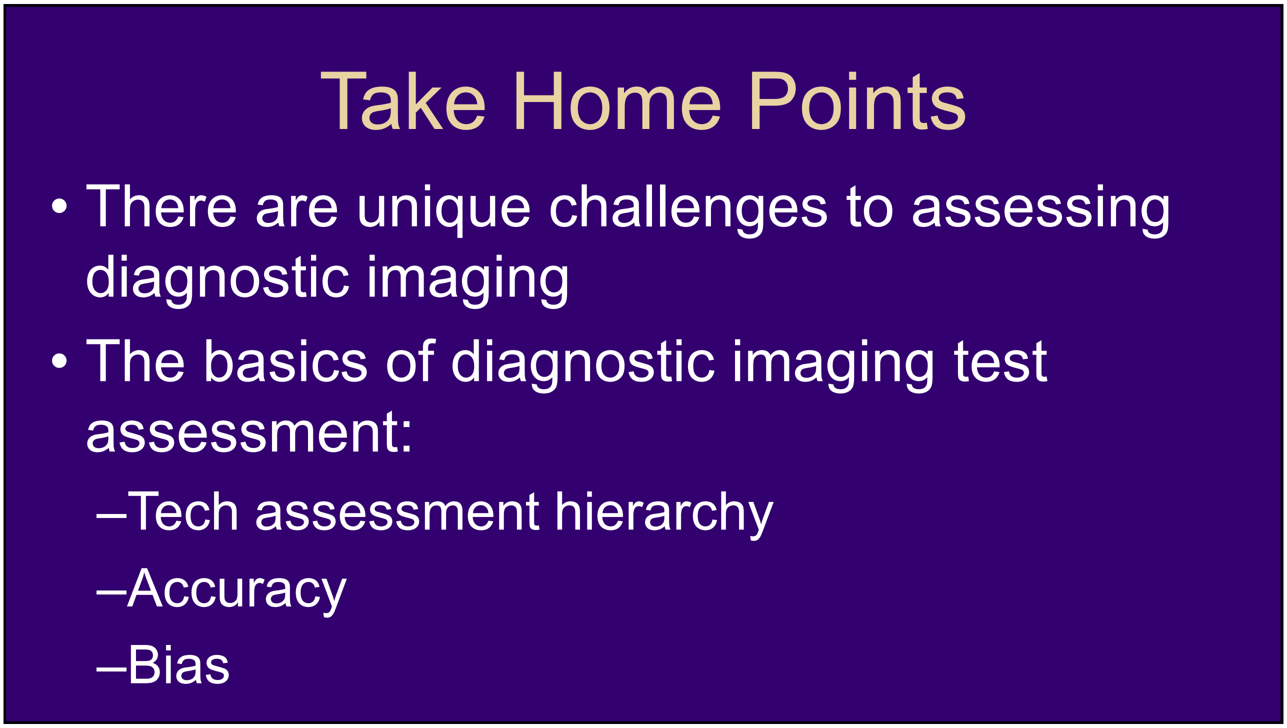
Acknowledgements

•NIH/NIAMS P30 AR072572

Disclosures (Jarvik)

- UpToDate
  - Contributing author
- Evidence-Based Neuroimaging Diagnosis and Treatment (Springer)
  - Co-Editor
- GE-AUR Radiology Research Academic Fellowship (GERRAF)

3



Take Home Points

- There are unique challenges to assessing diagnostic imaging
- The basics of diagnostic imaging test assessment:
  - Tech assessment hierarchy
  - Accuracy
  - Bias

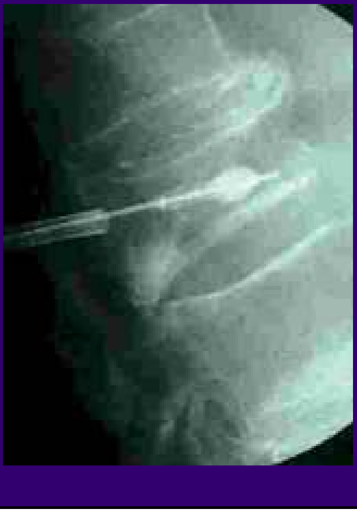
4

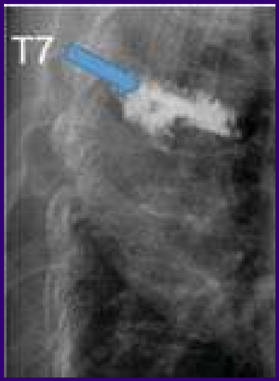
Take Home Points

- There are unique challenges to assessing diagnostic imaging
- The basics of diagnostic imaging test assessment:
  - –Tech assessment hierarchy
  - –Accuracy
  - –Bias

5

Typical Approach for Therapeutic Interventions: Kyphoplasty



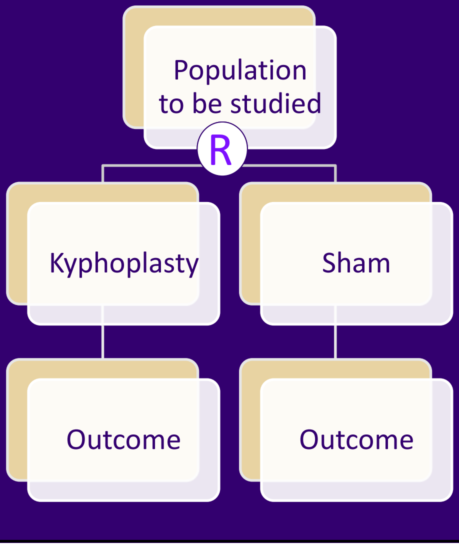


From Liu et al: Clinical Efficacy of Kyphoplasty with Zoledronic Acid of Osteoporotic Vertebral Fxs J invest Surg 2019

6

Simplified Study Design: INKTEST

Investigative Kyphoplasty Efficacy and Safety Trial



7

Reminder- No trial is simple!

Proposed INKTEST PICOT format

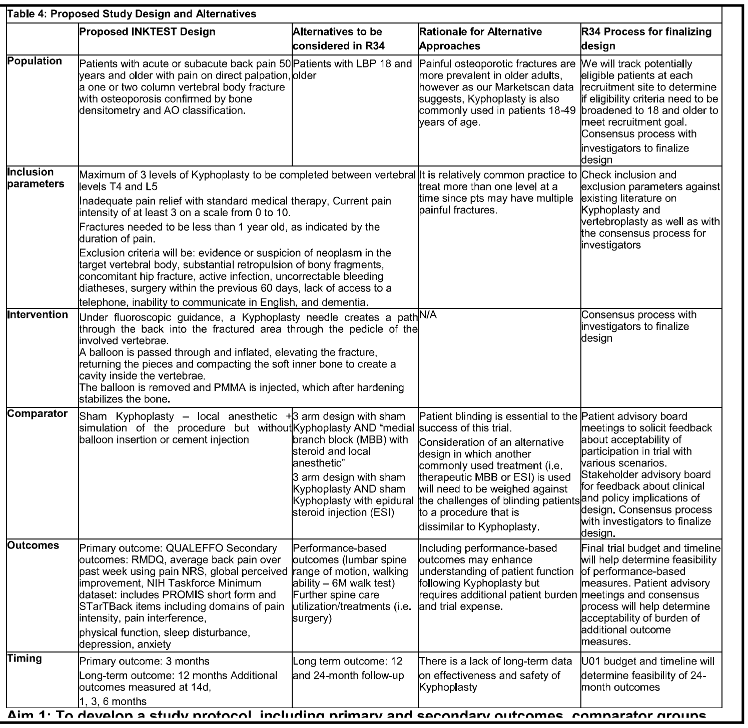


Table 4: Proposed Study Design and Alternatives

	Proposed INKTEST Design	Alternatives to be considered in R34	Rationale for Alternative Approaches	R34 Process for finalizing design
Population	Patients with acute or subacute back pain 50 years and older with pain on direct palpation, a one or two column vertebral body fracture with osteoporosis confirmed by bone densitometry and AO classification.	Patients with LBP 18 and older	Painful osteoporotic fractures are more prevalent in older adults, however as our Marketscan data suggests, Kyphoplasty is also commonly used in patients 18-49 years of age.	We will track potentially eligible patients at each recruitment site to determine if eligibility criteria need to be broadened to 18 and older to meet recruitment goal. Consensus process with investigators to finalize design
Inclusion parameters	Maximum of 3 levels of Kyphoplasty to be completed between vertebral levels T4 and L5 Inadequate pain relief with standard medical therapy, Current pain intensity of at least 3 on a scale from 0 to 10. Fractures needed to be less than 1 year old, as indicated by the duration of pain. Exclusion criteria will be: evidence or suspicion of neoplasm in the target vertebral body, substantial retropulsion of bony fragments, concomitant hip fracture, active infection, uncorrectable bleeding diatheses, surgery within the previous 60 days, lack of access to a telephone, inability to communicate in English, and dementia.		It is relatively common practice to treat more than one level at a time since pts may have multiple painful fractures.	Check inclusion and exclusion parameters against existing literature on Kyphoplasty and vertebroplasty as well as with the consensus process for investigators
Intervention	Under fluoroscopic guidance, a Kyphoplasty needle creates a path through the back into the fractured area through the pedicle of the involved vertebrae. A balloon is passed through and inflated, elevating the fracture, returning the pieces and compacting the soft inner bone to create a cavity inside the vertebrae. The balloon is removed and PMMA is injected, which after hardening stabilizes the bone.		N/A	Consensus process with investigators to finalize design
Comparator	Sham Kyphoplasty – local anesthetic + simulation of the procedure but without balloon insertion or cement injection	3 arm design with sham Kyphoplasty AND "medial branch block (MBB) with steroid and local anesthetic" 3 arm design with sham Kyphoplasty AND sham Kyphoplasty with epidural steroid injection (ESI)	Patient blinding is essential to the success of this trial. Consideration of an alternative design in which another commonly used treatment (i.e. therapeutic MBB or ESI) is used will need to be weighed against the challenges of blinding patients to a procedure that is dissimilar to Kyphoplasty.	Patient advisory board meetings to solicit feedback about acceptability of participation in trial with various scenarios. Stakeholder advisory board for feedback about clinical and policy implications of design. Consensus process with investigators to finalize design.
Outcomes	Primary outcome: QUALEFFO Secondary outcomes: RMDQ, average back pain over past week using pain NRS, global perceived improvement, NIH Taskforce Minimum dataset: includes PROMIS short form and STarTBack items including domains of pain intensity, pain interference, physical function, sleep disturbance, depression, anxiety	Performance-based outcomes (lumbar spine range of motion, walking ability – 6M walk test) Further spine care utilization/treatments (i.e. surgery)	Including performance-based outcomes may enhance understanding of patient function following Kyphoplasty but requires additional patient burden and trial expense.	Final trial budget and timeline will help determine feasibility of performance-based measures. Patient advisory meetings and consensus process will help determine acceptability of burden of additional outcome measures.
Timing	Primary outcome: 3 months Long-term outcome: 12 months Additional outcomes measured at 14d, 1, 3, 6 months	Long term outcome: 12 and 24-month follow-up	There is a lack of long-term data on effectiveness and safety of Kyphoplasty	U01 budget and timeline will determine feasibility of 24-month outcomes

Aim 1: To develop a study protocol, including primary and secondary outcomes, comparator groups

8

Diagnostic Tests: Even Less Simple

Hard to demonstrate the impact of a diagnostic test on patient outcome, or…
"Many a slip twixt cup and lip"

9

Can't Show a Link Between a Diagnostic Test and Patient Outcomes? Who's To Blame?

blame the test (it really isn't useful)

10

Can’t Show a Link Between a Diagnostic Test and Patient Outcomes? Who’s To Blame?

blame the test (it really isn’t useful)

blame the radiologist (useful test but bad interpretation)

11

Can’t Show a Link Between a Diagnostic Test and Patient Outcomes? Who’s To Blame?

- blame the test (it really isn’t useful)
- blame the radiologist (useful test but bad interpretation)

blame the clinician (bad use of helpful info)

12

Can’t Show a Link Between a Diagnostic Test and Patient Outcomes? Who’s To Blame?

- blame the test (it really isn’t useful)
- blame the radiologist (useful test but bad interpretation)

blame the clinician (bad use of helpful info)

blame the therapy (available Rx ineffective)

Can’t Show a Link Between a Diagnostic Test and Patient Outcomes? Who’s To Blame?

- blame the test (it really isn’t useful)
- blame the radiologist (useful test but bad interpretation)
- blame the clinician (bad use of helpful info)

blame the therapy (available Rx ineffective)

blame the patient (non-compliance)

Can’t Show a Link Between a Diagnostic Test and Patient Outcomes? Who’s To Blame?

- blame the test (it really isn’t useful)
- blame the radiologist (useful test but bad interpretation)
- blame the clinician (bad use of helpful info)
- blame the therapy (available Rx ineffective)

blame the patient (non-compliance)

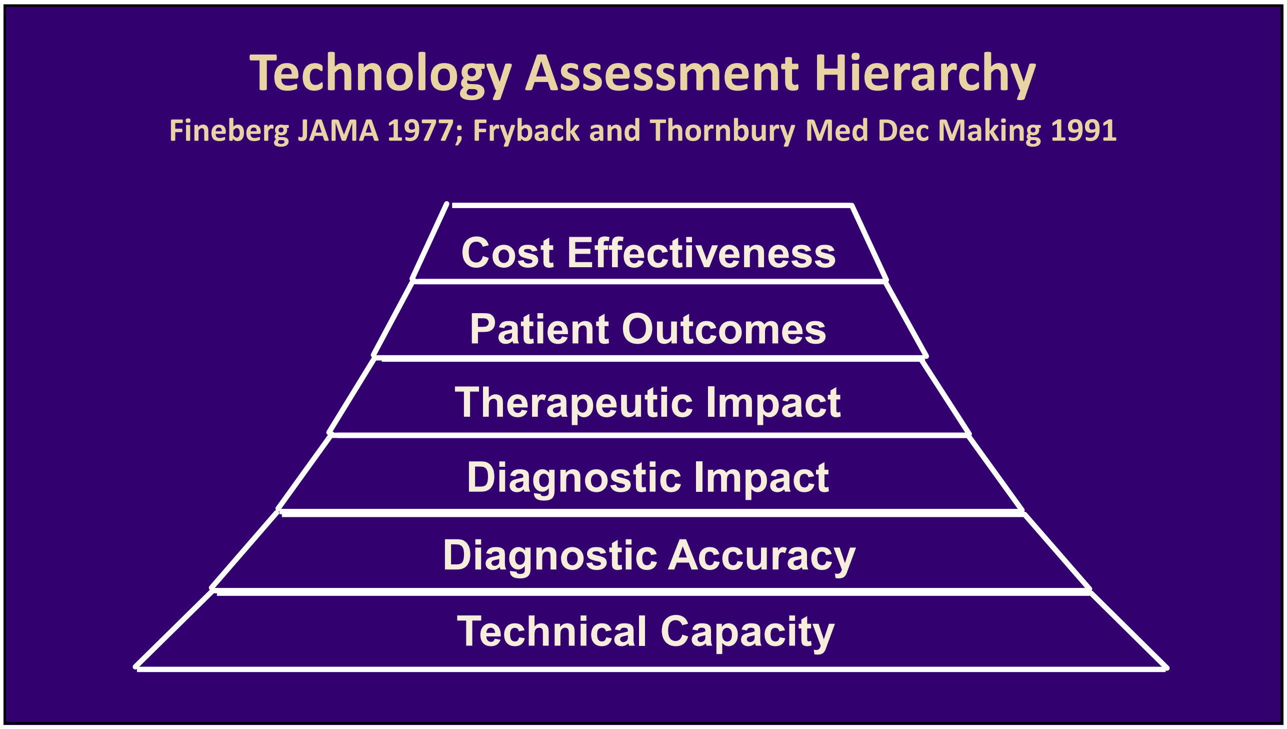
blame the system (lack of access)

15

Take Home Points

- There are unique challenges to assessing diagnostic imaging
- The basics of diagnostic imaging test assessment:
  - Tech assessment hierarchy
  - Accuracy
  - Bias

16



Technology Assessment Hierarchy

Fineberg JAMA 1977; Fryback and Thornbury Med Dec Making 1991

Cost Effectiveness

Patient Outcomes

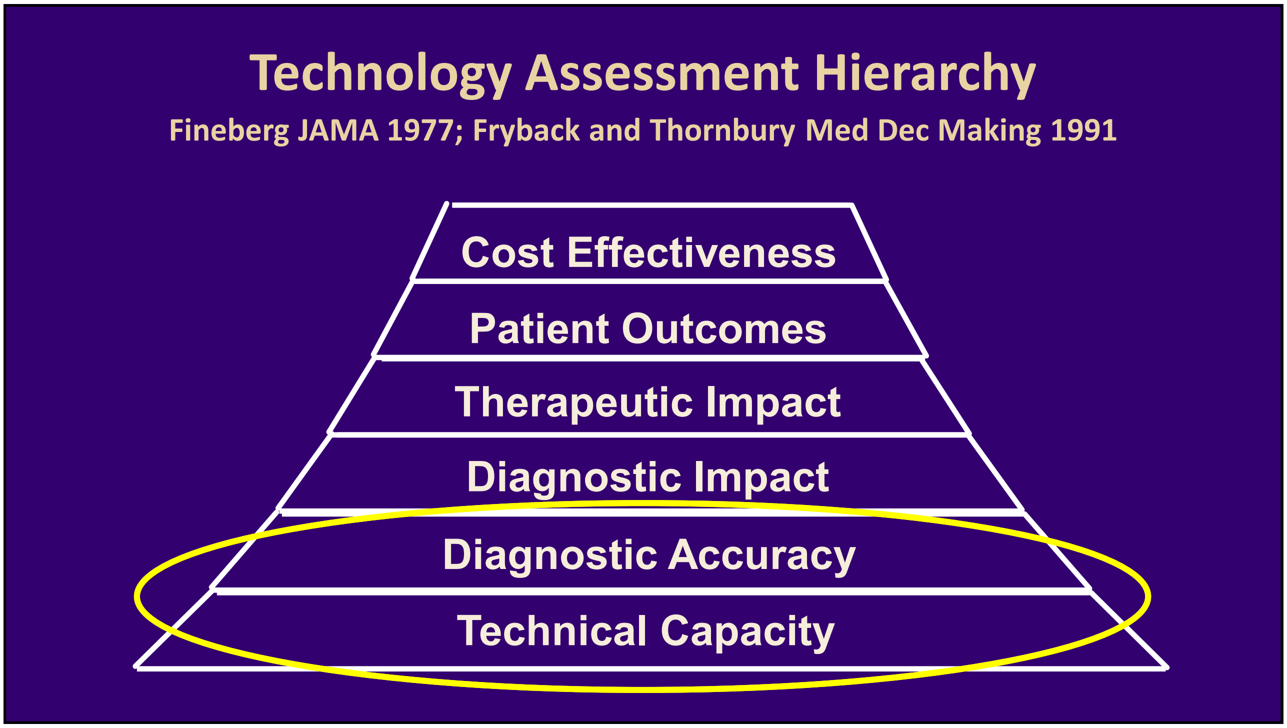
Therapeutic Impact

Diagnostic Impact

Diagnostic Accuracy

Technical Capacity

17



Technology Assessment Hierarchy

Fineberg JAMA 1977; Fryback and Thornbury Med Dec Making 1991

Cost Effectiveness

Patient Outcomes

Therapeutic Impact

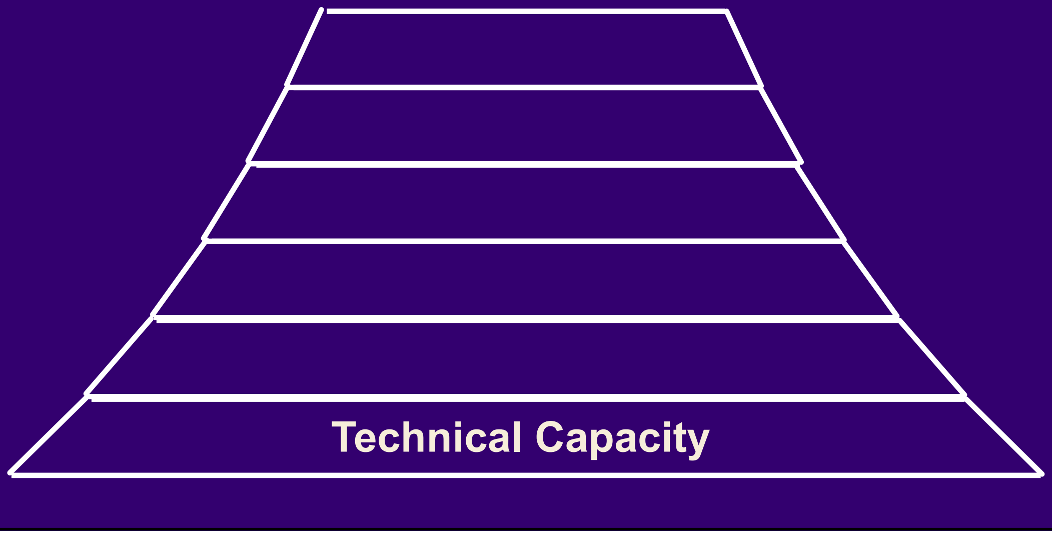
Diagnostic Impact

Diagnostic Accuracy

Technical Capacity

18

Technology Assessment Pyramid



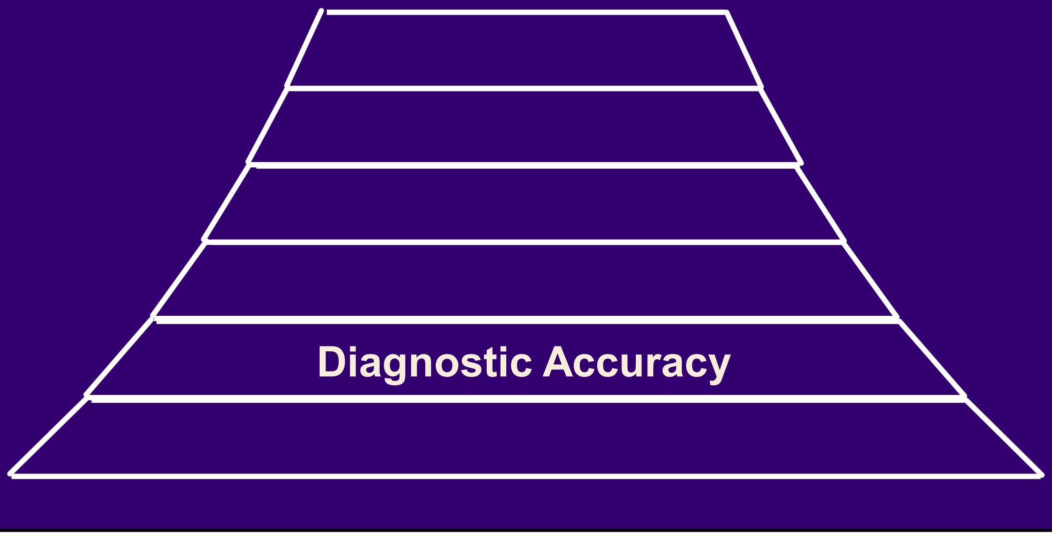
19

Technical Capacity

- laboratory phase
- standardize technical parameters of test
- phantom studies
- reliability

20

Technology Assessment Pyramid



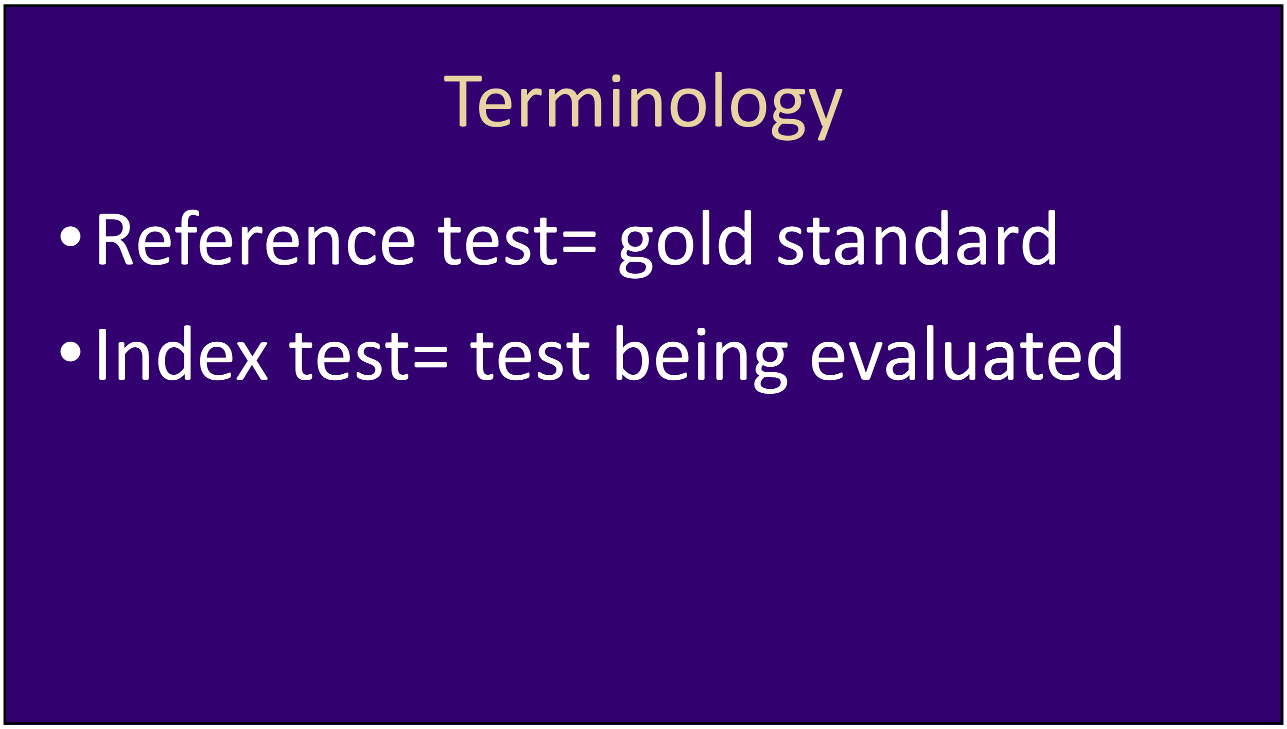
Diagnostic Accuracy

21

Diagnostic Accuracy

- sensitivity
- specificity
- predictive value
- likelihood ratios

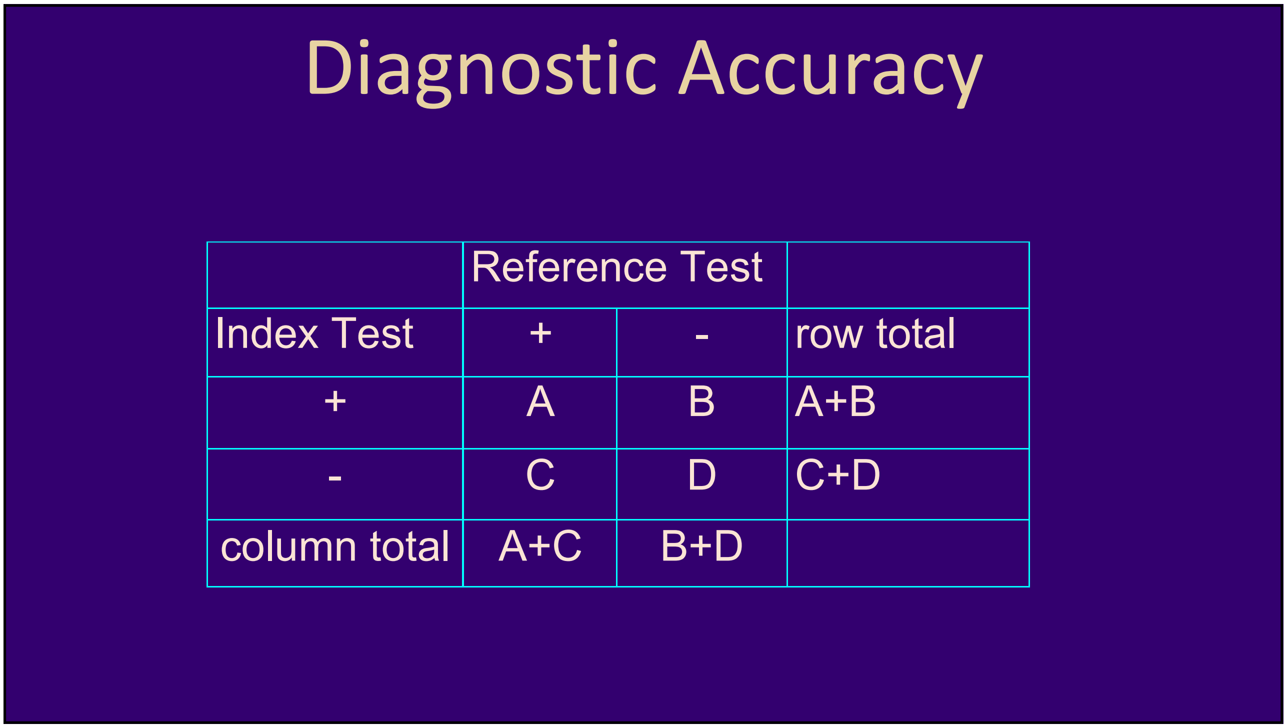
22



Terminology

- Reference test= gold standard
- Index test= test being evaluated

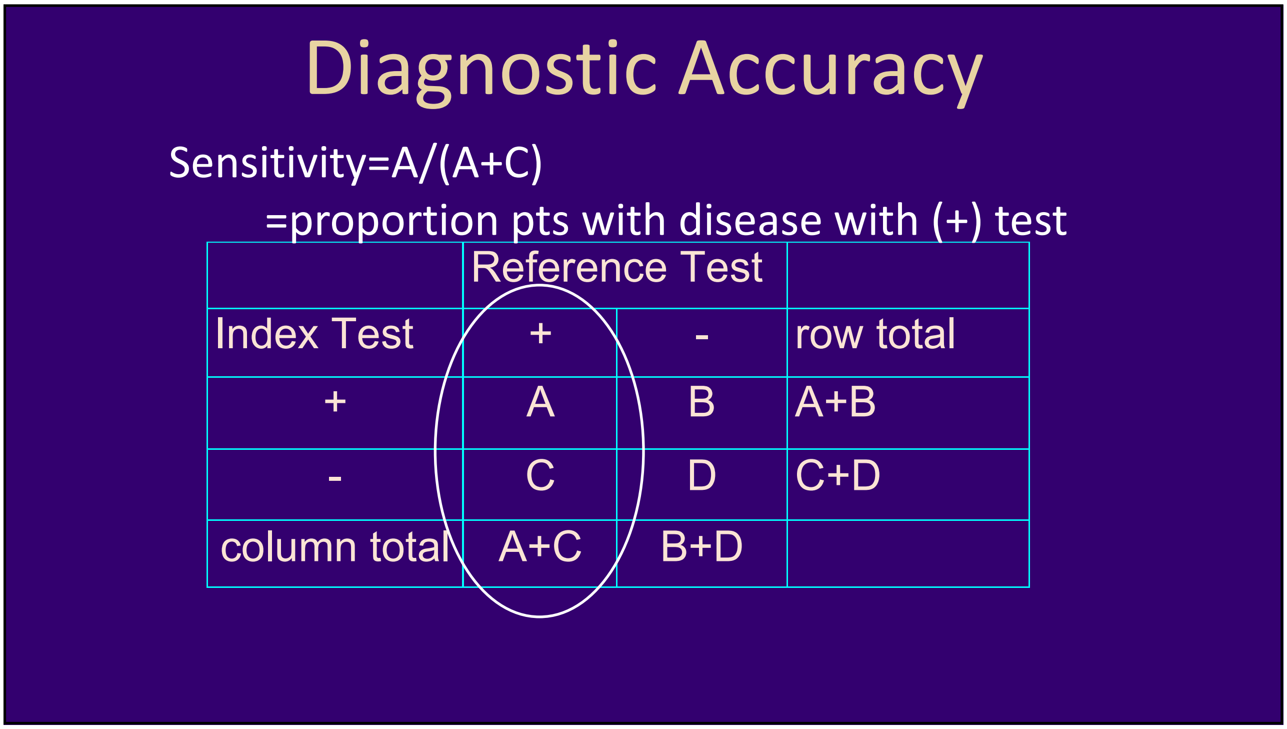
23



Diagnostic Accuracy

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

24



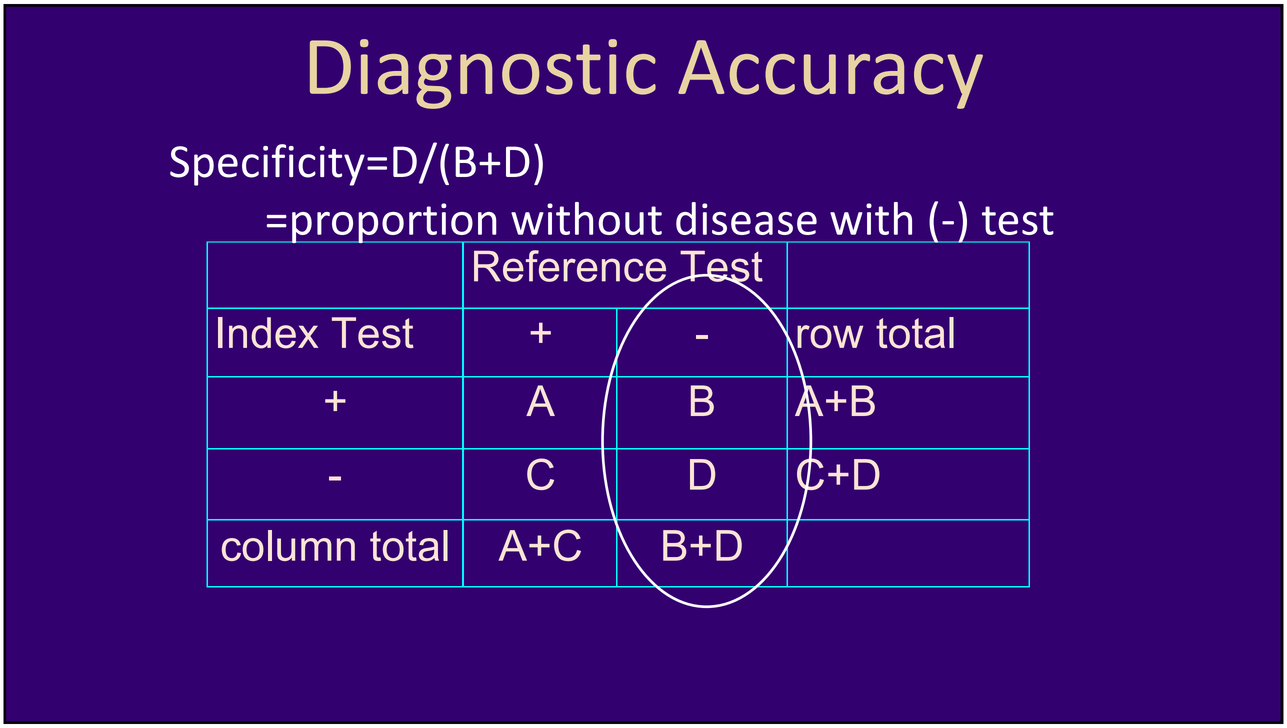
Diagnostic Accuracy

Sensitivity=A/(A+C)

=proportion pts with disease with (+) test

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

25



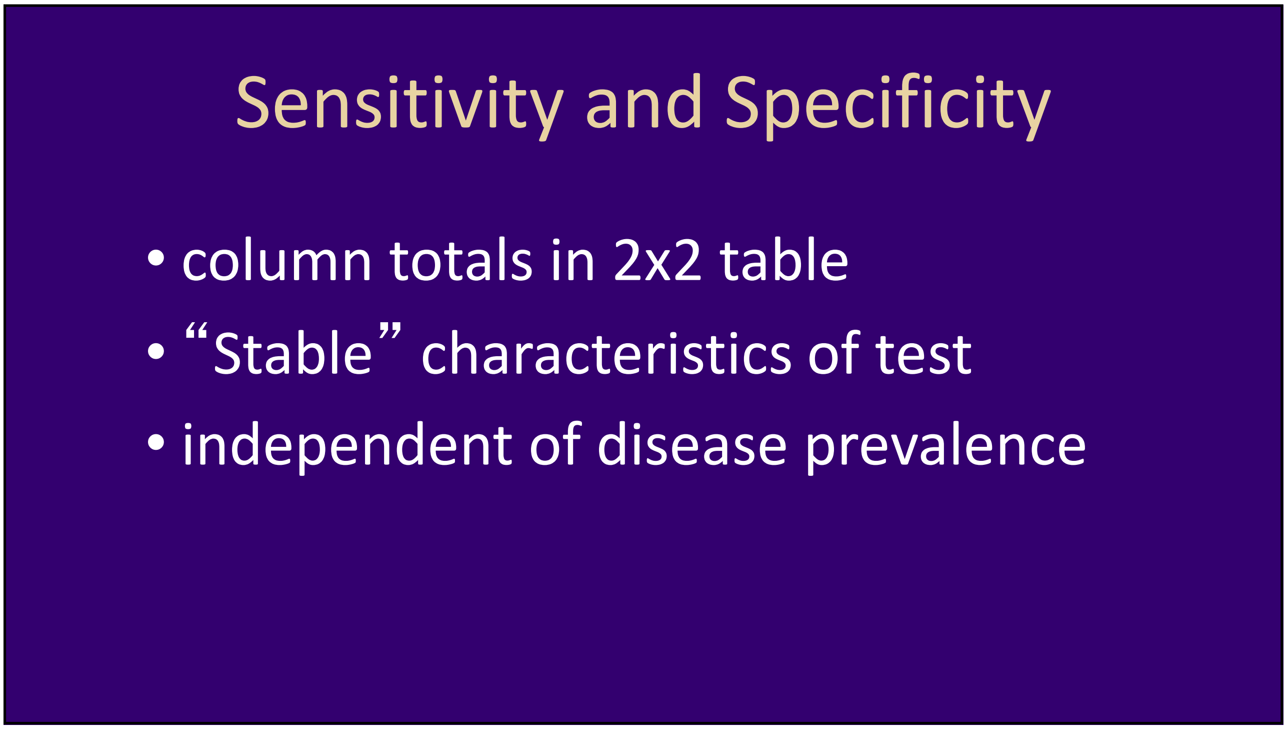
Diagnostic Accuracy

Specificity=D/(B+D)

=proportion without disease with (-) test

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

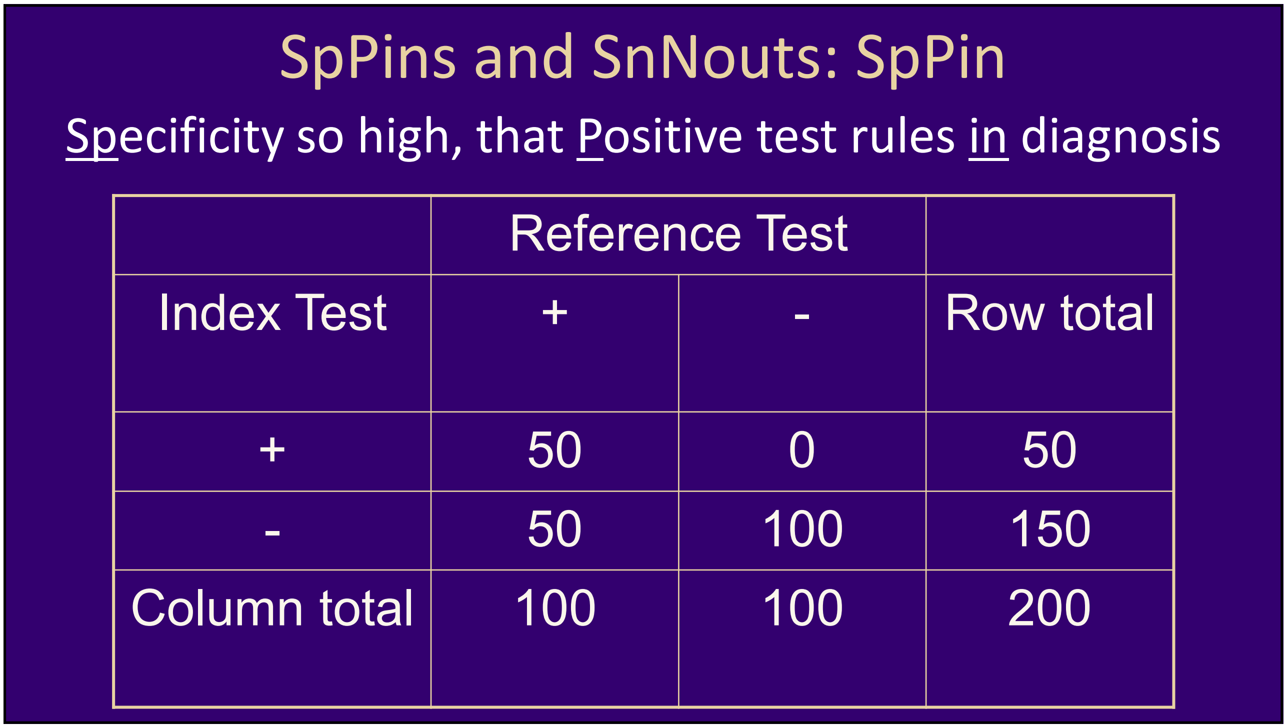
26



Sensitivity and Specificity

- column totals in 2x2 table
- “Stable” characteristics of test
- independent of disease prevalence

27

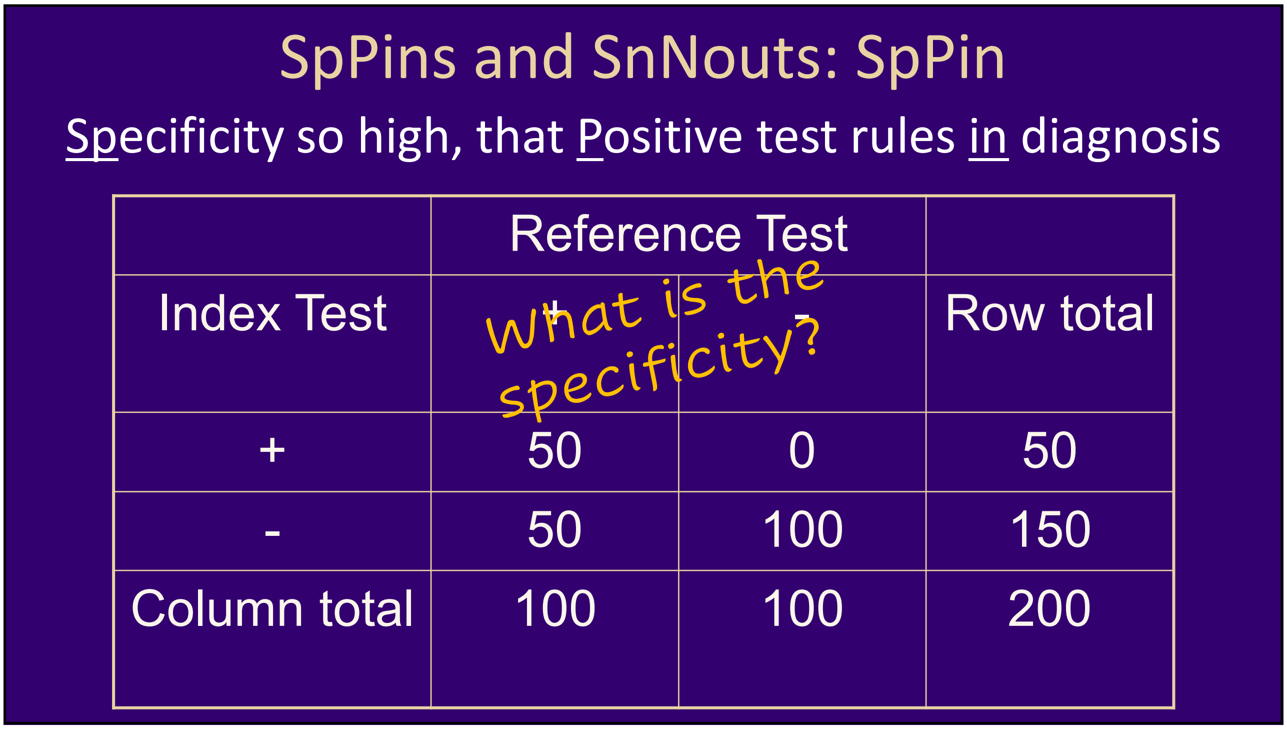


SpPins and SnNouts: SpPin

Specificity so high, that Positive test rules in diagnosis

	Reference Test		
Index Test	+	-	Row total
+	50	0	50
-	50	100	150
Column total	100	100	200

28



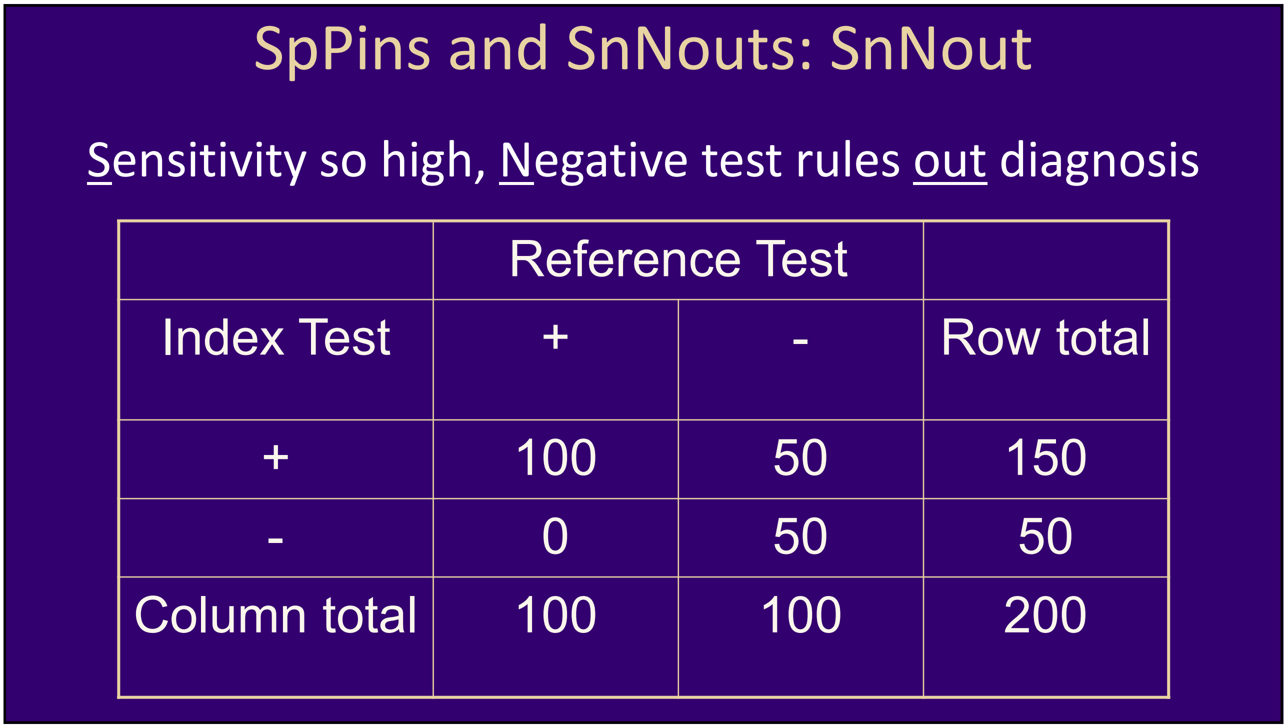
SpPins and SnNouts: SpPin

Specificity so high, that Positive test rules in diagnosis

	Reference Test		
Index Test	+	-	Row total
+	50	0	50
-	50	100	150
Column total	100	100	200

What is the specificity?

29

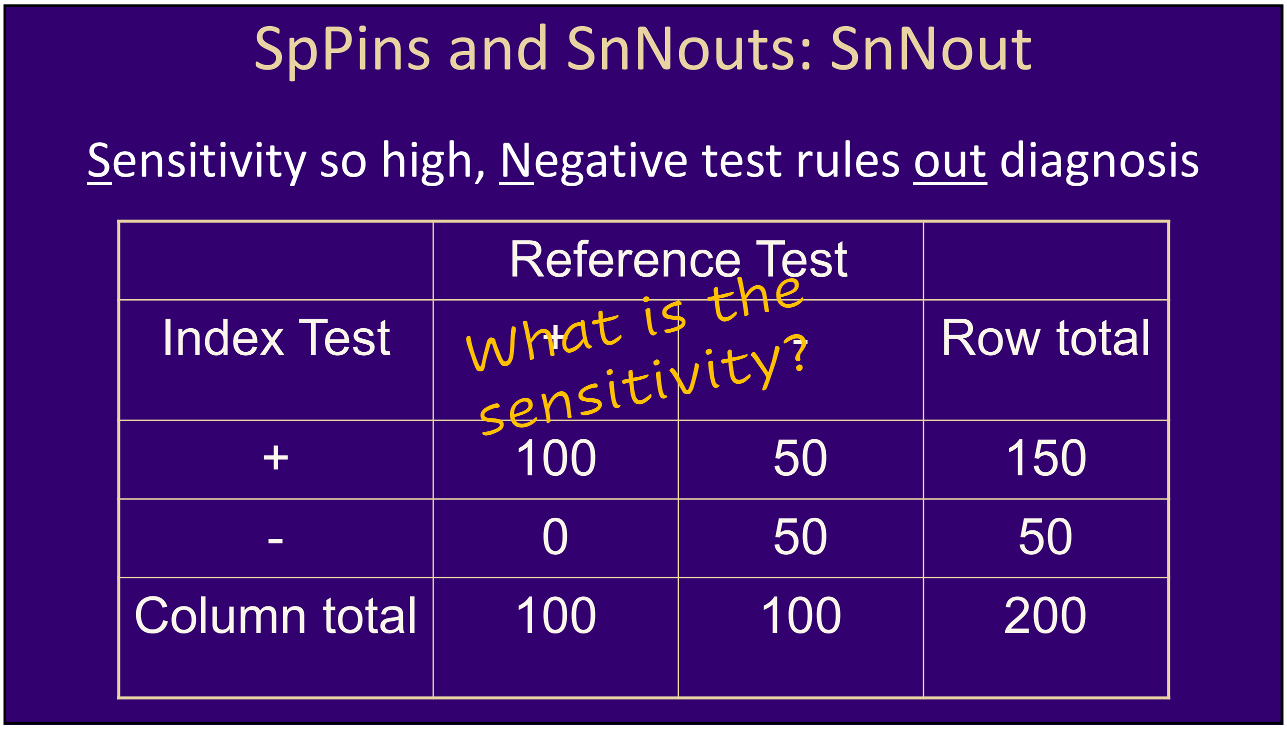


SpPins and SnNouts: SnNout

Sensitivity so high, Negative test rules out diagnosis

	Reference Test		
Index Test	+	-	Row total
+	100	50	150
-	0	50	50
Column total	100	100	200

30



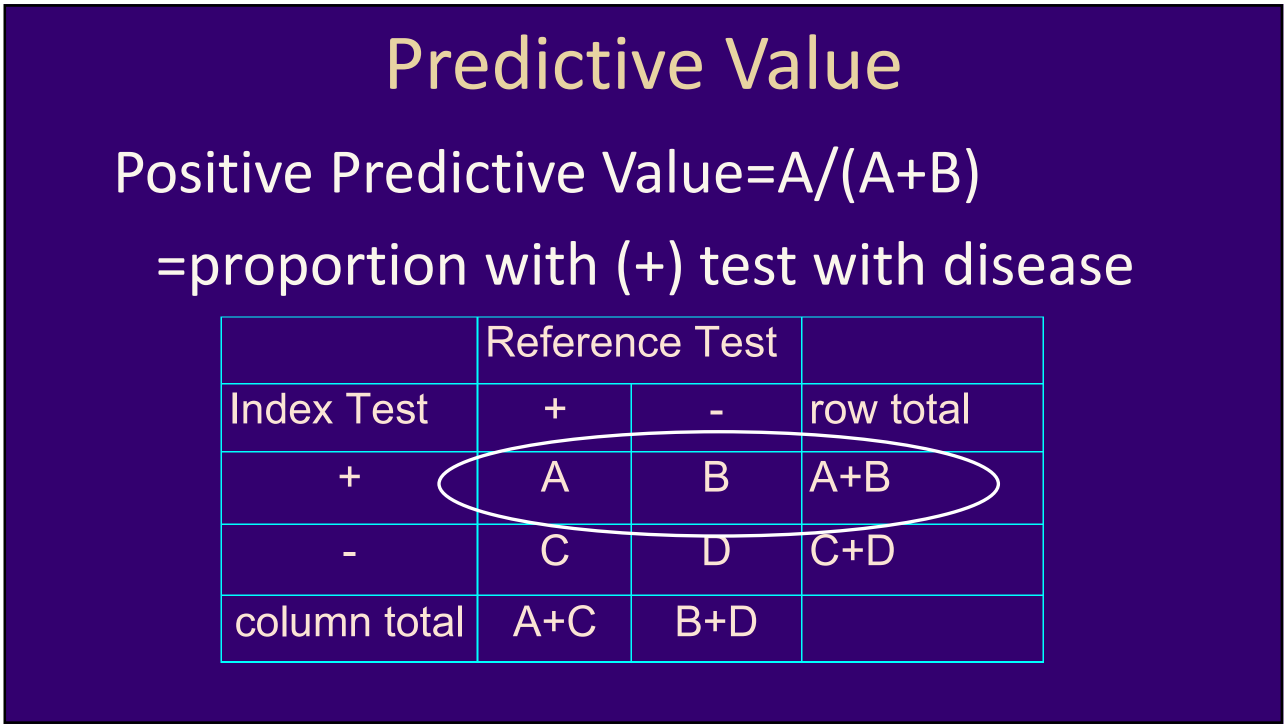
SpPins and SnNouts: SnNout

Sensitivity so high, Negative test rules out diagnosis

	Reference Test		
Index Test	+	-	Row total
+	100	50	150
-	0	50	50
Column total	100	100	200

What is the sensitivity?

31



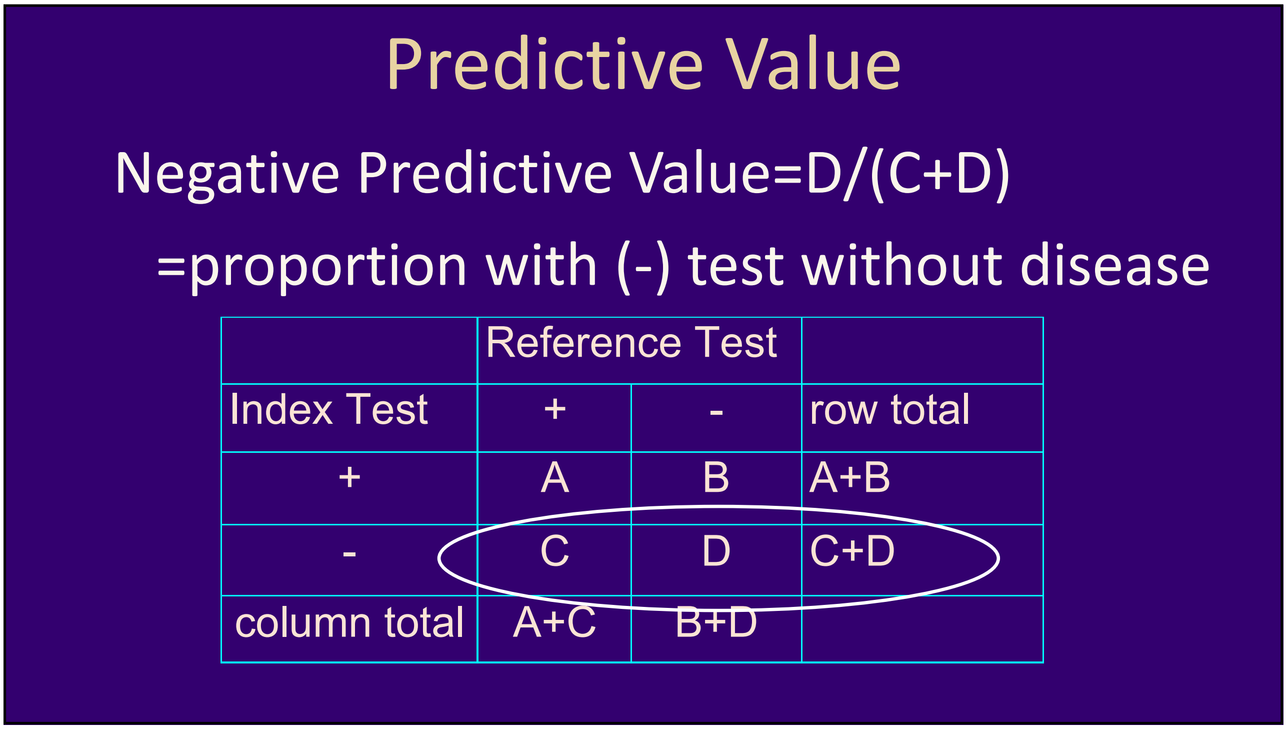
Predictive Value

Positive Predictive Value=A/(A+B)

=proportion with (+) test with disease

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

32



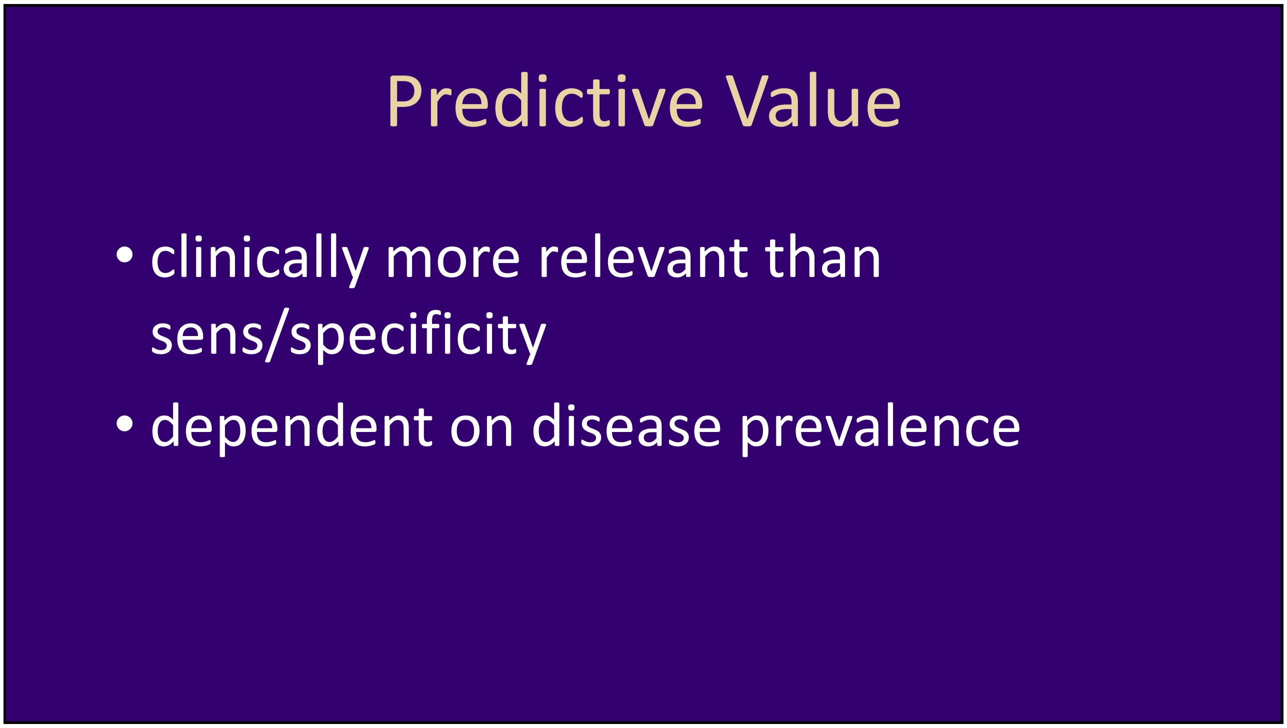
Predictive Value

Negative Predictive Value=D/(C+D)

=proportion with (-) test without disease

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

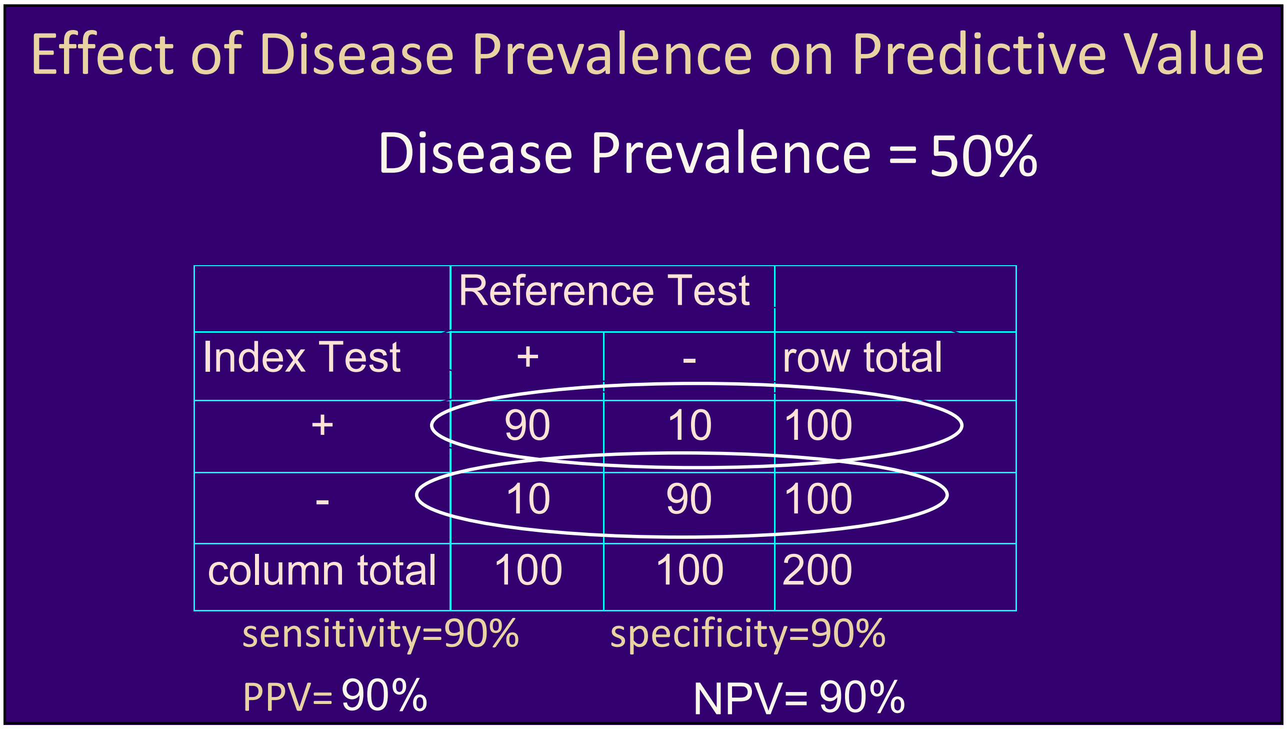
33



Predictive Value

- clinically more relevant than sens/specificity
- dependent on disease prevalence

34



Effect of Disease Prevalence on Predictive Value

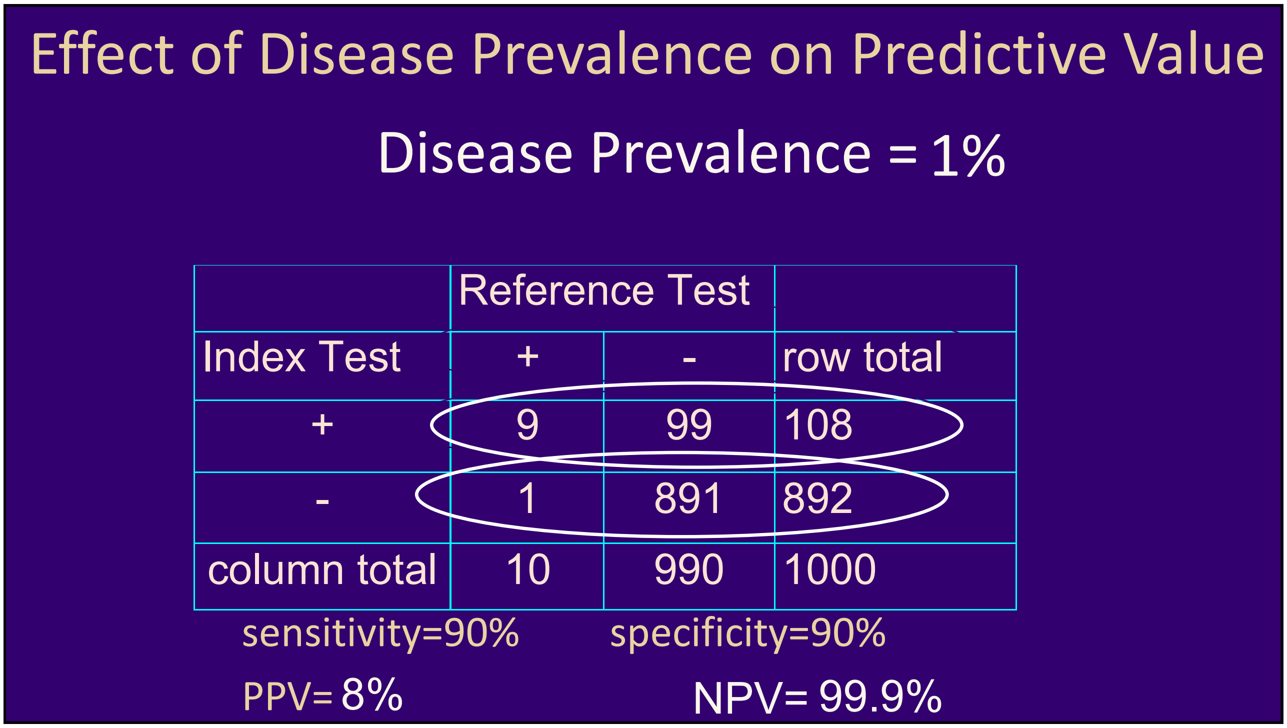
Disease Prevalence = 50%

	Reference Test		
Index Test	+	-	row total
+	90	10	100
-	10	90	100
column total	100	100	200

sensitivity=90% specificity=90%

PPV= 90% NPV= 90%

35



Effect of Disease Prevalence on Predictive Value

Disease Prevalence = 1%

	Reference Test		
Index Test	+	-	row total
+	9	99	108
-	1	891	892
column total	10	990	1000

sensitivity=90% specificity=90%

PPV= 8% NPV= 99.9%

36

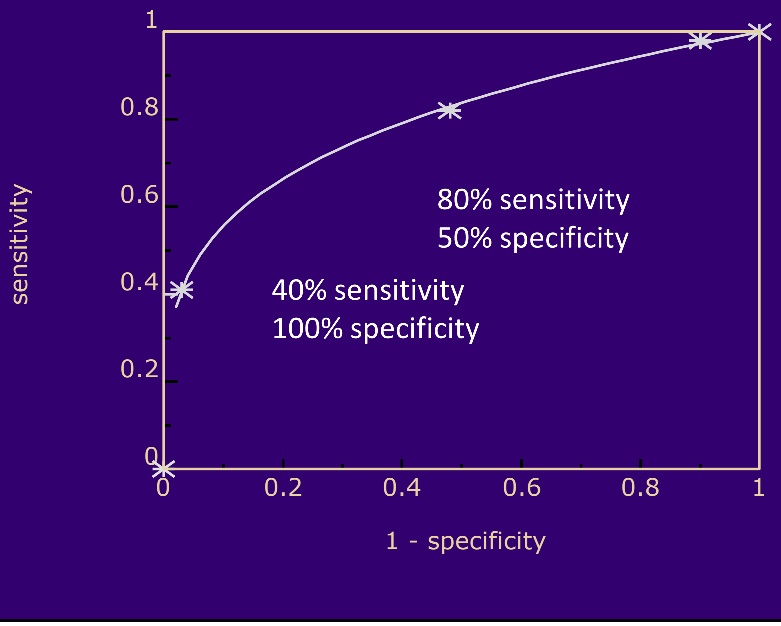
Likelihood Ratio (positive)

- Prob. of +test in those with dx divided by prob of +test in those without the dx
- [A/(A+C)]/[B/(B+D)]
- sensitivity/(1-specificity) (look familiar?)

	Reference Test		
Index Test	+	-	row total
+	A	B	A+B
-	C	D	C+D
column total	A+C	B+D	

37

Receiver Operator Characteristics (ROC)

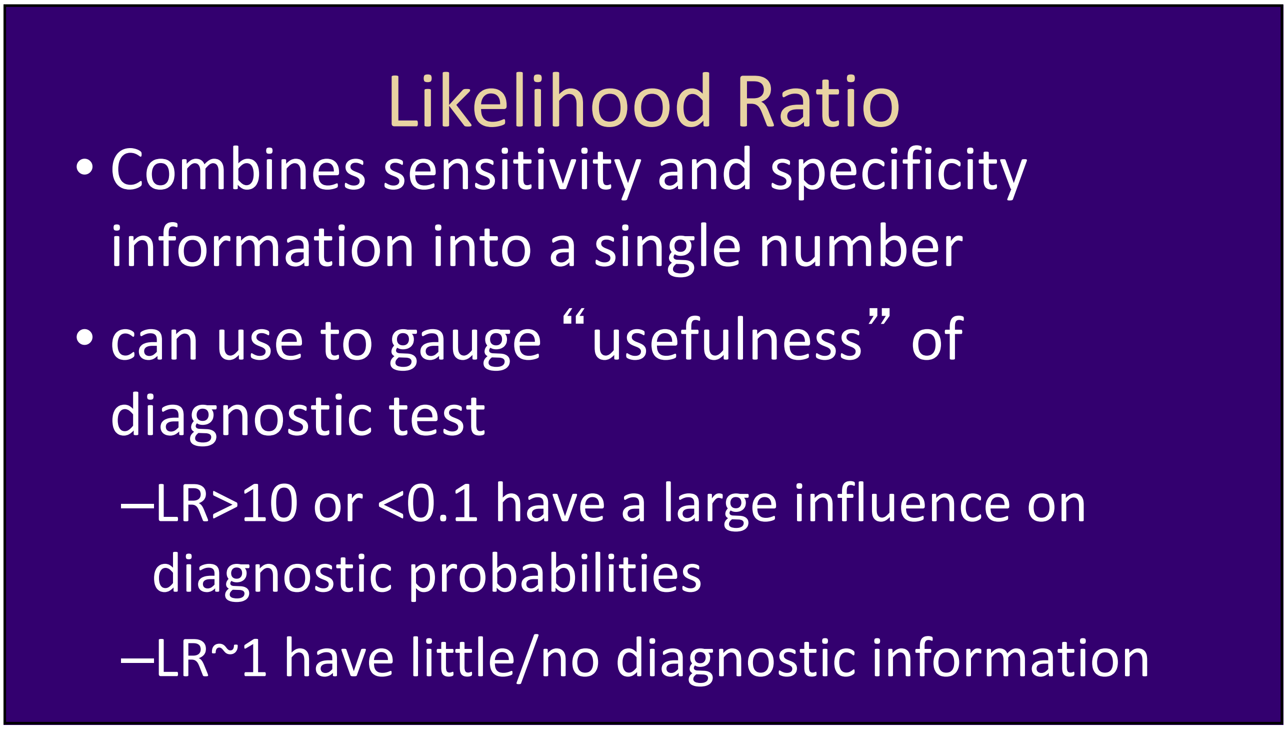


80% sensitivity
50% specificity

40% sensitivity
100% specificity

- Stay tuned- more on this later in the week

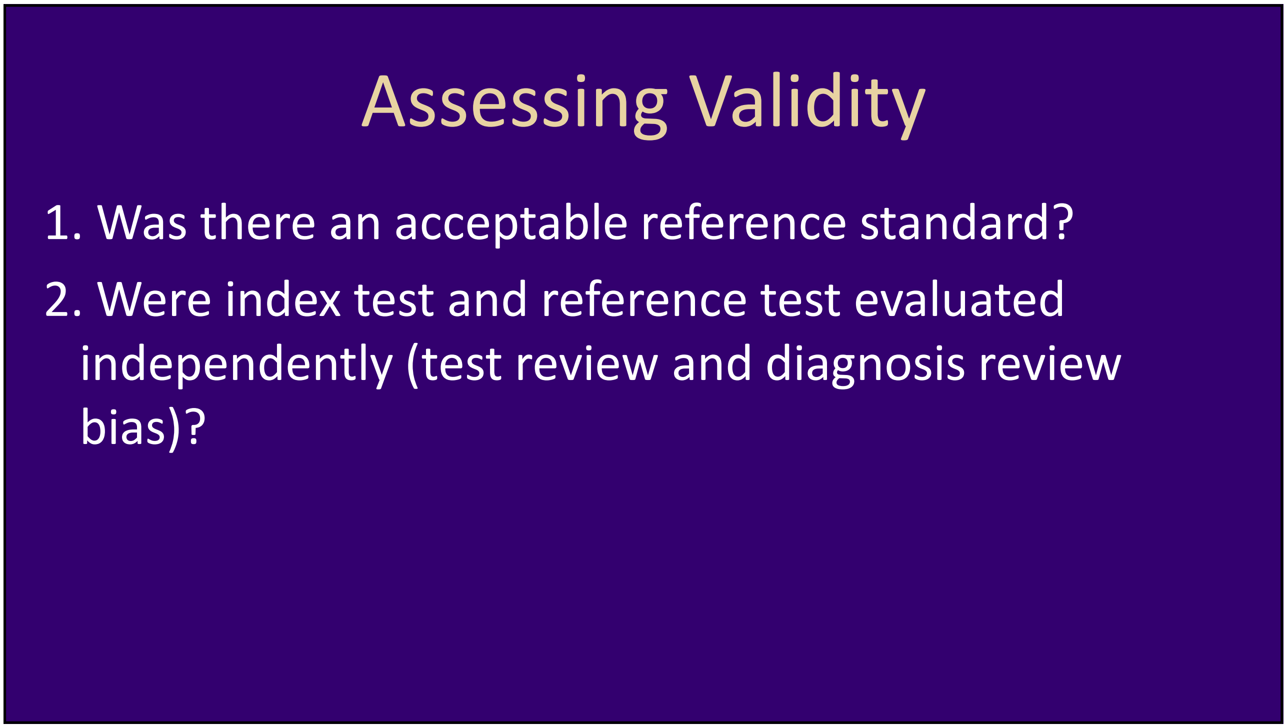
38



Likelihood Ratio

- Combines sensitivity and specificity information into a single number
- can use to gauge “usefulness” of diagnostic test
  - –LR>10 or <0.1 have a large influence on diagnostic probabilities
  - –LR~1 have little/no diagnostic information

39



Assessing Validity

1. Was there an acceptable reference standard?
2. Were index test and reference test evaluated independently (test review and diagnosis review bias)?

40

Take Home Points

- There are unique challenges to assessing diagnostic imaging
- The basics of diagnostic imaging test assessment:
  - Tech assessment hierarchy
  - Accuracy
  - Bias

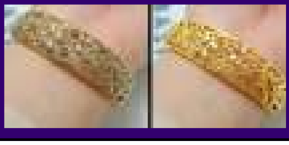
41

Test Review Bias

- Index test reviewed knowing results of reference test

Diagnosis Review Bias

- Reference test reviewed knowing results of index test →



Tarnished gold standard

42

Assessing Validity

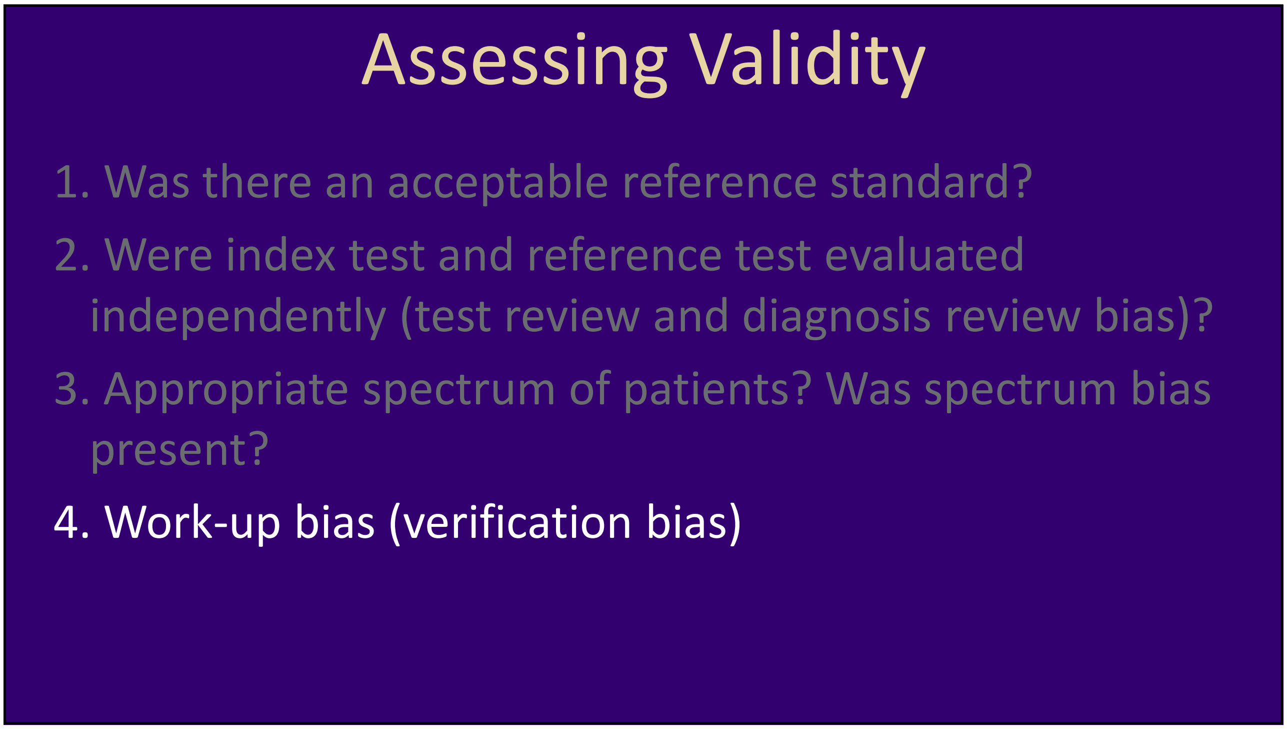
1. Was there an acceptable reference standard?
2. Were index test and reference test evaluated independently (test review and diagnosis review bias)?
3. Appropriate spectrum of patients? Was spectrum bias present?

43

Spectrum Bias

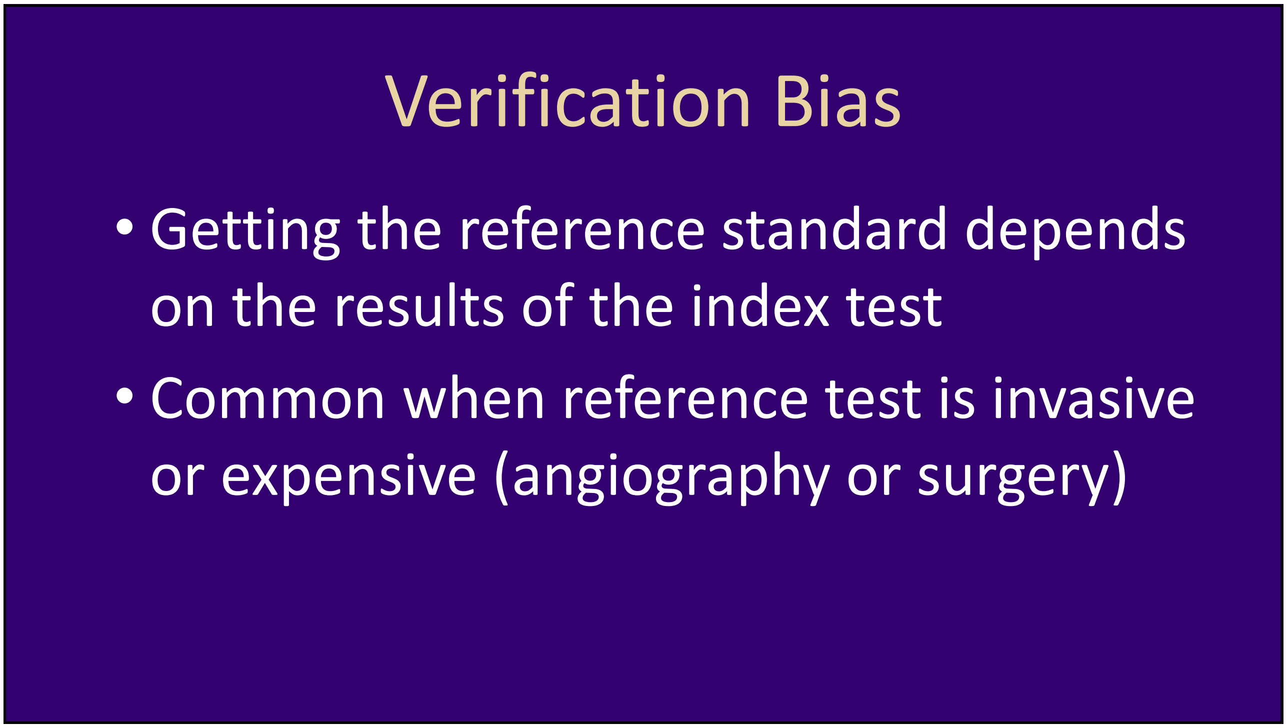
- Common sampling bias in radiology
- Compare “sickest of sick with wellest of well”
- e.g. testing the ability of tau imaging to discriminate between healthy med student volunteers and elderly cohort with severe dementia

44



Assessing Validity

1. Was there an acceptable reference standard?
2. Were index test and reference test evaluated independently (test review and diagnosis review bias)?
3. Appropriate spectrum of patients? Was spectrum bias present?
4. Work-up bias (verification bias)



Verification Bias

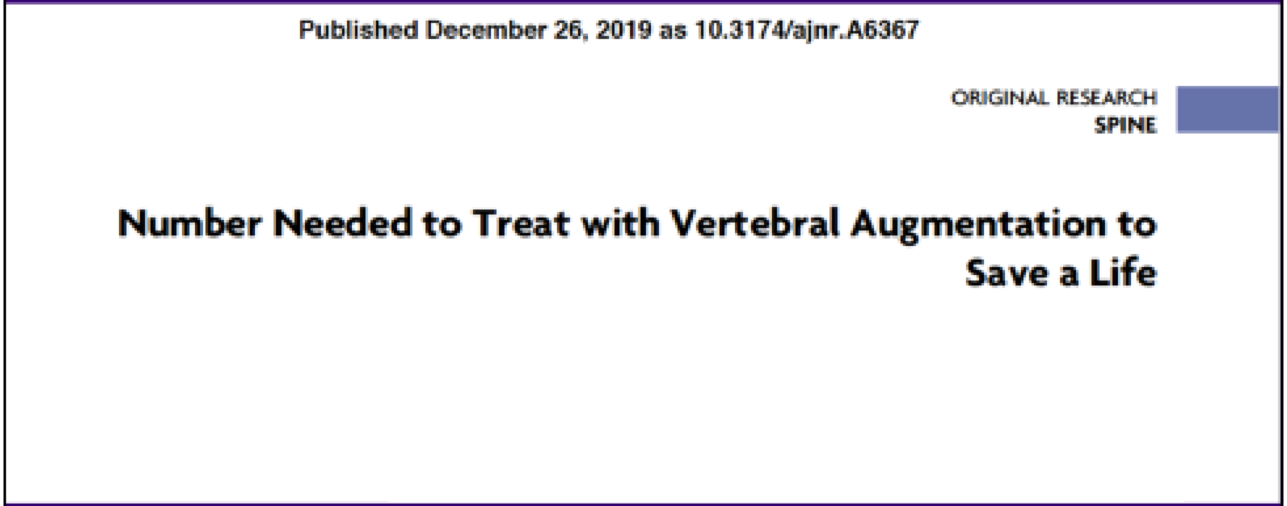
- Getting the reference standard depends on the results of the index test
- Common when reference test is invasive or expensive (angiography or surgery)

Immortal Time Bias

- Type of survivor/ascertainment bias
- Cool name!

47

Example of Immortal Time Bias



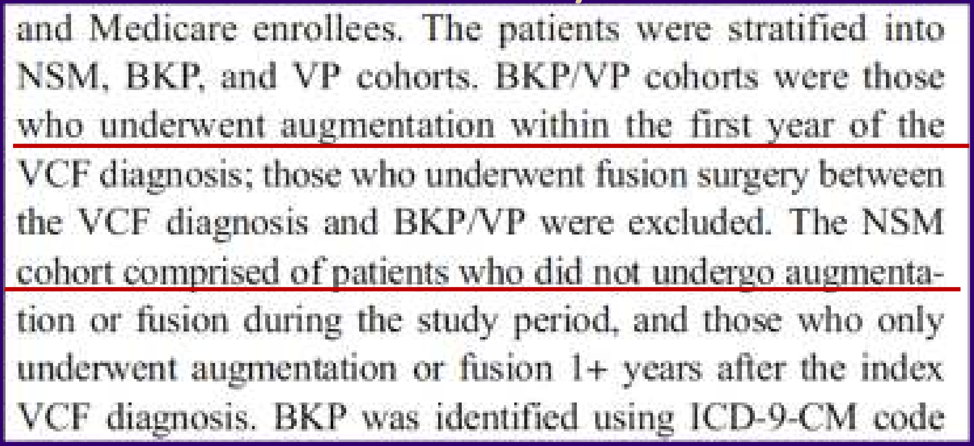
Published December 26, 2019 as 10.3174/ajnr.A6367

ORIGINAL RESEARCH
SPINE

Number Needed to Treat with Vertebral Augmentation to Save a Life

48

Balloon Kyphoplasty (BKP)/ Vertebroplasty (VP) Study



and Medicare enrollees. The patients were stratified into NSM, BKP, and VP cohorts. BKP/VP cohorts were those who underwent augmentation within the first year of the VCF diagnosis; those who underwent fusion surgery between the VCF diagnosis and BKP/VP were excluded. The NSM cohort comprised of patients who did not undergo augmentation or fusion during the study period, and those who only underwent augmentation or fusion 1+ years after the index VCF diagnosis. BKP was identified using ICD-9-CM code

AUG = (observed to have BKP/VP w/in 1 year)
NSM = (no BKP/VP w/in one year)

49

Immortal Time Bias

AUG = (had BKP/VP w/in 1 year)
NSM= (no BKP/VP w/in one year)

Q: What happens if you died in the first year before AUG?

A: You didn't live long enough to get AUG so you're put in NSM group.

NSM is enriched with deaths due to bias in assigning group membership!

AKA Immortal Time Bias: intervention group is "immortal" during 1st yr

50

Better Call Saul (BCS) & Immortal Time Bias

- Prequel to *Breaking Bad (BB)*
- Violent dramedy
- High risk of death

Question: Do all characters have equal chance of dying?

51

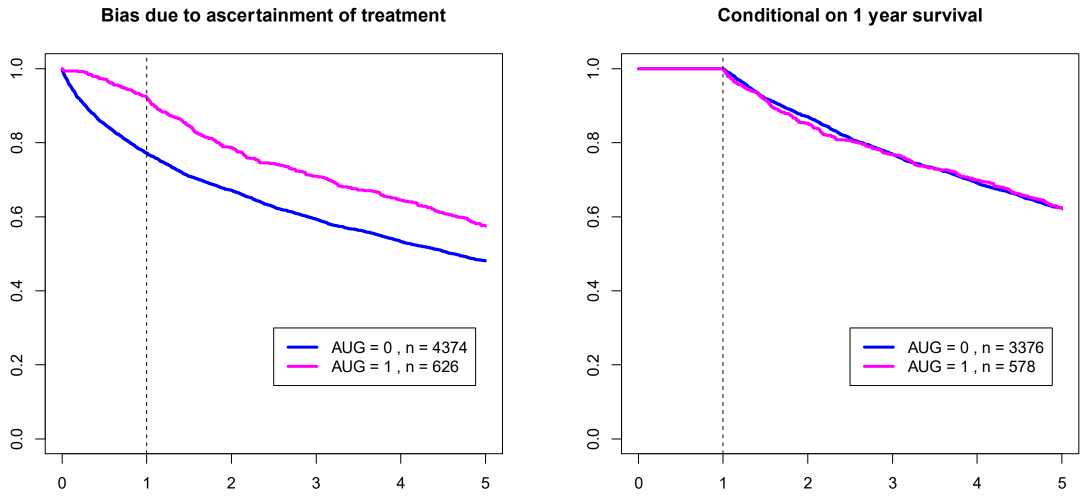
Main Characters *Better Call Saul*

Appear in *BB*→ immortal in *BCS*
(Had AUG→ immortal till AUG)

Not in *BB*→ mortal in *BCS*
(No AUG→ can die in yr1)

If they die in *BCS*, CANNOT be in *BB* (If die in YR1 before AUG, need to be in non-AUG group)

52



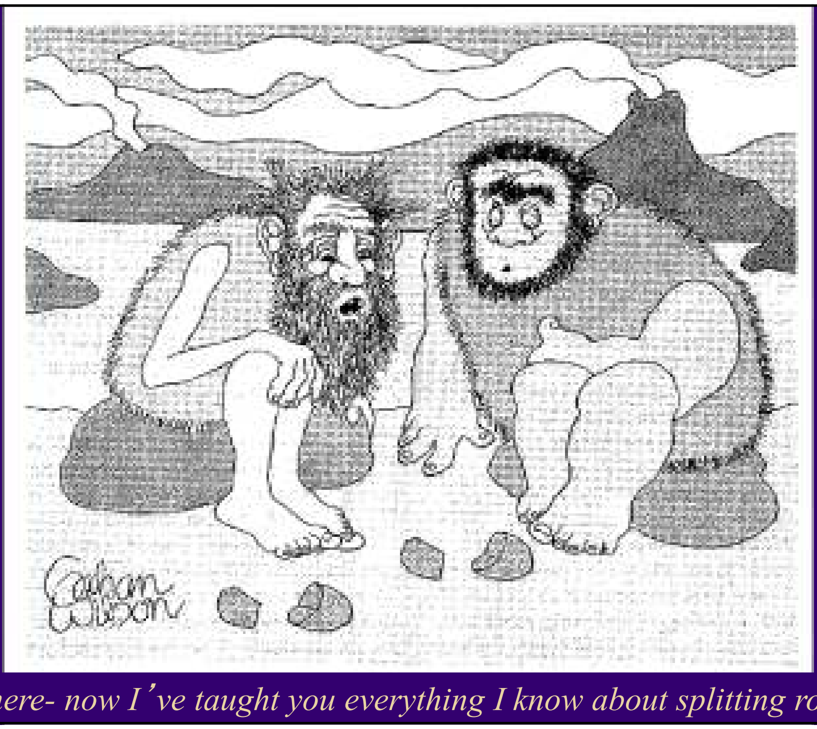
Unpublished analysis by P. Heagerty, PhD UW

53

Take Home Points

- There are unique challenges to assessing diagnostic imaging
- The basics of diagnostic imaging test assessment:
  - –Tech assessment hierarchy
  - –Accuracy
  - –Bias

54



“There- now I’ve taught you everything I know about splitting rocks.”

55

56