Bias in trials comparing paired continuous tests can cause researchers to choose the wrong screening modality

Deborah Glueck

Department of Biostatistics

Colorado School of Public Health

Denver, Colorado

United States of America

Collaborators and mentors

Todd Alonzo

John Brinton

Molly Lamb

John Lewin

Keith Muller

Colin O'Donnell

Etta Pisano

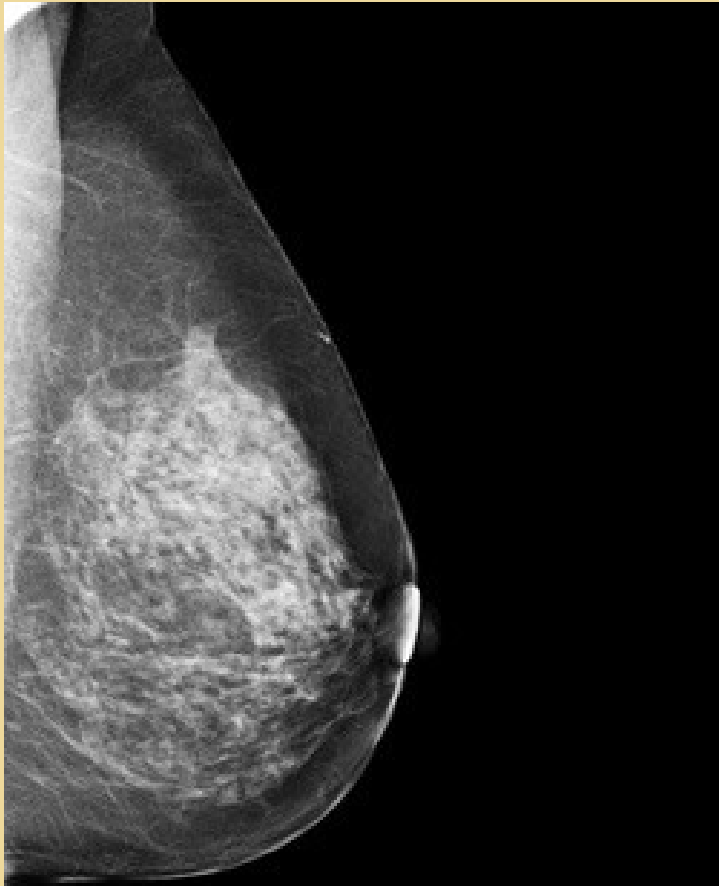Brandy Ringham

Outline

Science

- **Scientific background**
- **Scientific question**
- **ROC curves**
- ~~**Hypothesis**~~ **testing**
- **Study design**

Statistics

- **Limitations of design**
- **Is there bias?**
- **Assumptions**
- **Derivations**
- **Results**
- ~~**Conclusions**~~

Mammography

- **A screening technique for early detection of breast cancer**

- **An x-ray of the breast**

- **Screen film mammography uses film**

- **Digital mammography uses a digital detector**

http://www.hologic.com/wh/whccl.cfm

Different images produced by screen-film and digital mammography of the same breast

Full field digital mammogram
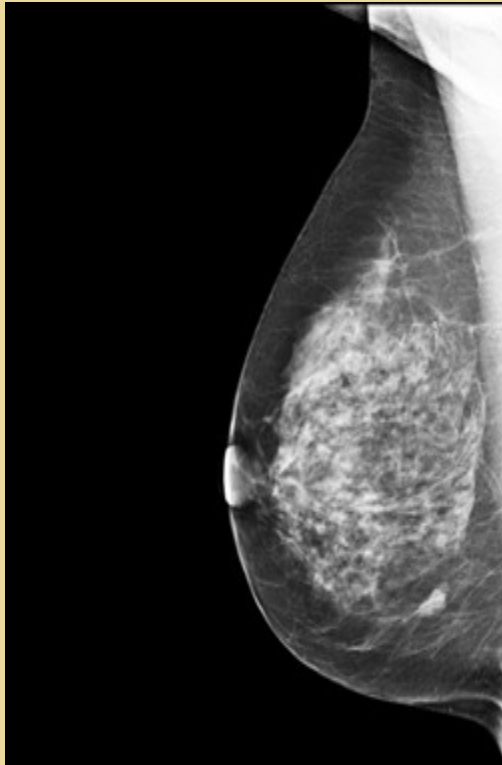
Screen-film mammogram
http://www.hologic.com/wh/cc-0507
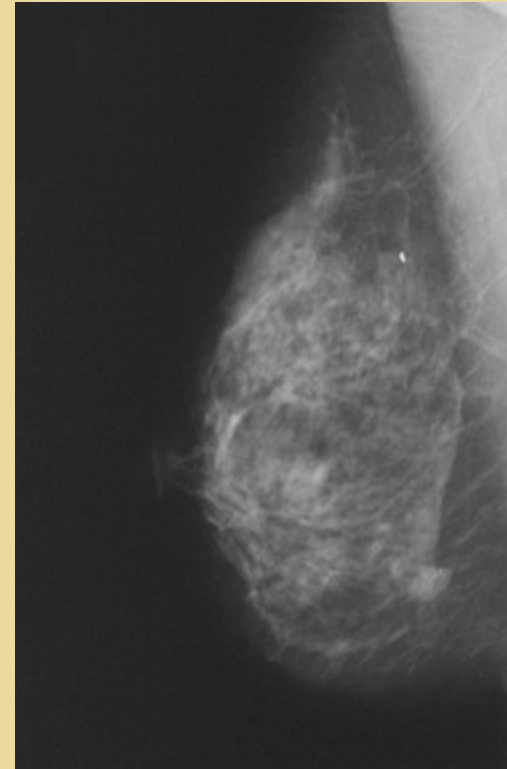
Invasive ductal carcinoma detected
on digital but not screen film mammography

Full field digital
mammogram

Screen-film
mammogram

Lewin et al.
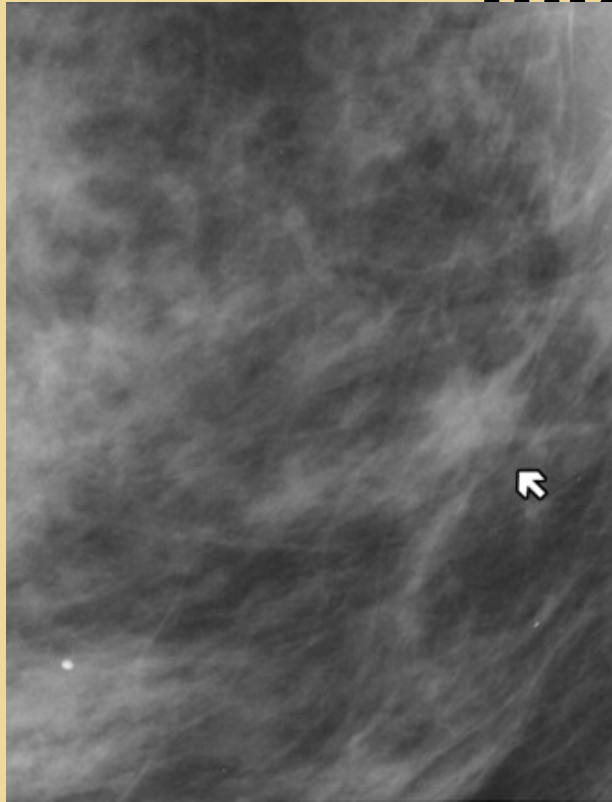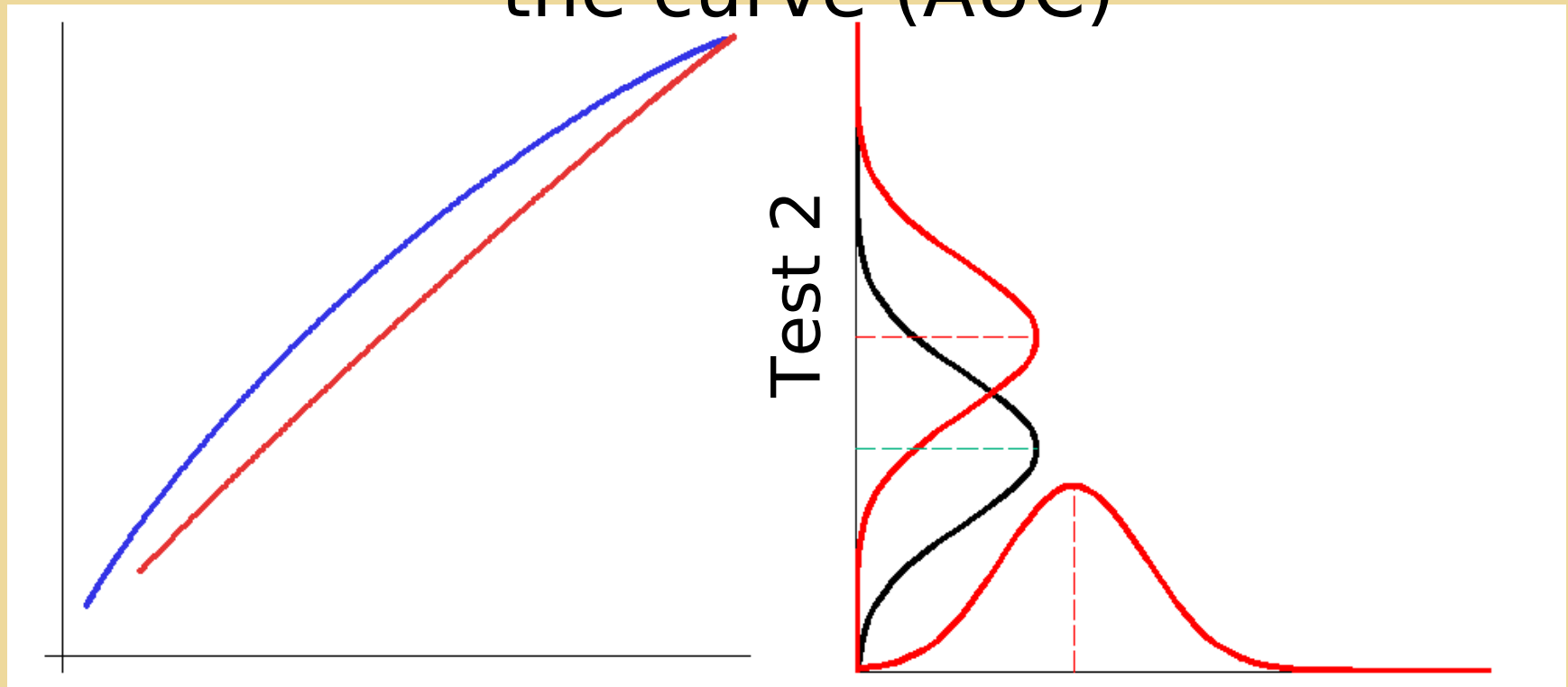
Which screening method is best?

Truth

+   **-**

Test

	+	**-**
+	**a**	**b**
-	**c**	**d**

Sensitivity = a / (a+c)

Specificity = d / (b+d)
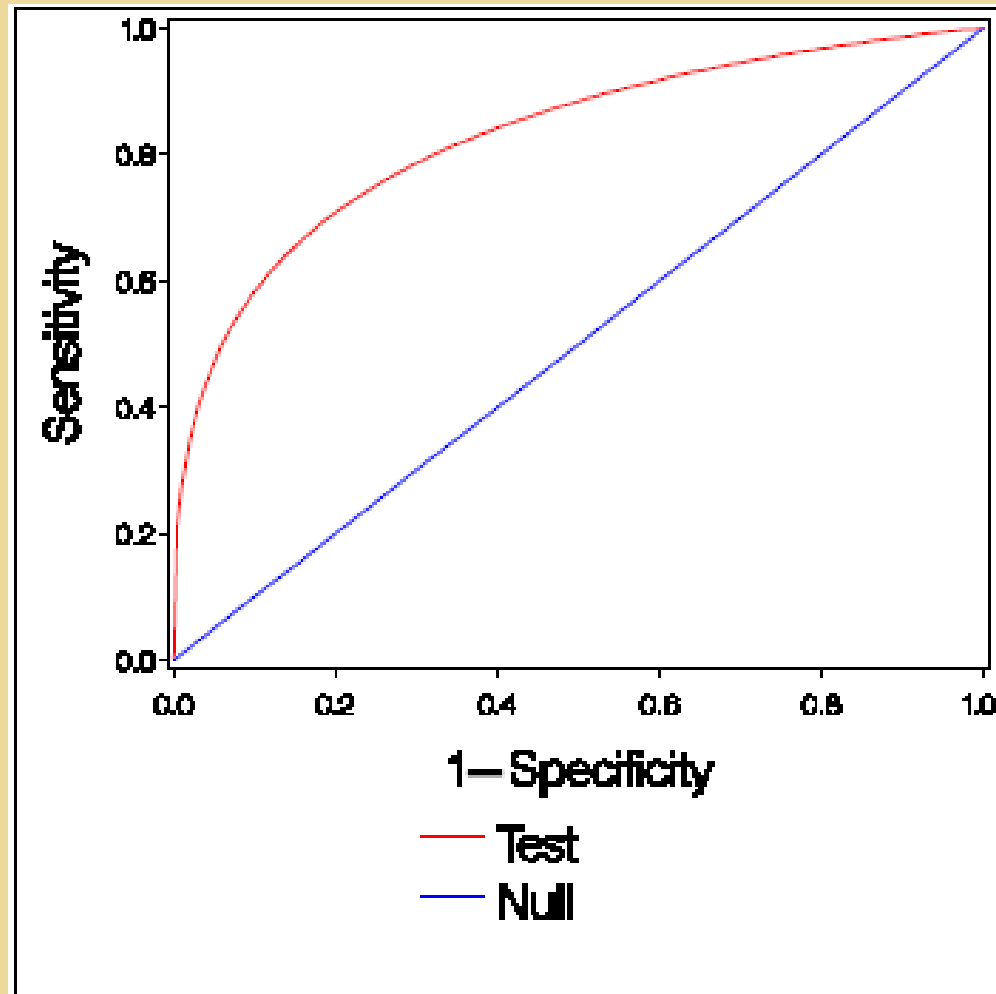
Difference between ROC distributions affects area under the curve (AUC)
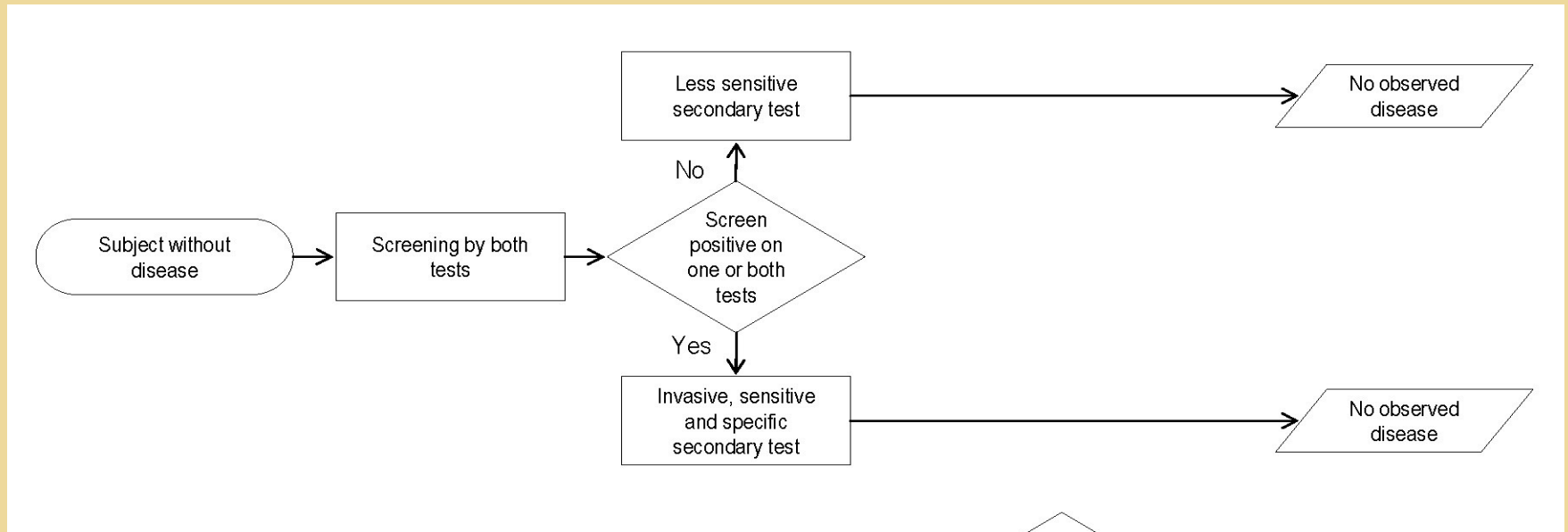

Receiver operating characteristic curve (ROC)
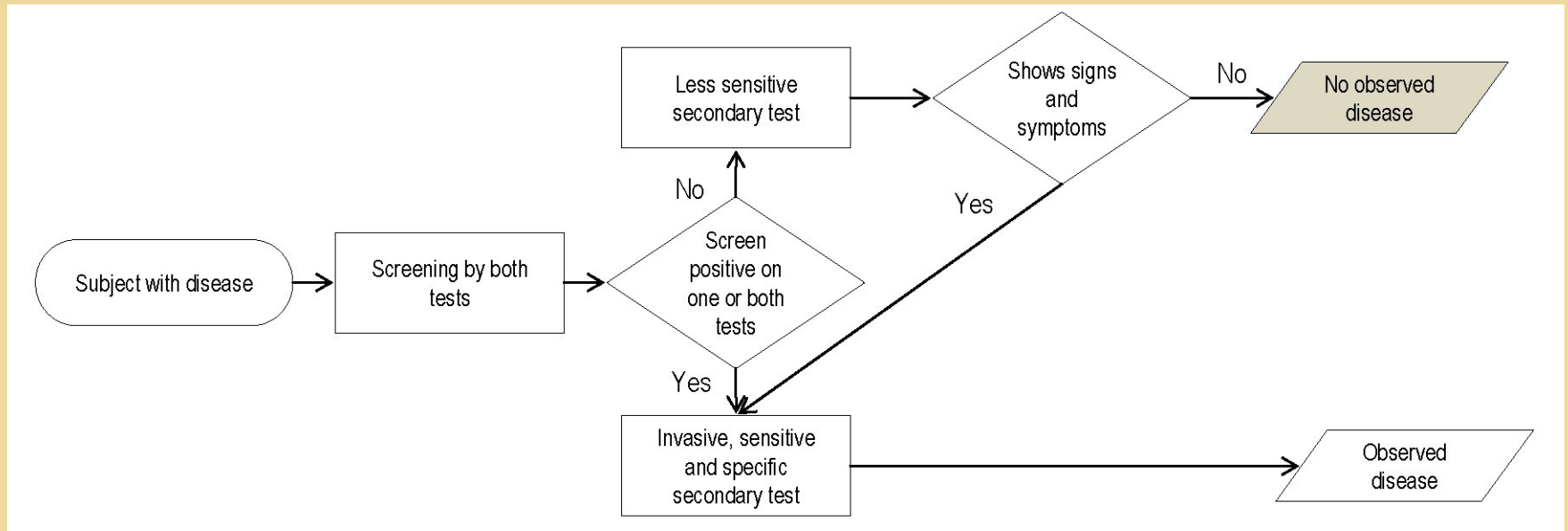
Test 1

| Test 1 —— | Test 2 —— |

| Disease —— | Non-diseased —— |

Design flowchart: participants without disease

Design flowchart: participants with disease

Disease observed in one of four ways

- **Screening test 1 is positive**
- **Screening test 2 is positive**
- **Both screening tests positive**
- **Both screening tests negative, but signs and symptoms occur**

Question

- **Is this trial design affected by paired screening trial bias?**

- **Can the study designer make the wrong decision as to which test is better?**

Two points of view

- **Omniscient
  (know true disease status)**
- **Study coordinator
  (know only observed disease status)**

Classify Participants by

- **Score on Test 1**
- **Score on Test 2**
- **Signs and symptoms**
- **Which secondary test used**

  *Infallible, invasive, sensitive and specific

  *Non-invasive, but less sensitive
- **Observed disease status**
- **True disease status**

Derive

- **True** sensitivity for screening test 1
- **True** specificity for screening test 1
- **True** sensitivity for screening test 2
- **True** specificity for screening test 2
- **Observed** sensitivity for screening test 1
- **Observed** specificity for screening test 1
- **Observed** sensitivity for screening test 2
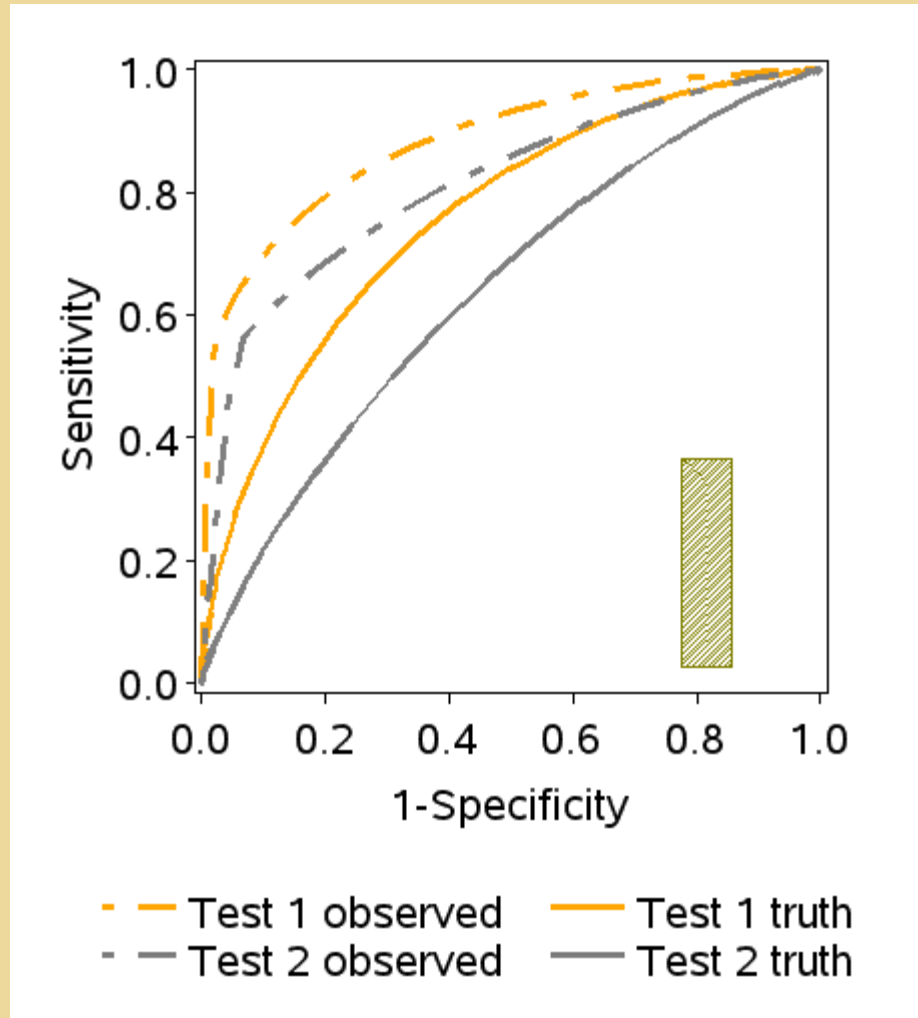- **Observed** specificity for screening

Paired screening trial bias decreases
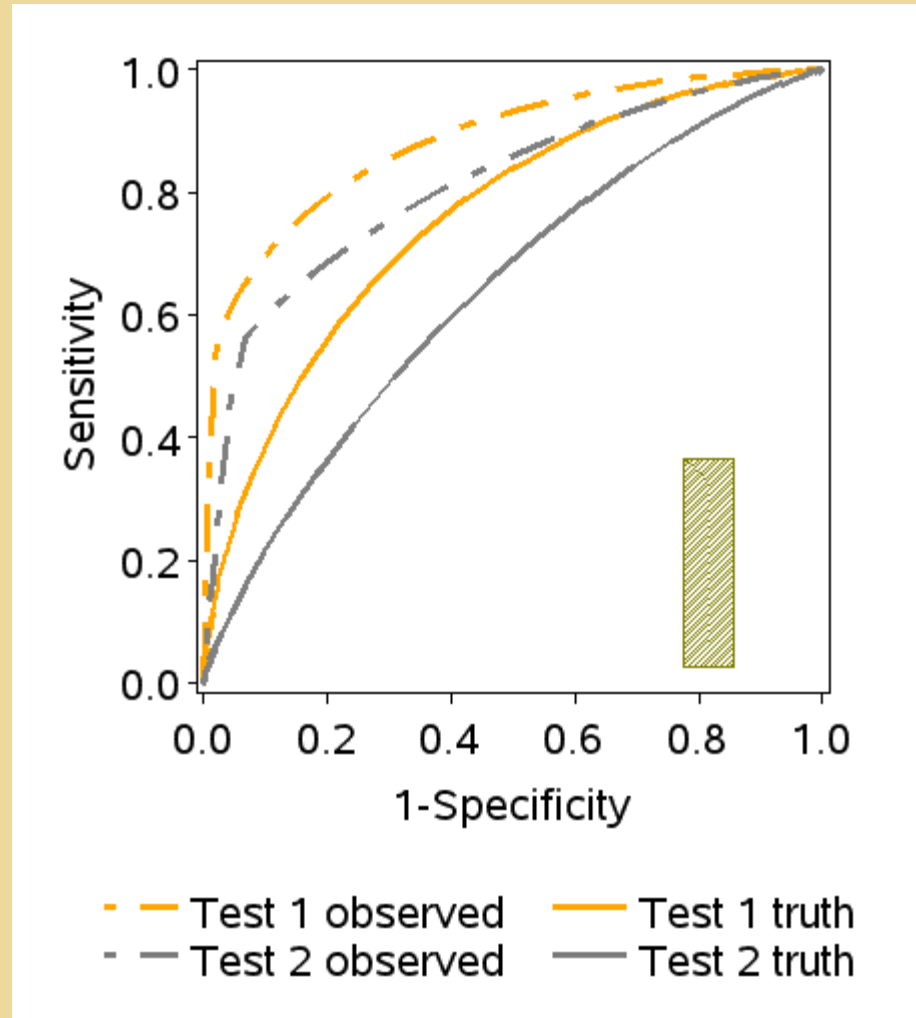
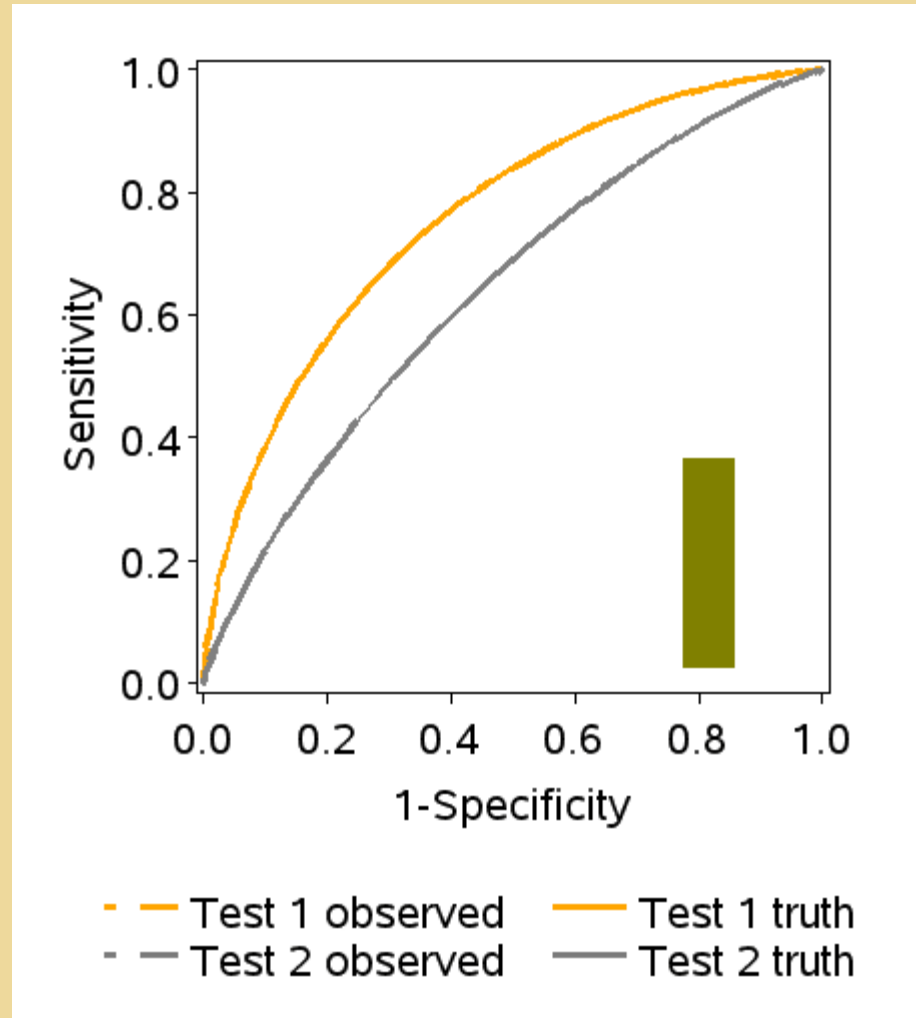as the chance of signs and symptoms increases
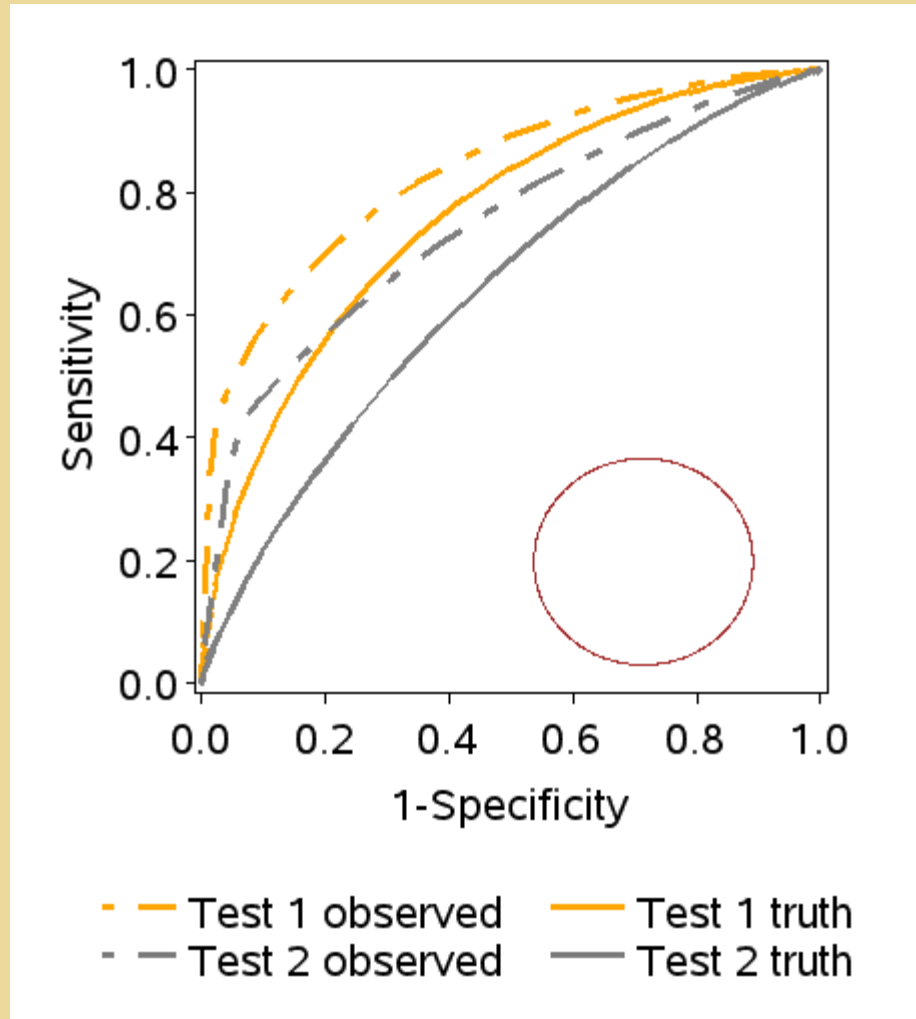
Paired screening trial bias is high

when the rate of signs and symptoms is low
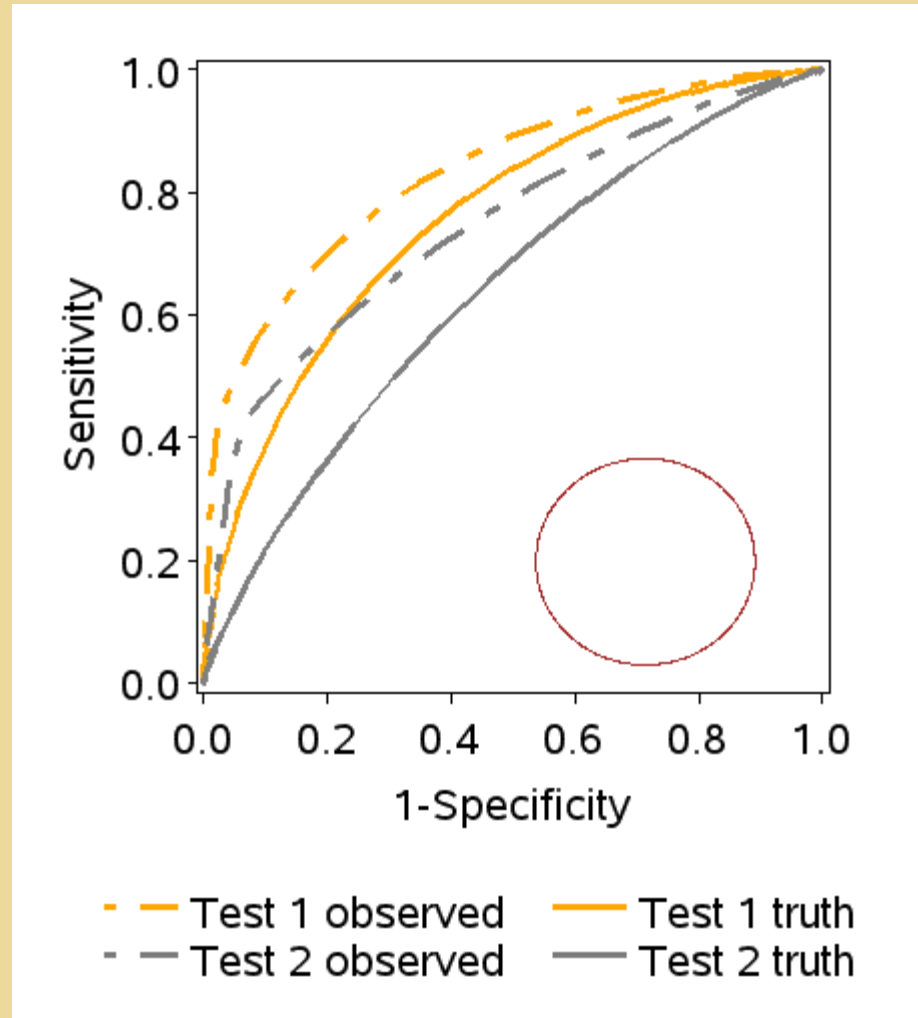
Paired screening trial bias is low

when the rate of signs and symptoms is high

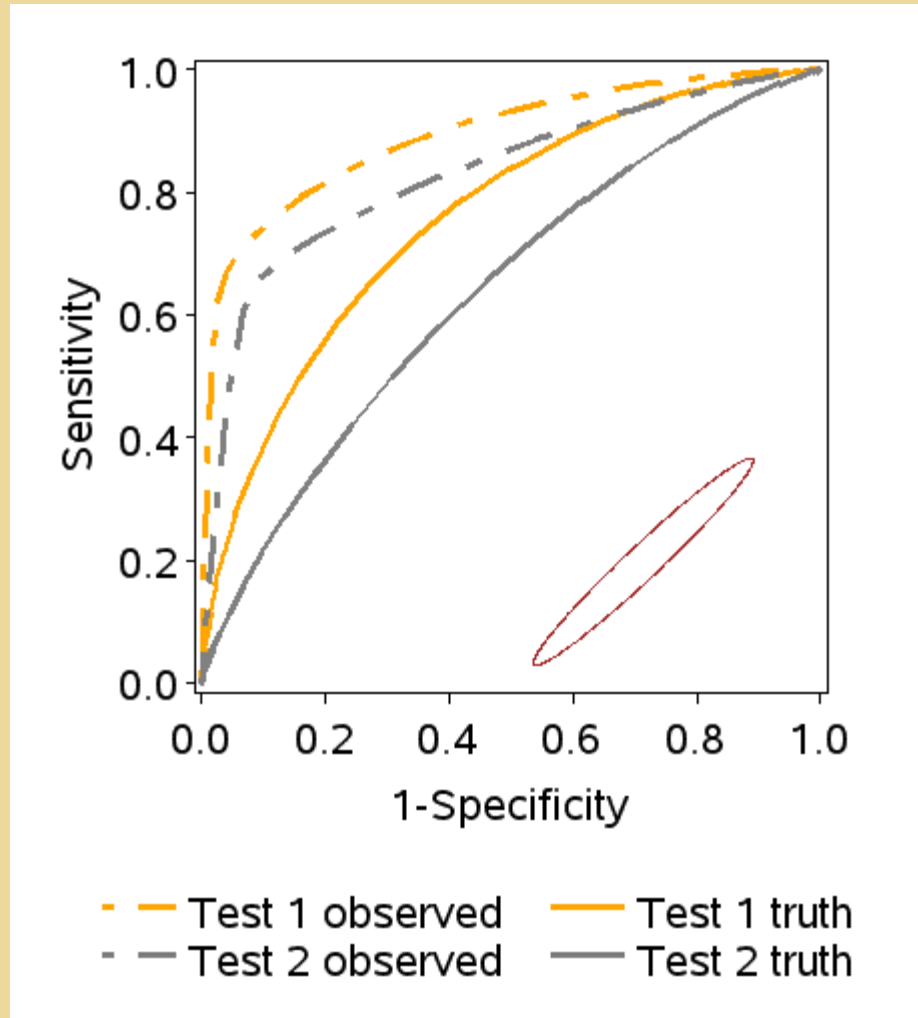# Paired screening trial bias **increases**

he correlation between screening tests increa

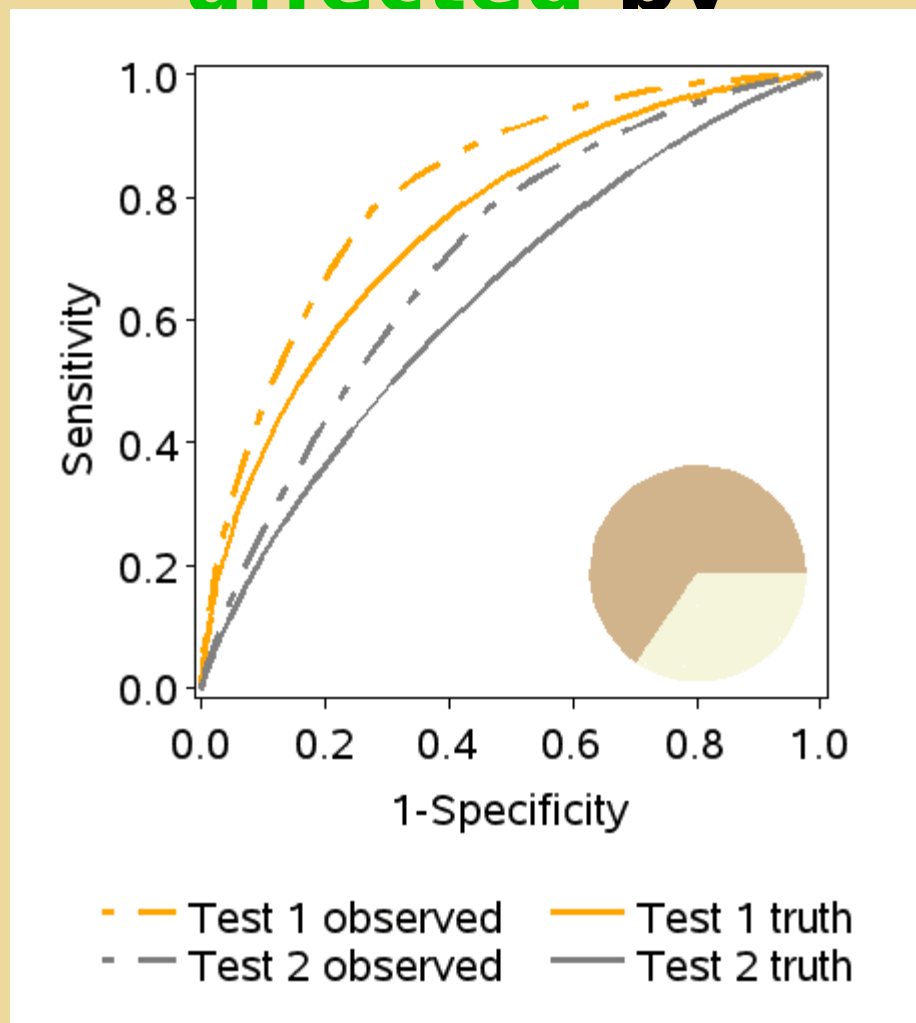# Paired screening trial bias is low

when the correlation is low
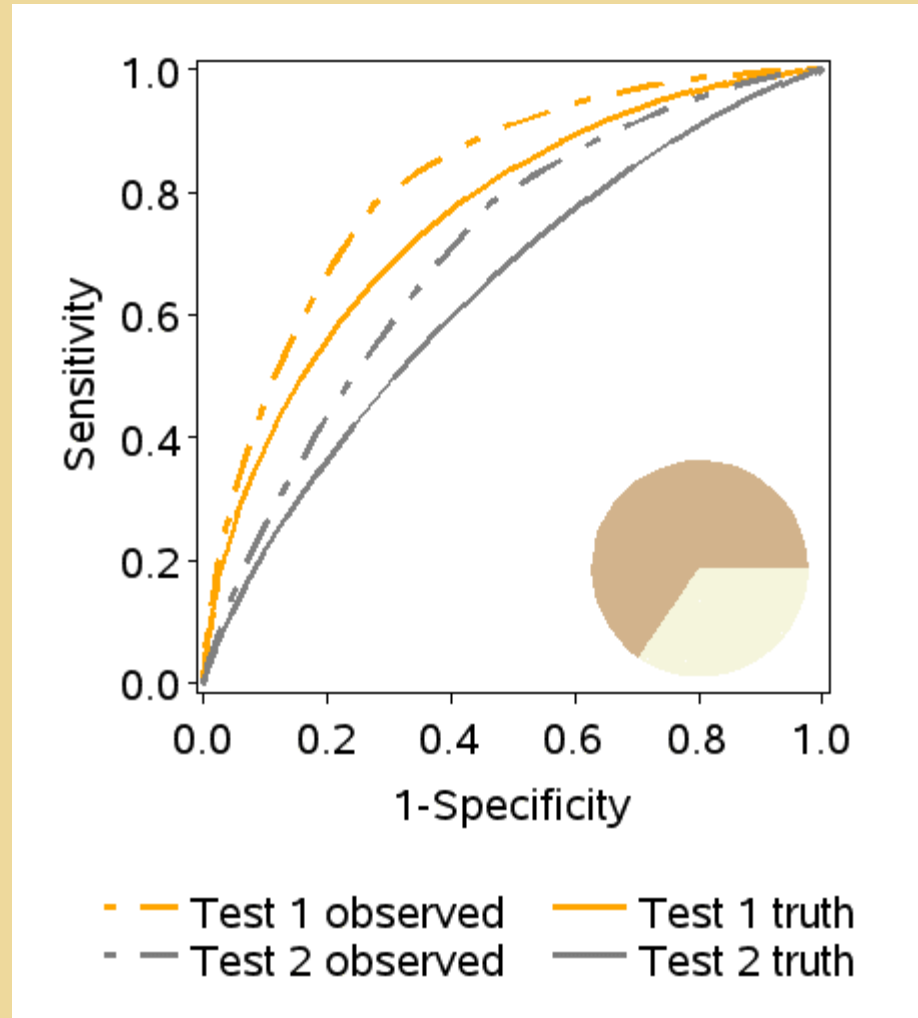
Paired screening trial bias is high

when the correlation is high

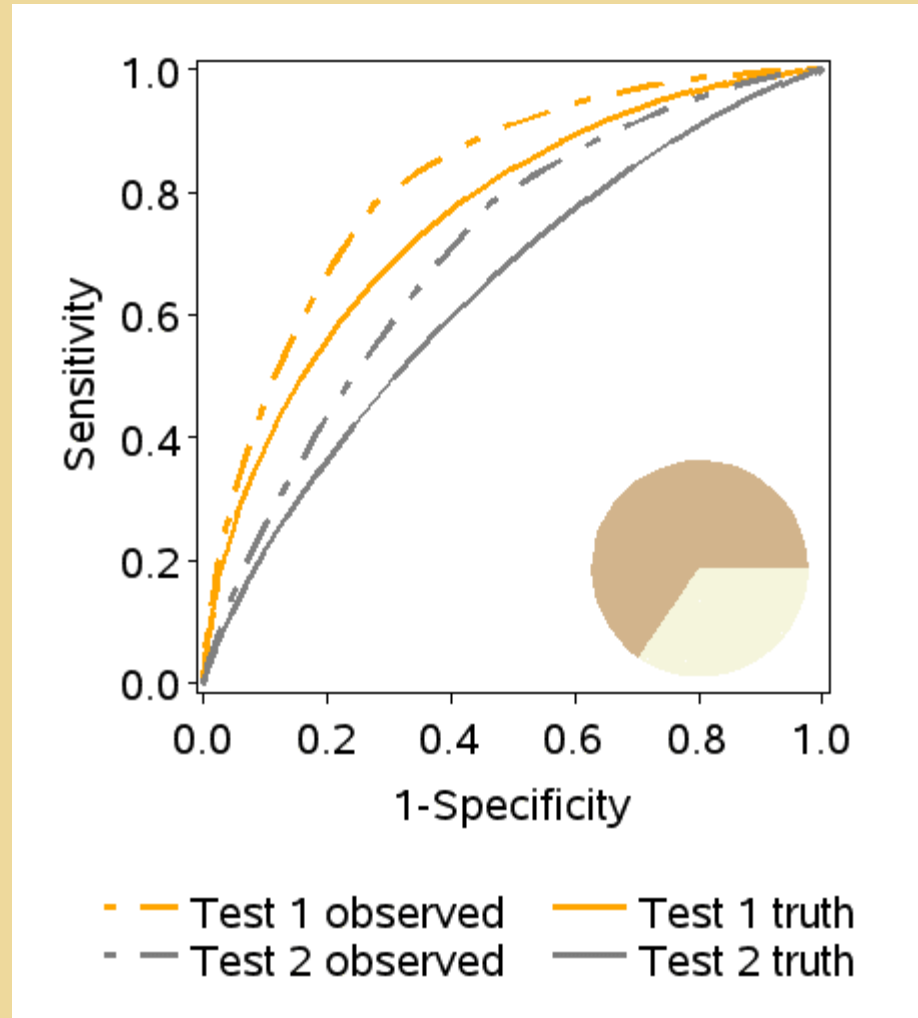# Paired screening trial bias is strongly affected by

the threshold for recall for the infallible test

Paired screening trial bias increases

as fewer cases are observed

Paired screening trial bias gets smaller
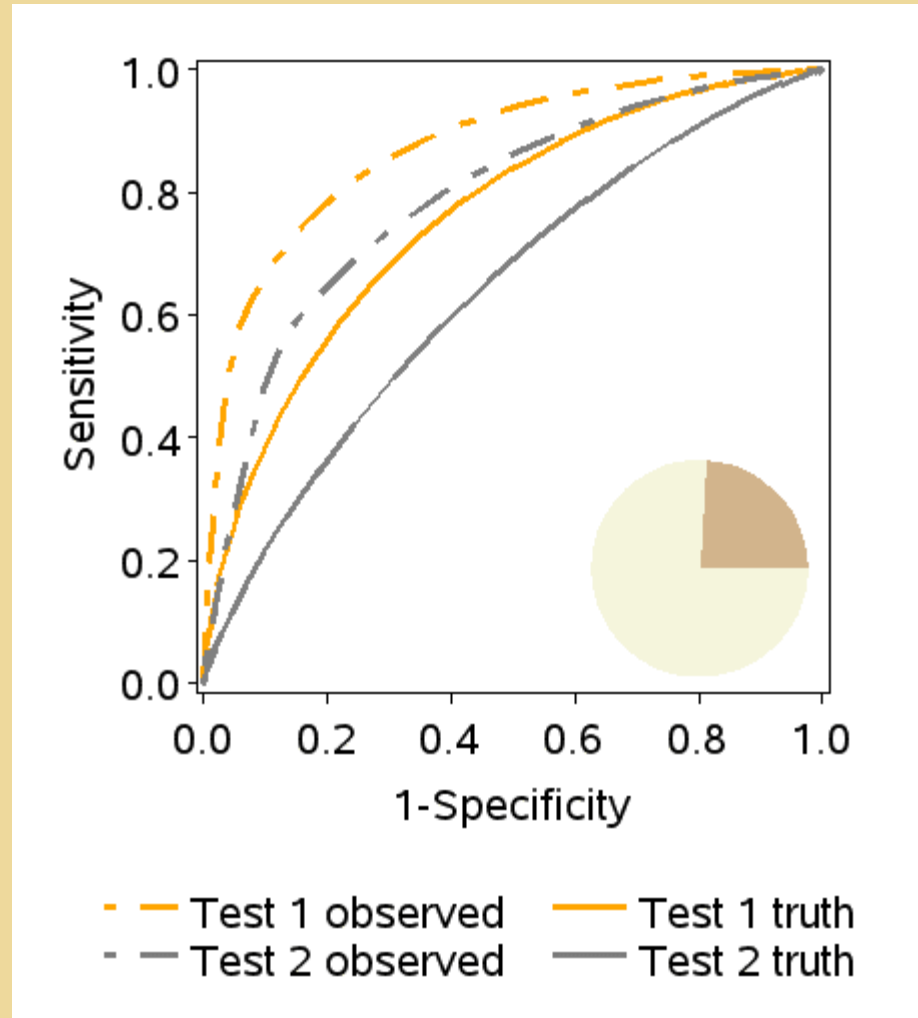

as more cases are observed

Paired screening trial bias gets larger
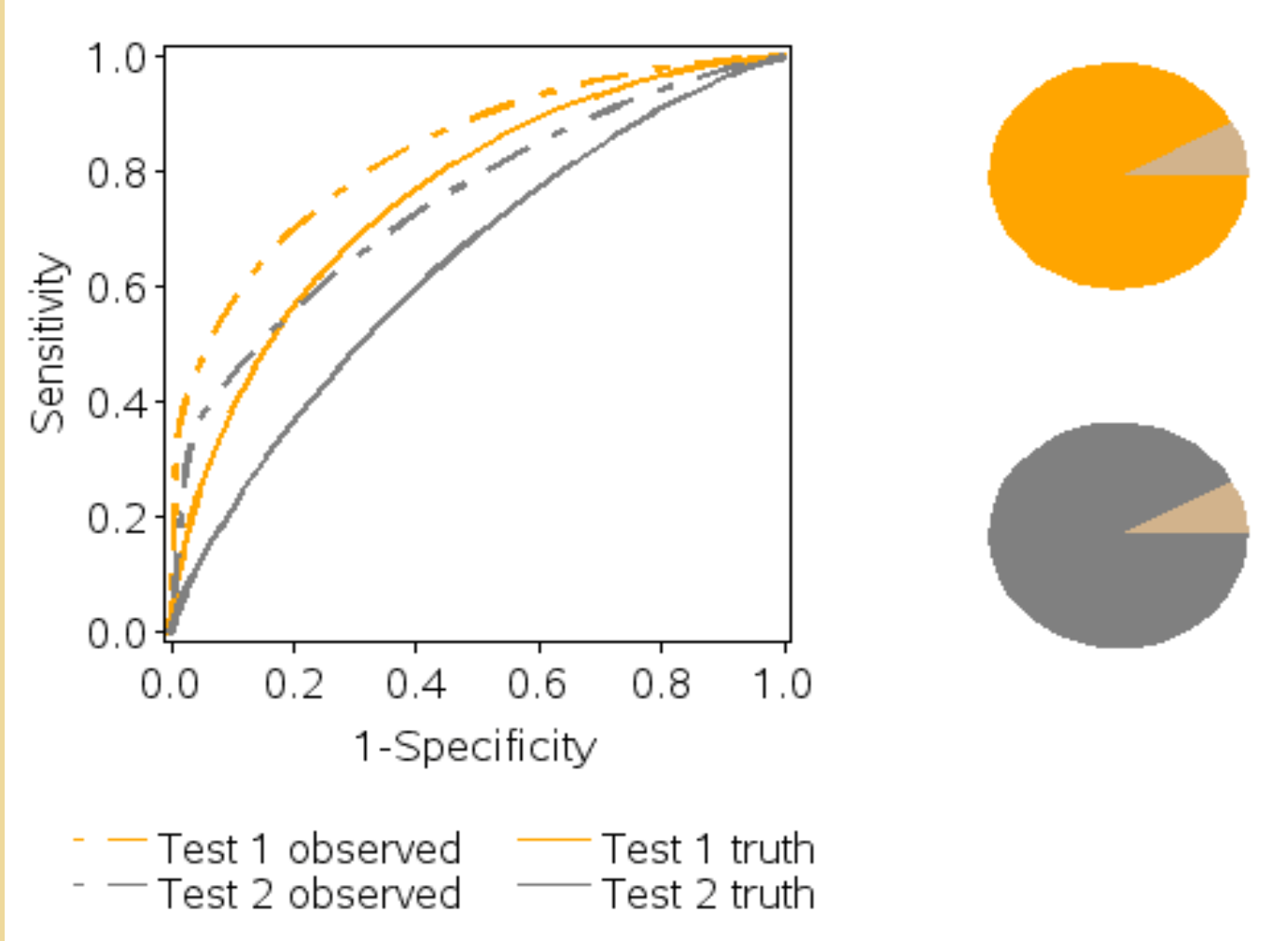

as fewer cases are observed

until...

Paired screening trial bias decreases
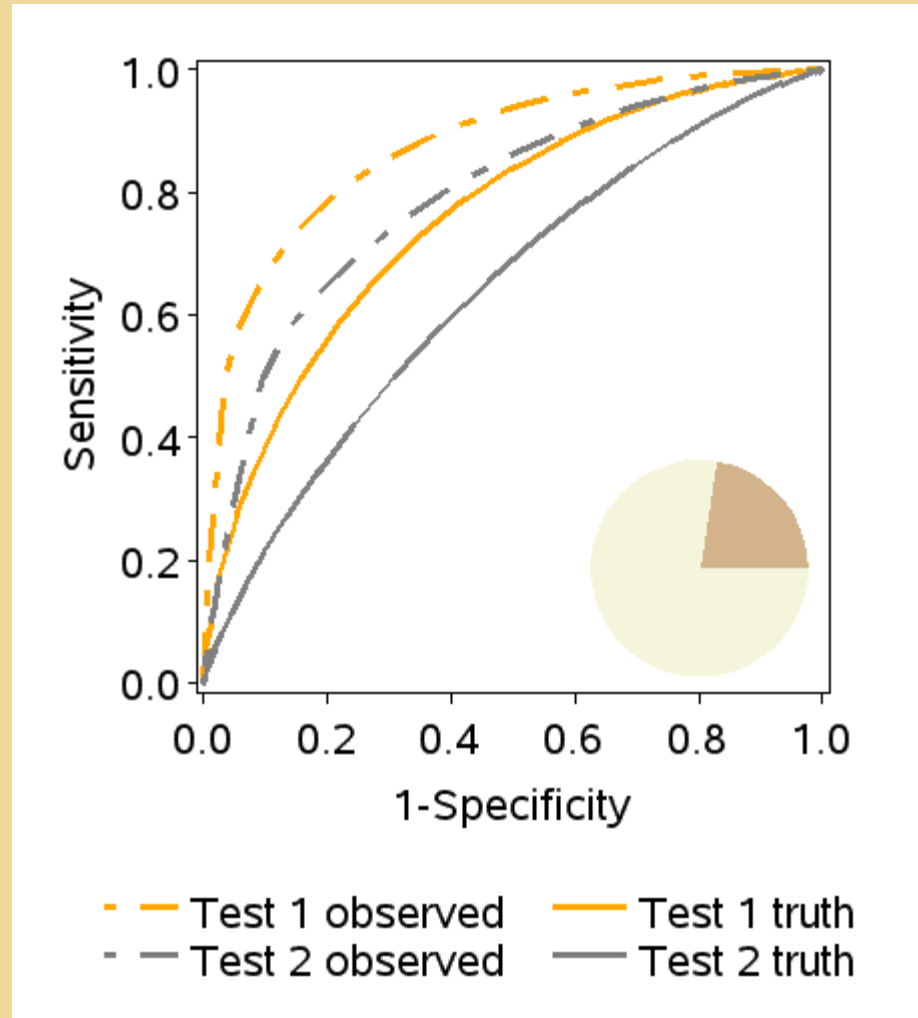
as fewer cases are found

The scientific conclusion may change when

the chance of a case being recalled differs

The scientific conclusion is **correct** when
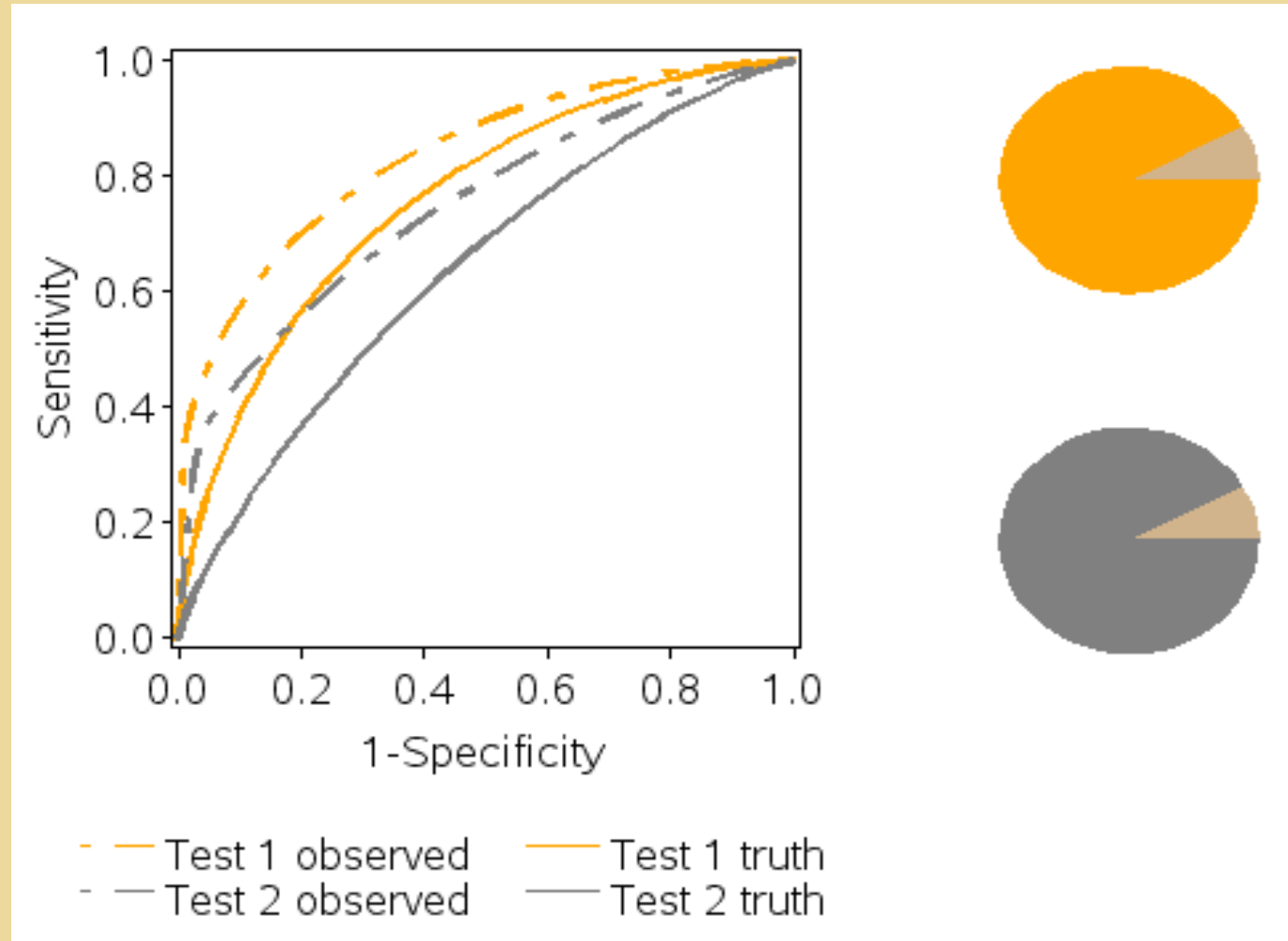
the chance of a case being recalled is the same

The scientific conclusion is wrong when
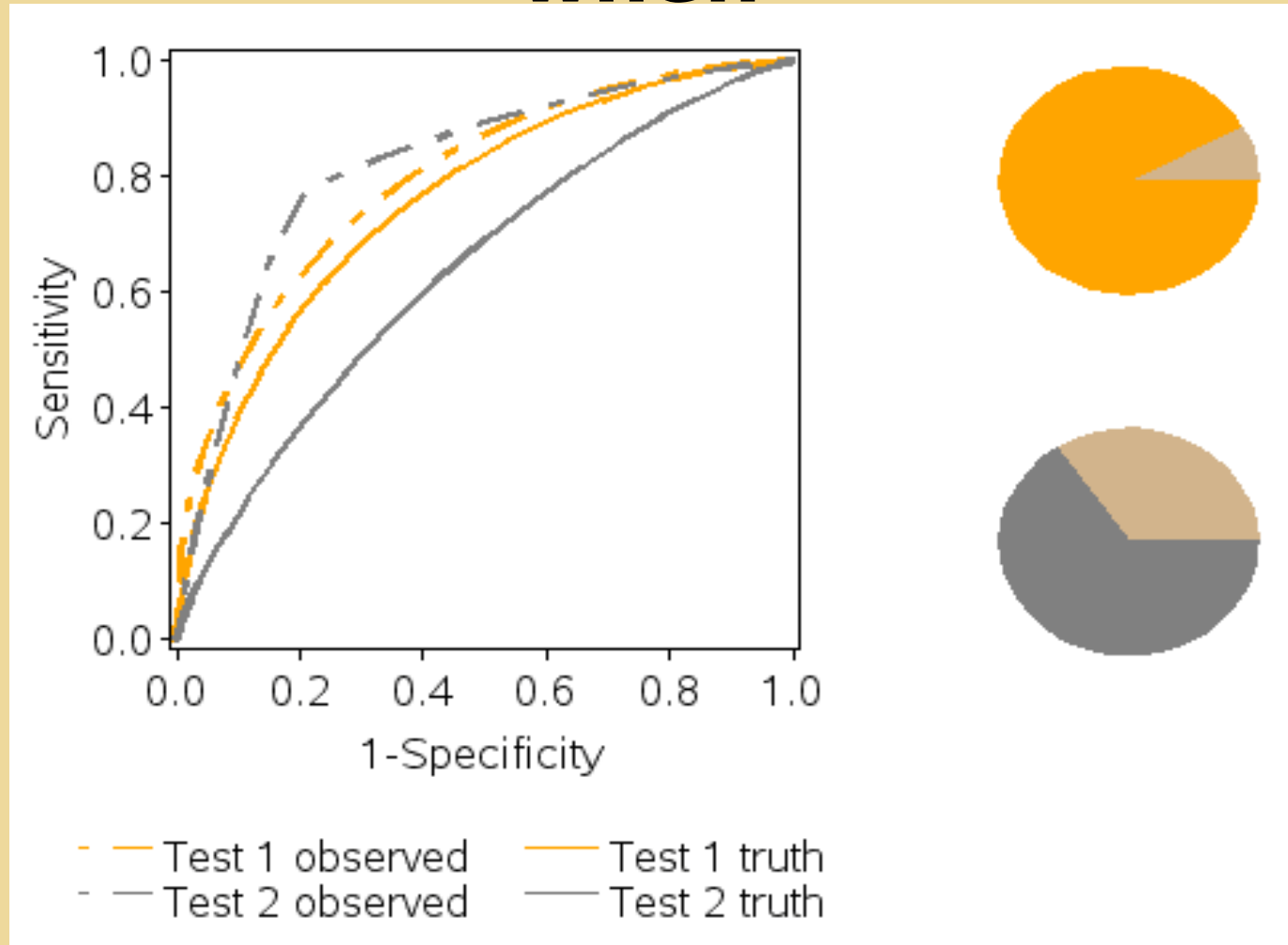
the chance of a case being recalled differs

Conclusions: Paired screening trial bias

- **Decreases** as chance of signs and symptoms for those with disease increases

- **Increases** as correlation between screening tests increases

- Is **strongly affected** by the threshold for recall for the infallible test

Conclusions

- **Paired screening trial bias can lead the study designer to decide screening Test 1 is better than screening Test 2, when in fact, screening Test 2 is better than screening Test 1**

- **Decisions made in paired screening trials affect the health of millions of people**

- **There is a need for mathematical bias correction**

References

Berg WA, Blume JD, Cormack JB, Mendelson EB, Lehrer D, Böhm-Vélez M, Pisano ED, Jong RA, Evans WP, Morton MJ, Mahoney MC, Larsen LH, Barr RG, Farria DM, Marques HS, Boparai K for the ACRIN 6666 Investigators: Combined screening with ultrasound and mammography vs mammography alone in women at elevated risk of breast cancer. J Am Med Assoc 2008, 299: 2151-2163.

Berg WA, Gutierrez L, NessAiver MS, Carter WB, Bhargavan M, Lewis RS, Ioffe OB.Radiology. 2004 Dec;233(3):830-49. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer.

DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837–845.

Glueck DH, Lamb MM, O'Donnell CI, Ringham BM, Brinton JT, Muller KE, Lewin JM, Alonzo TA, Pisano ED: Bias in trials comparing paired continuous tests can cause researchers to choose the wrong screening modality. To appear in BMC Medical Research Methodology, 2009.

Metz, C.;Wang, P.; Kronman, HA. New approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F. , editor. Information processing in medical imaging. The Hague, the Netherlands: Nijihoff; 1984.

Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, Conant EF, Fajardo LL, Basett L, D'Orsi C, Jong R and Rebner M: Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005; 353:1773-1783.