Reducing Decision Errors in the Paired Comparison of the Diagnostic Accuracy of Continuous Screening Tests

Brandy M. Ringham,[1] Todd A. Alonzo,[2] John T. Brinton,[1]
Aarti Munjal,[1] Keith E. Muller,[3] Deborah H. Glueck[1]

[1]Department of Biostatistics and Informatics, University of Colorado Denver
[2]Department of Preventive Medicine, University of Southern California
[3]Department of Health Outcomes and Policy, University of Florida

Acknowledgements

The project described was supported by Award Number NCI 1R03CA136048-01A1 from the National Cancer Institute, and by Award Number NIDCR 3R01DE020832-01A1S1 from the National Institute of Dental and Craniofacial Research. The content is solely the responsibility of the authors, and does not necessarily represent the official views of the National Cancer Institute, the National Institute of Dental and Craniofacial Research nor the National Institutes of Health.

Outline

Science	• Case Study • Cancer Screening Trial Design • Cancer Screening Analysis
Statistics	• Bias Correction Algorithm • Evaluation Studies • Oral Cancer Screening Demonstration

Oral Cancer Screening Case Study

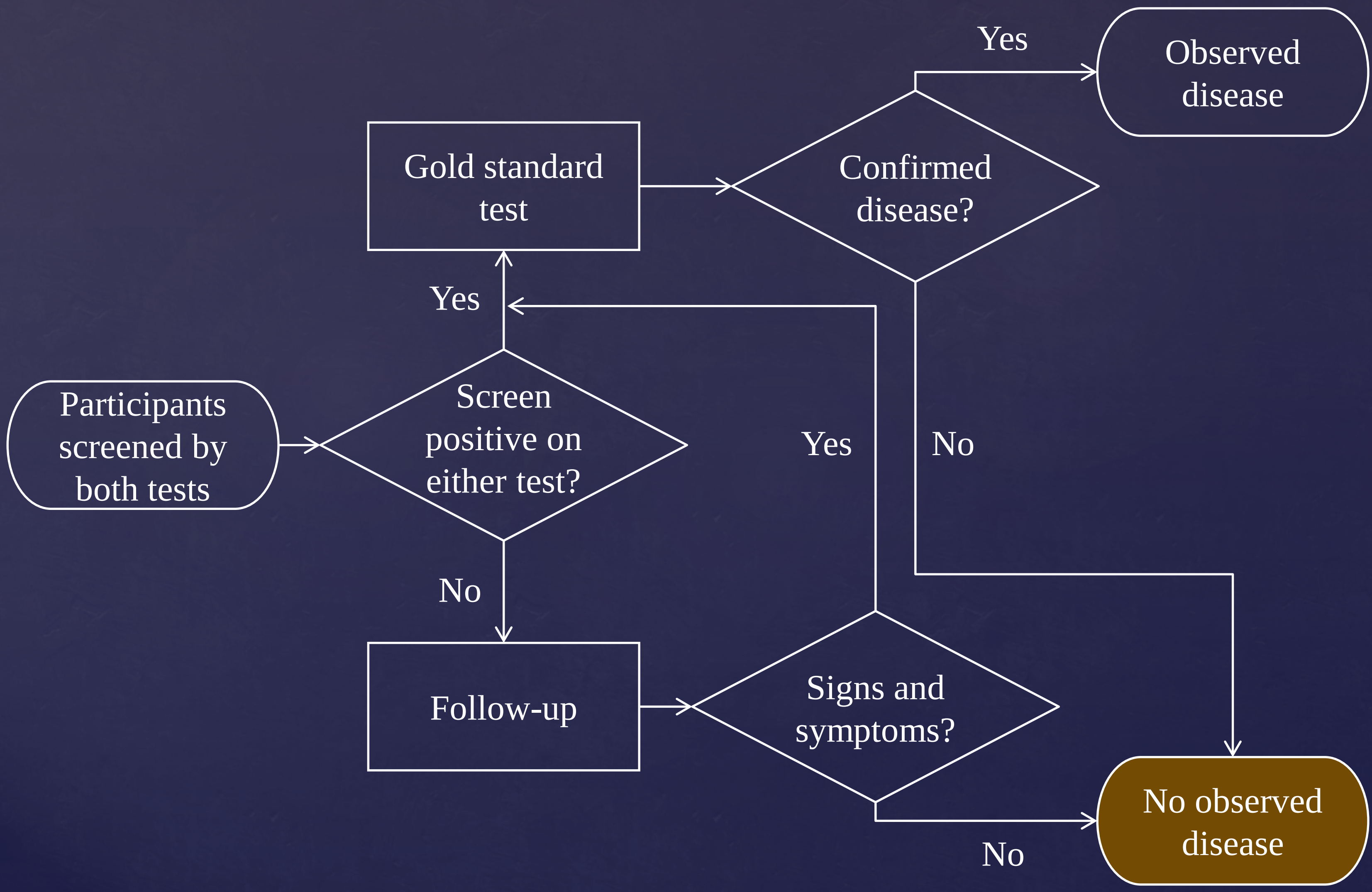
VISIBLE LIGHT

No visible lesion

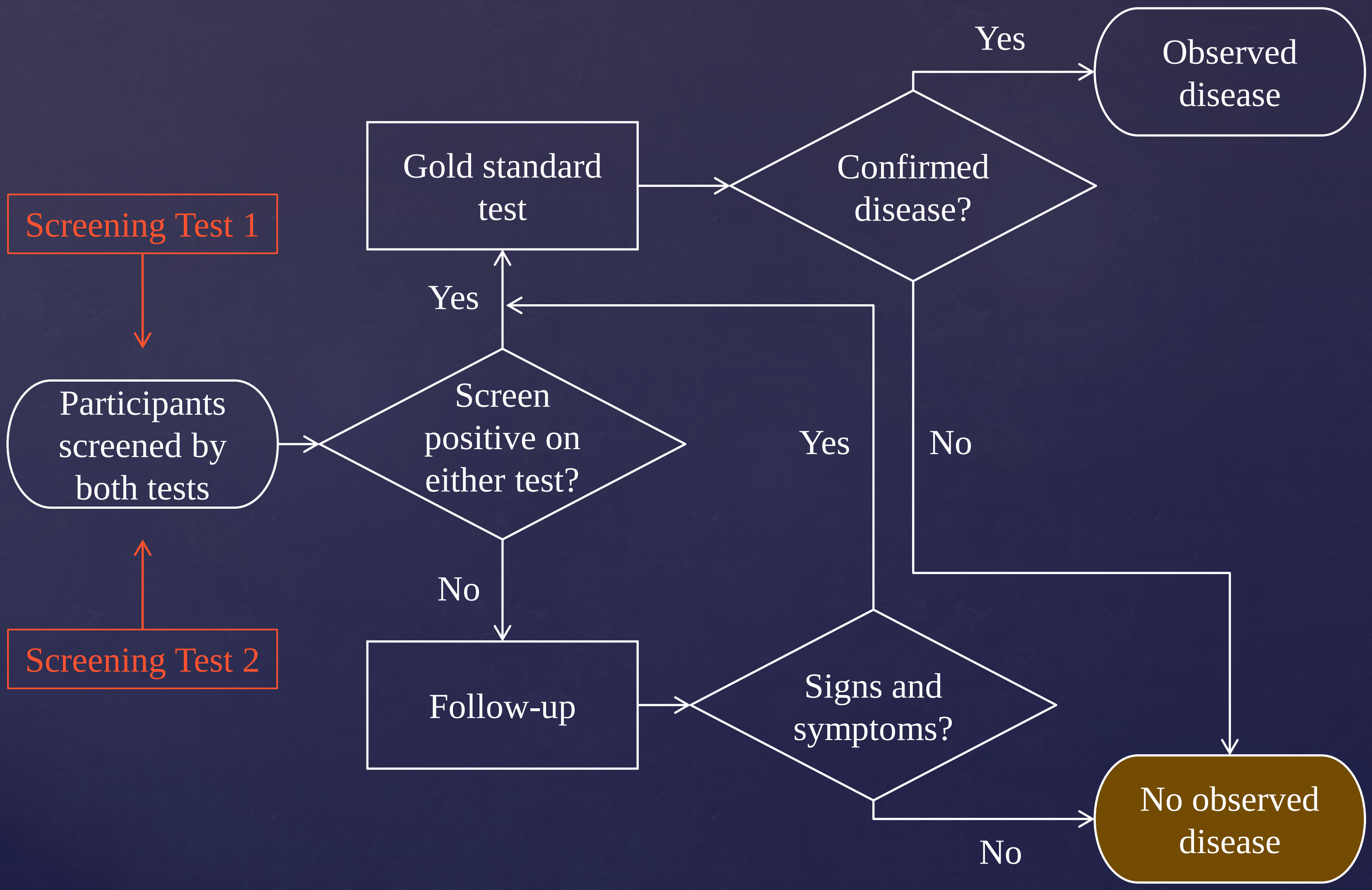
AUTOFLUORESCENCE

Dark region confirmed to be carcinoma in situ

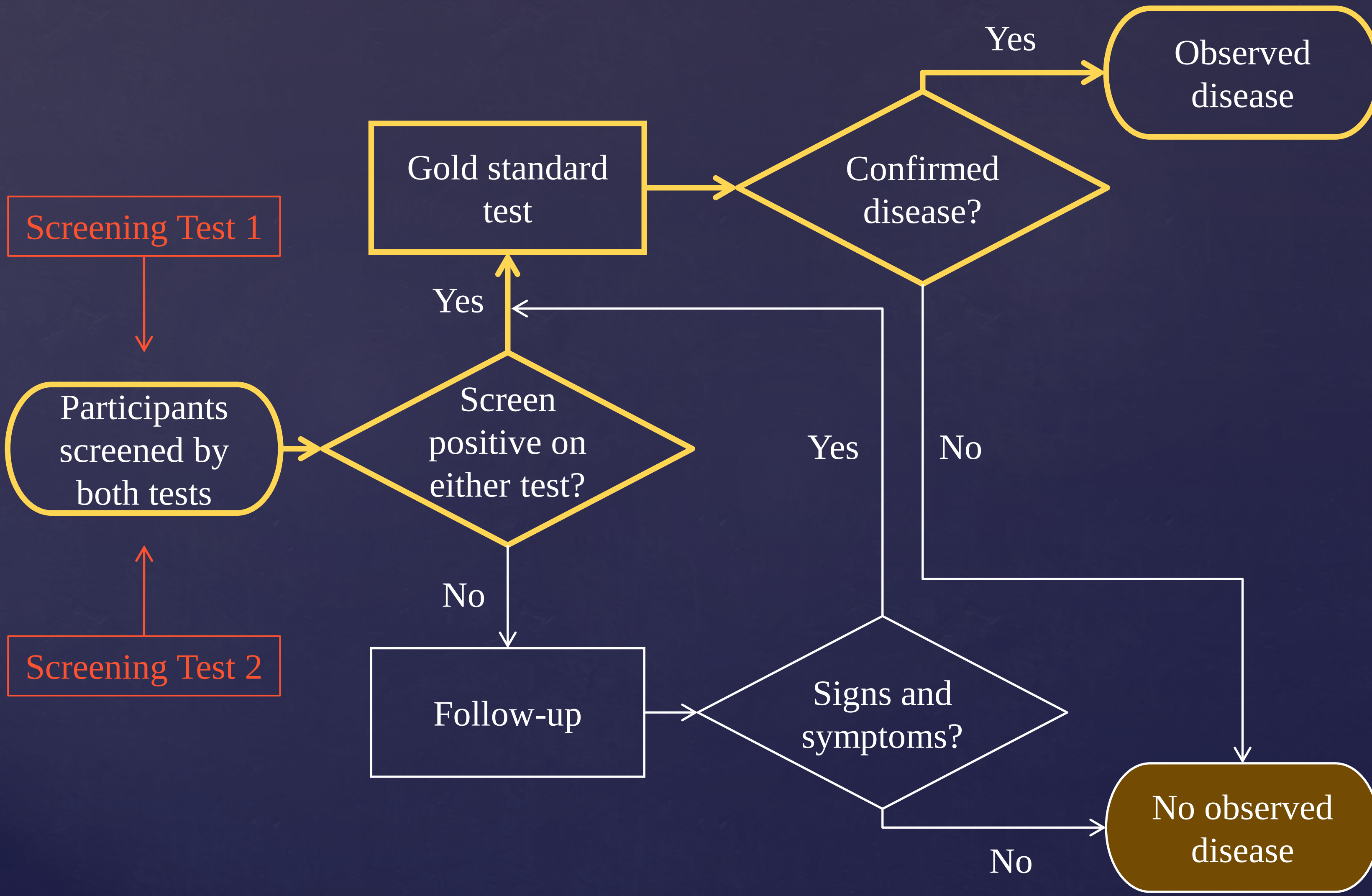
Paired Cancer Screening Trial



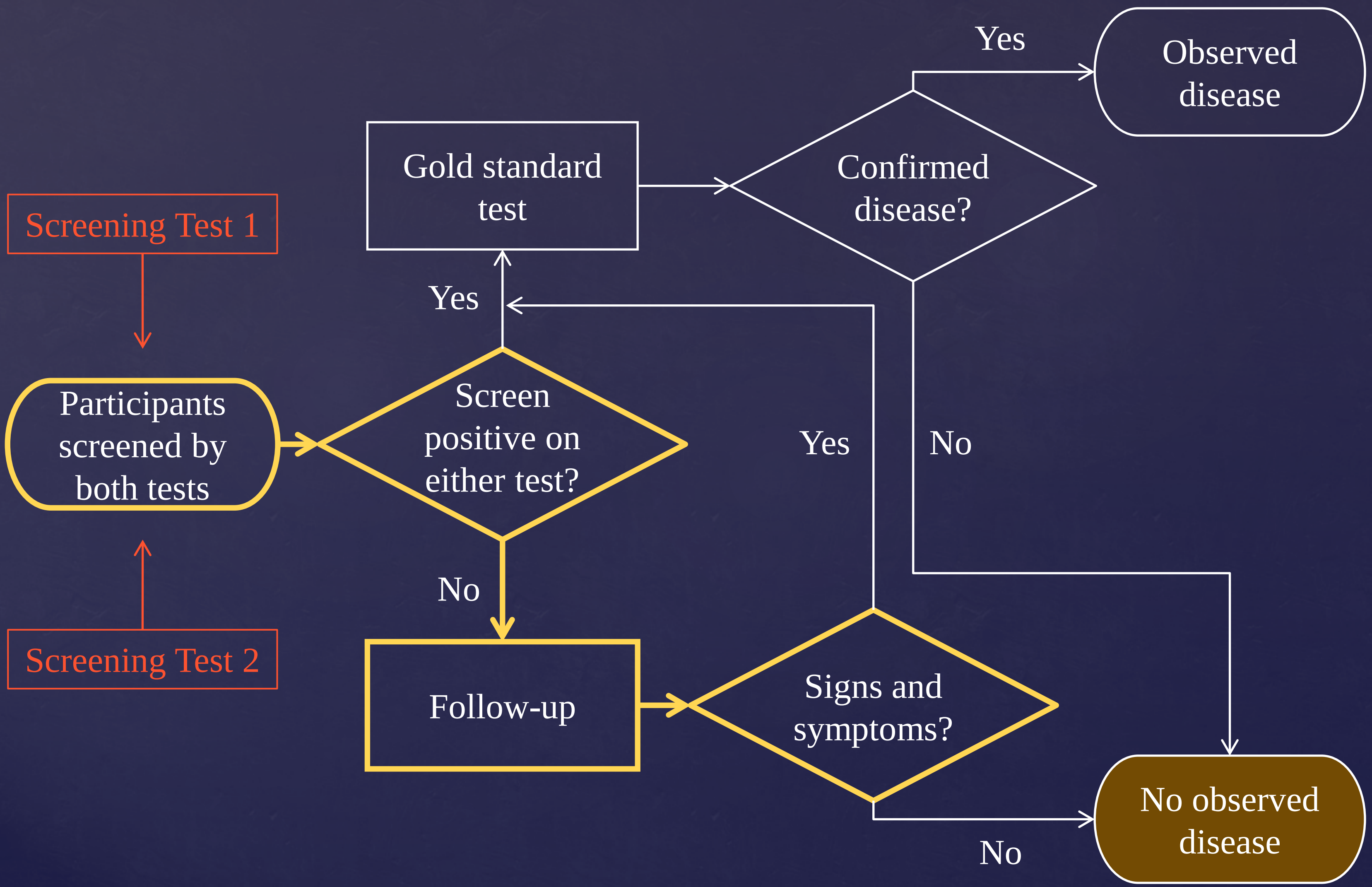
Paired Cancer Screening Trial



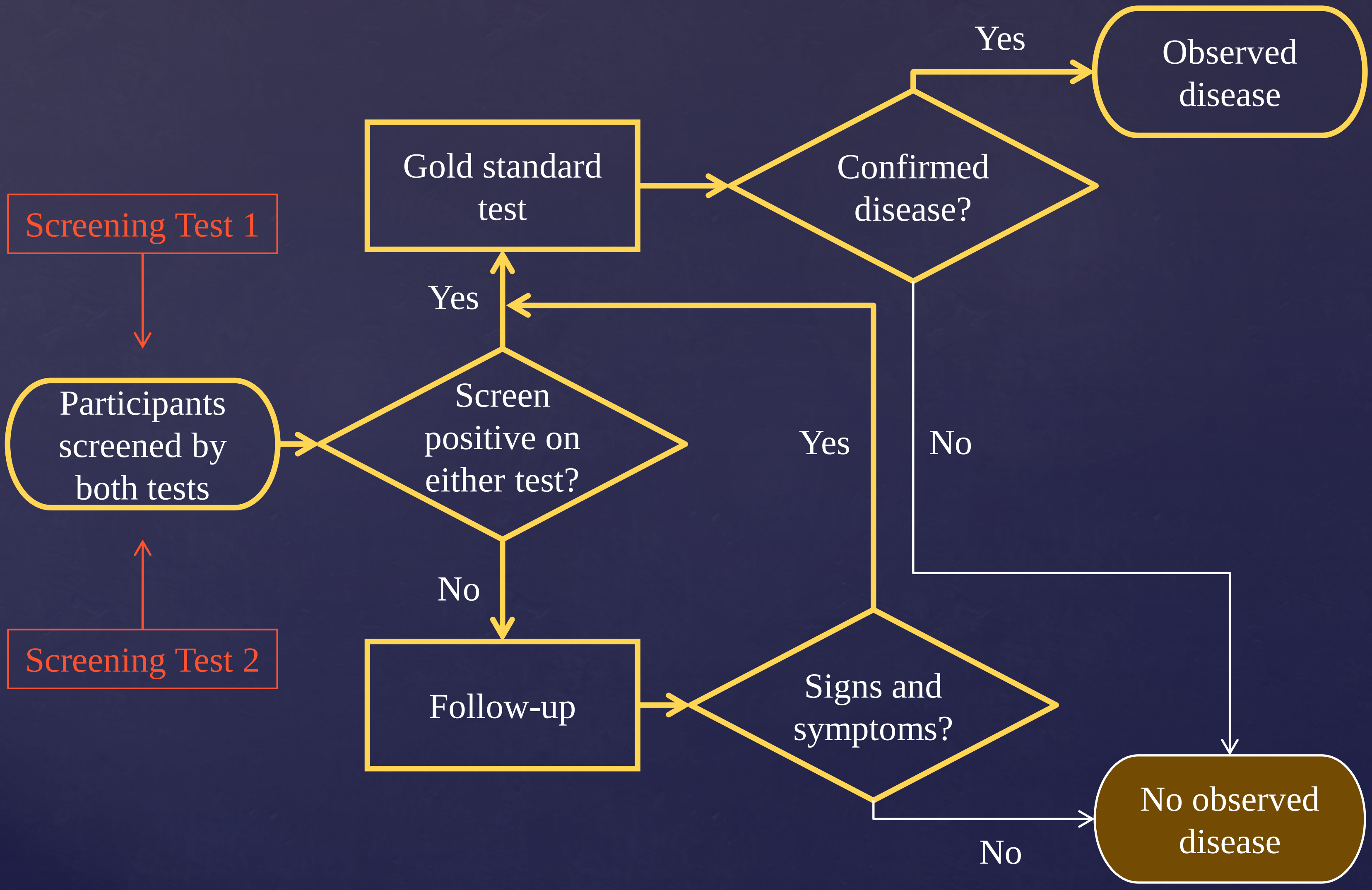
Paired Cancer Screening Trial



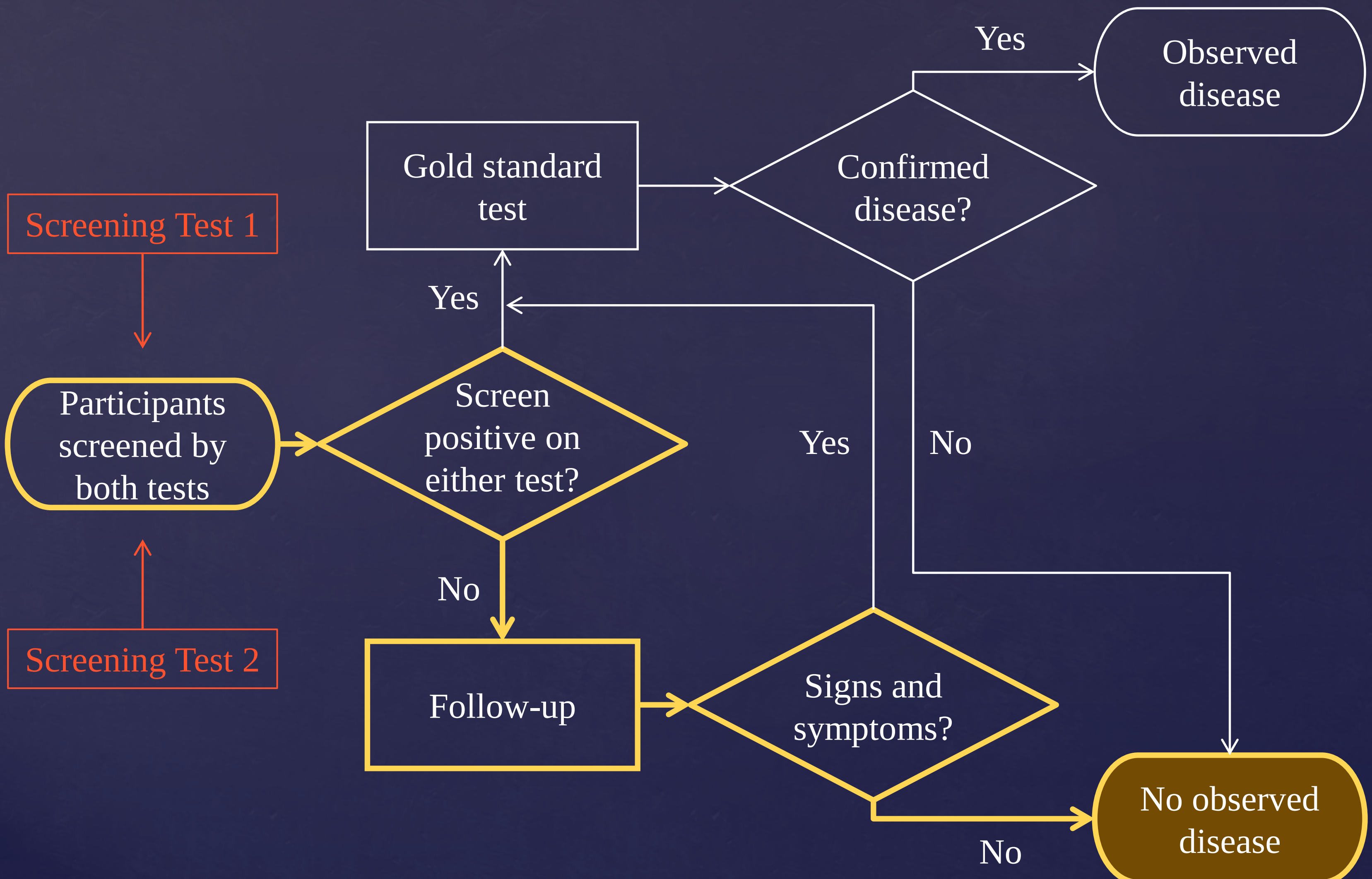
Paired Cancer Screening Trial



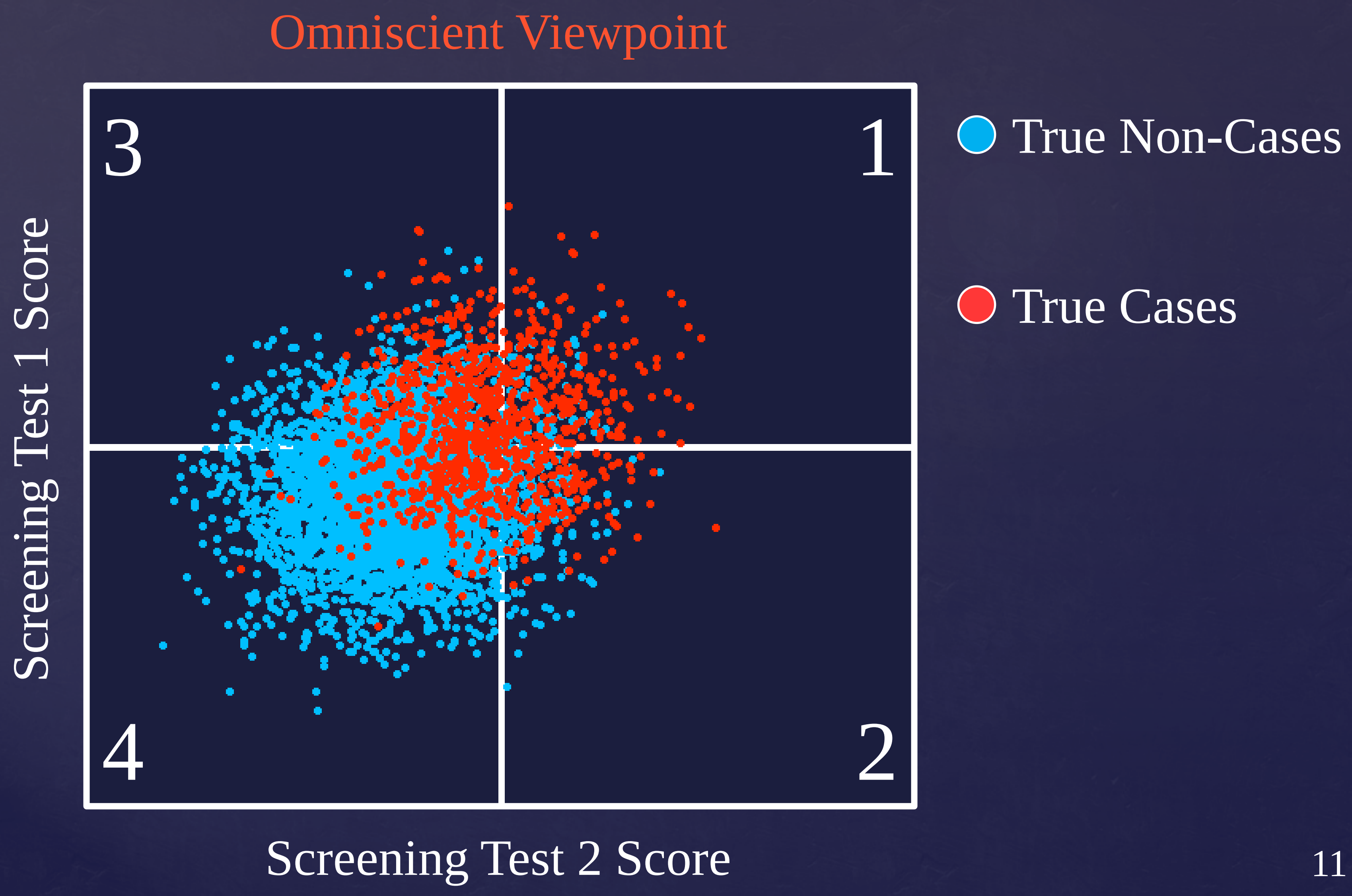
Paired Cancer Screening Trial



Paired Cancer Screening Trial

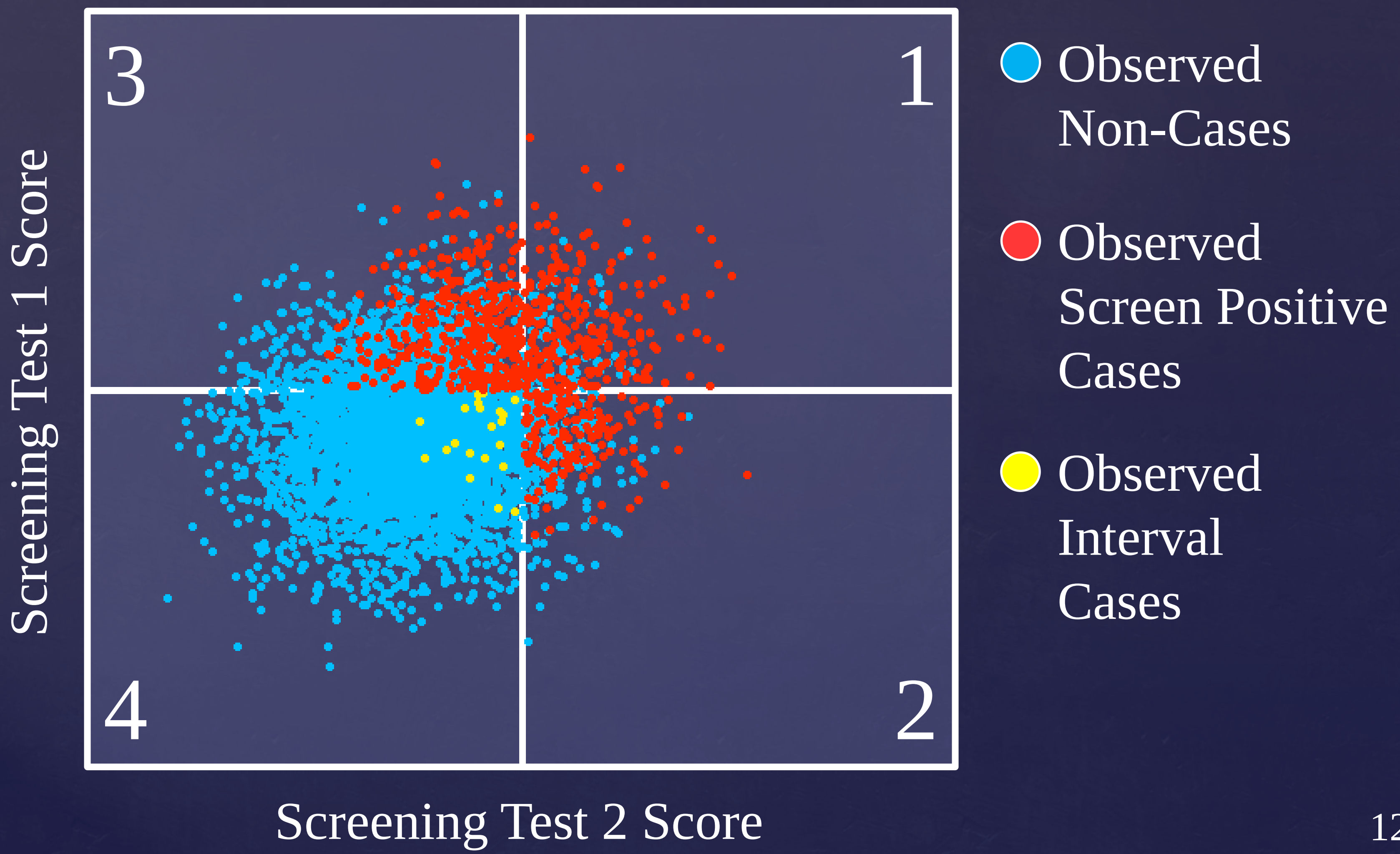


Hypothetical Cancer Screening Data

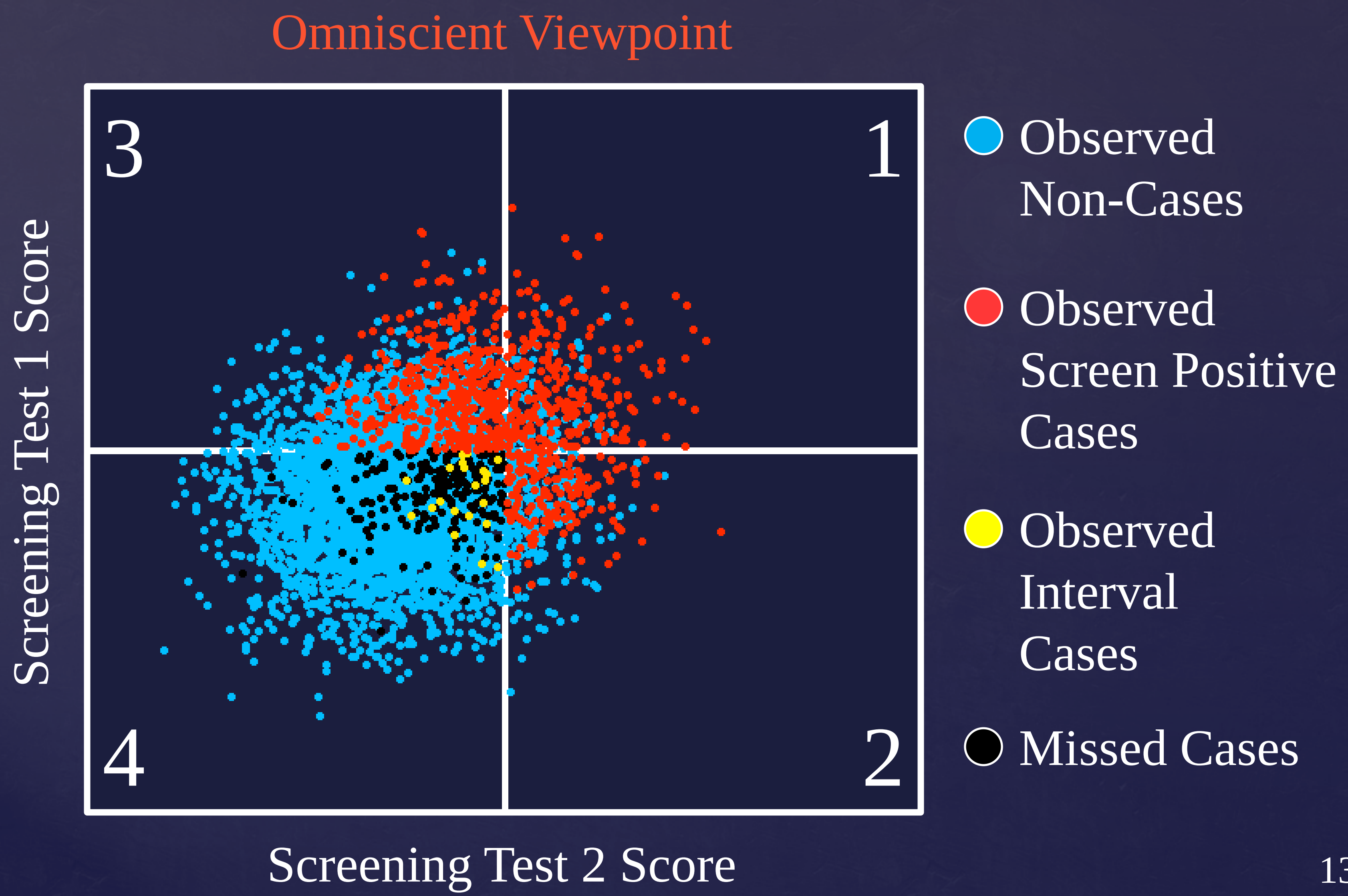


Hypothetical Cancer Screening Data

Study Investigator's Viewpoint

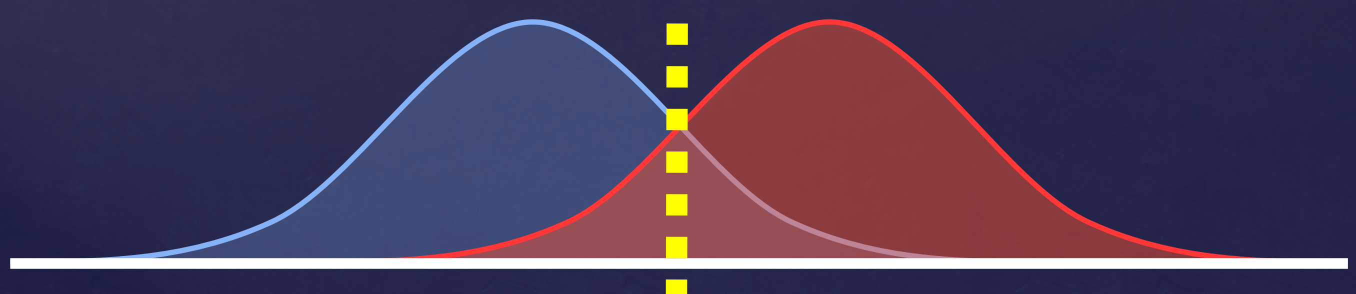


Hypothetical Cancer Screening Data



Analysis of Cancer Screening Studies

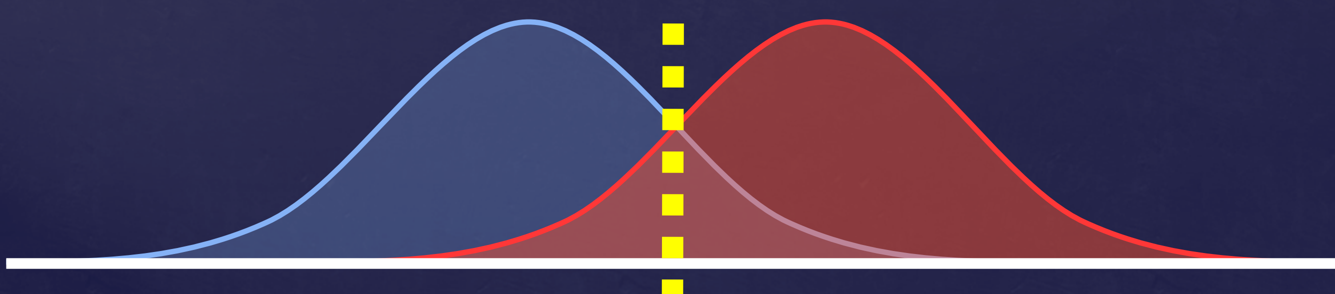
Screening Test	Disease Status +	Disease Status -
+	a	b
-	c	d



Analysis of Cancer Screening Studies

Screening Test	Disease Status +	Disease Status -
+	a	b
-	c	d

$$\text{Sensitivity} = \frac{a}{a+c}$$

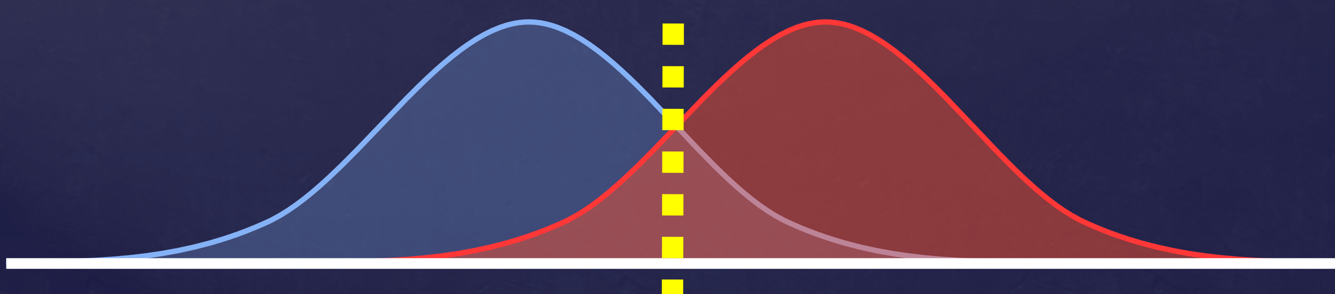


Analysis of Cancer Screening Studies

Screening Test	Disease Status +	Disease Status -
+	a	b
-	c	d

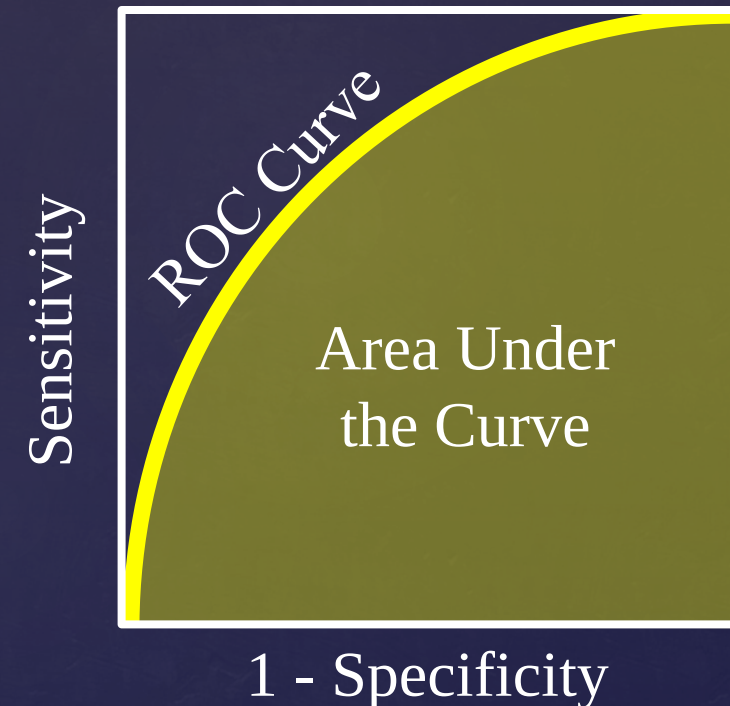
$$\text{Sensitivity} = \frac{a}{a+c}$$

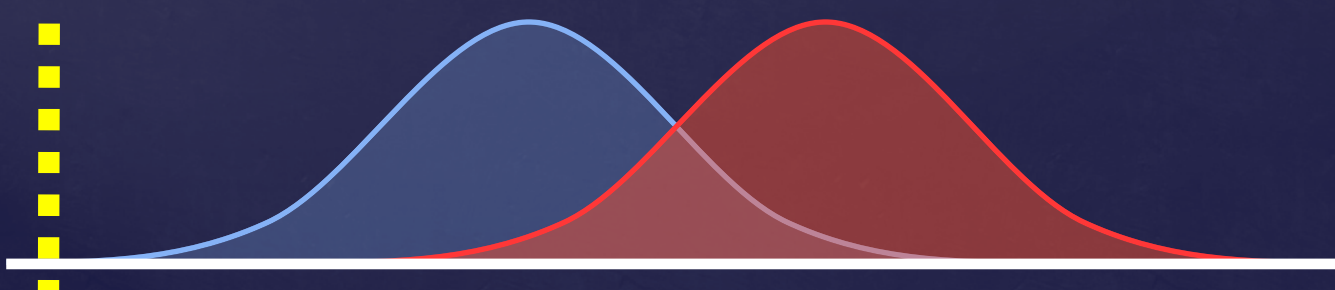
$$\text{Specificity} = \frac{d}{b+d}$$



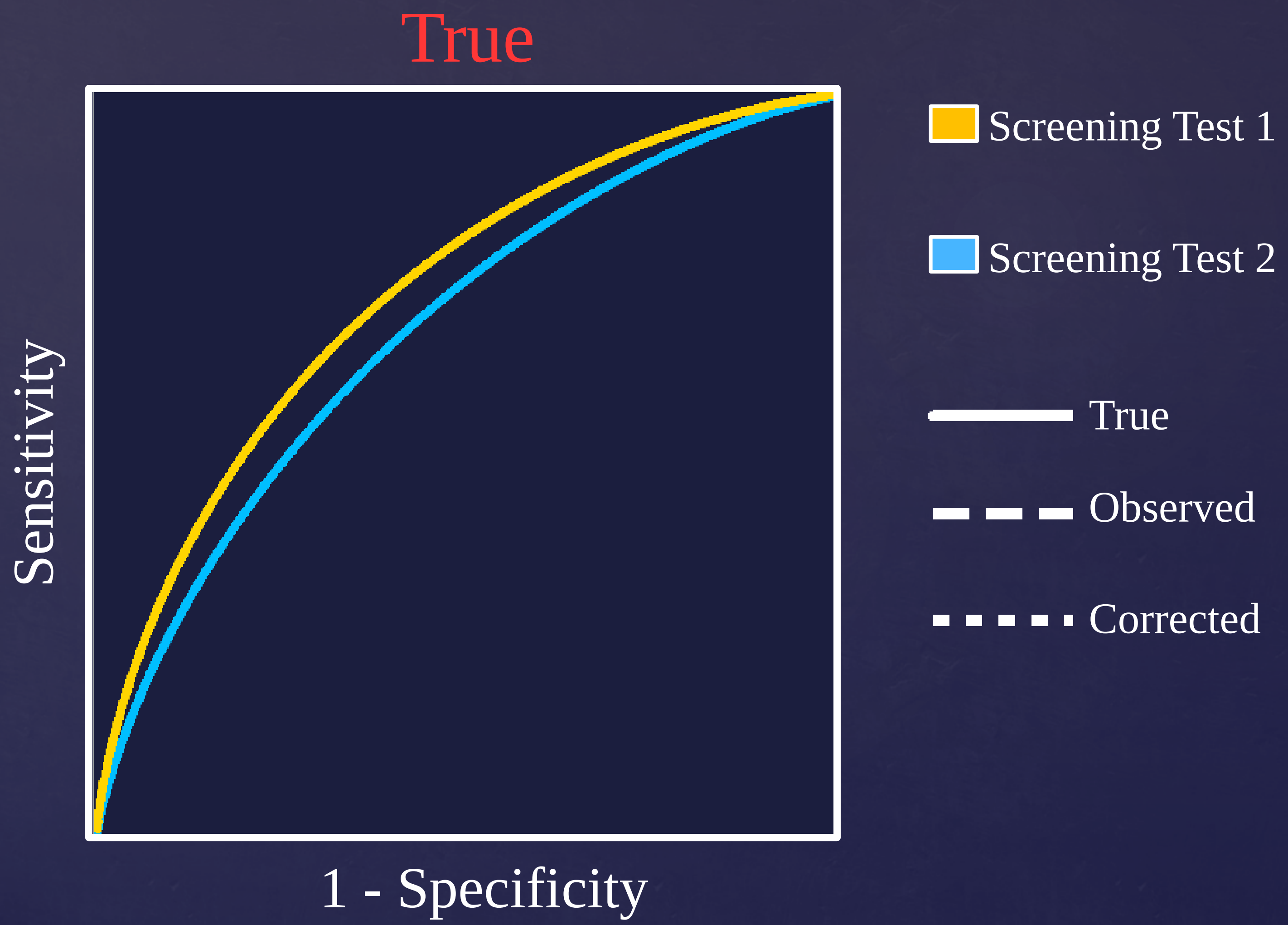
Analysis of Cancer Screening Studies

		Disease Status	
		+	-
Screening Test	+	a	b
	-	c	d





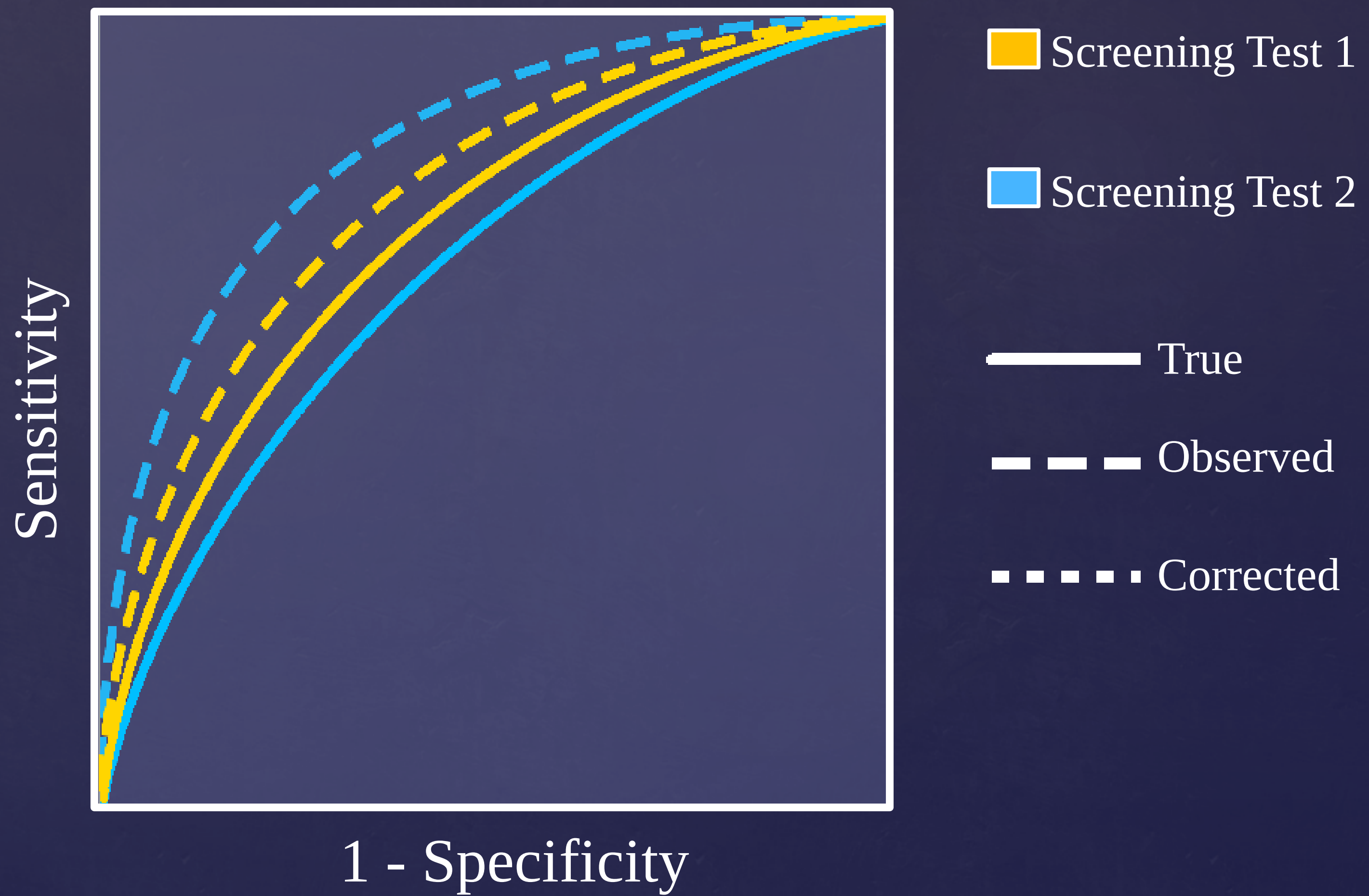
Comparing Two Continuous Screening Tests



(Glueck *et al.*, 2009)

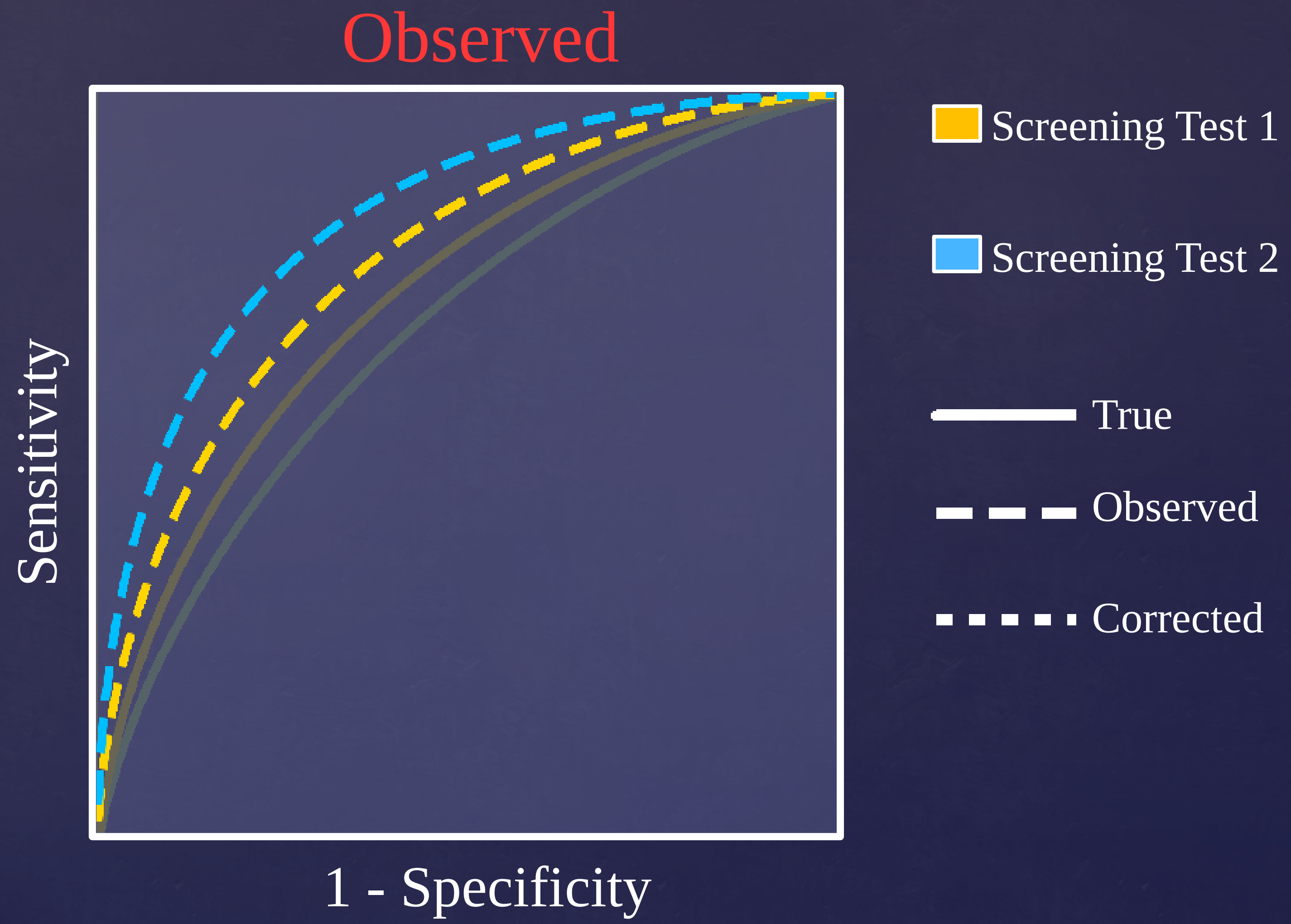
Comparing Two Continuous Screening Tests

Observed



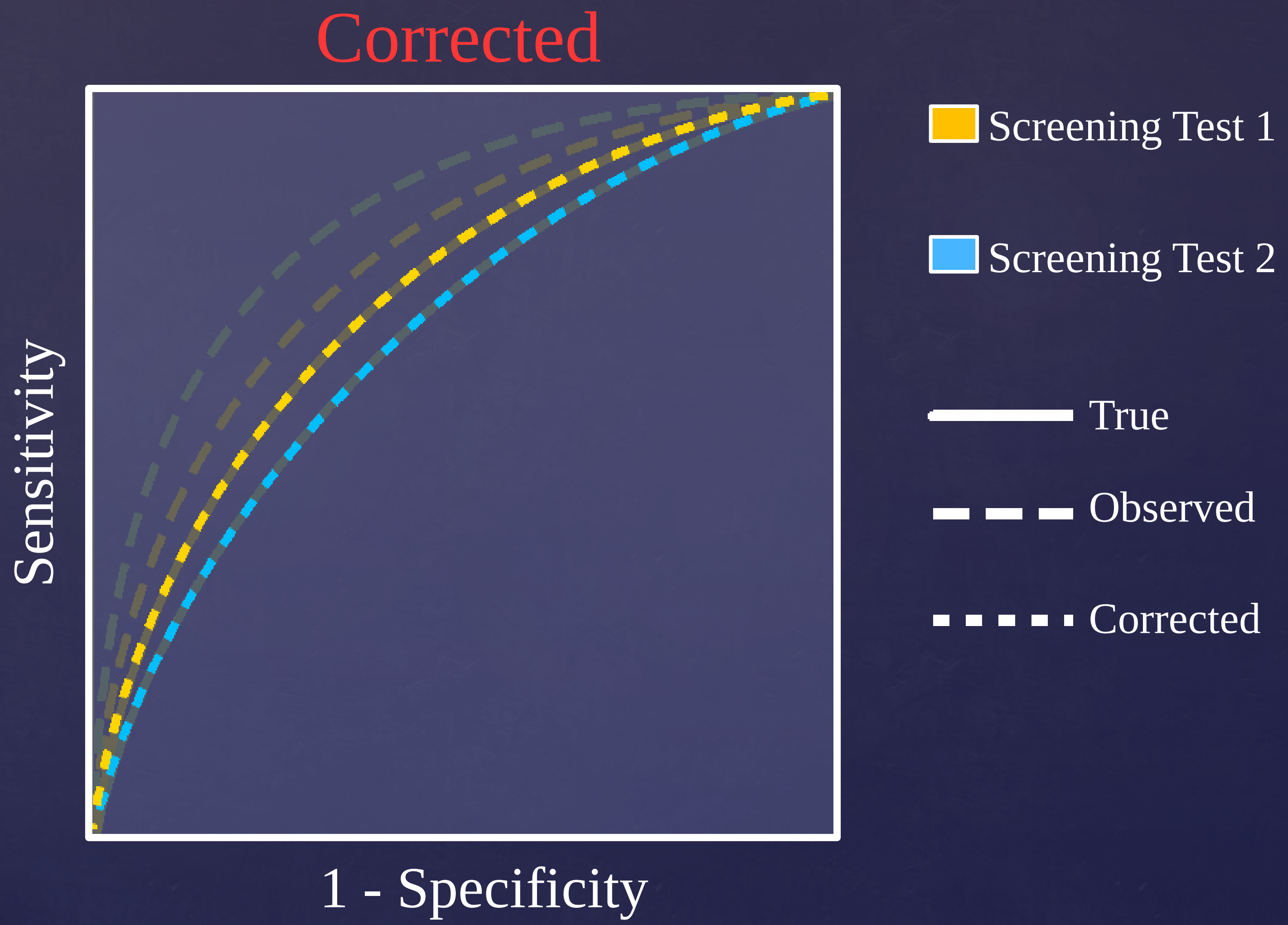
(Glueck *et al.*, 2009)

Comparing Two Continuous Screening Tests



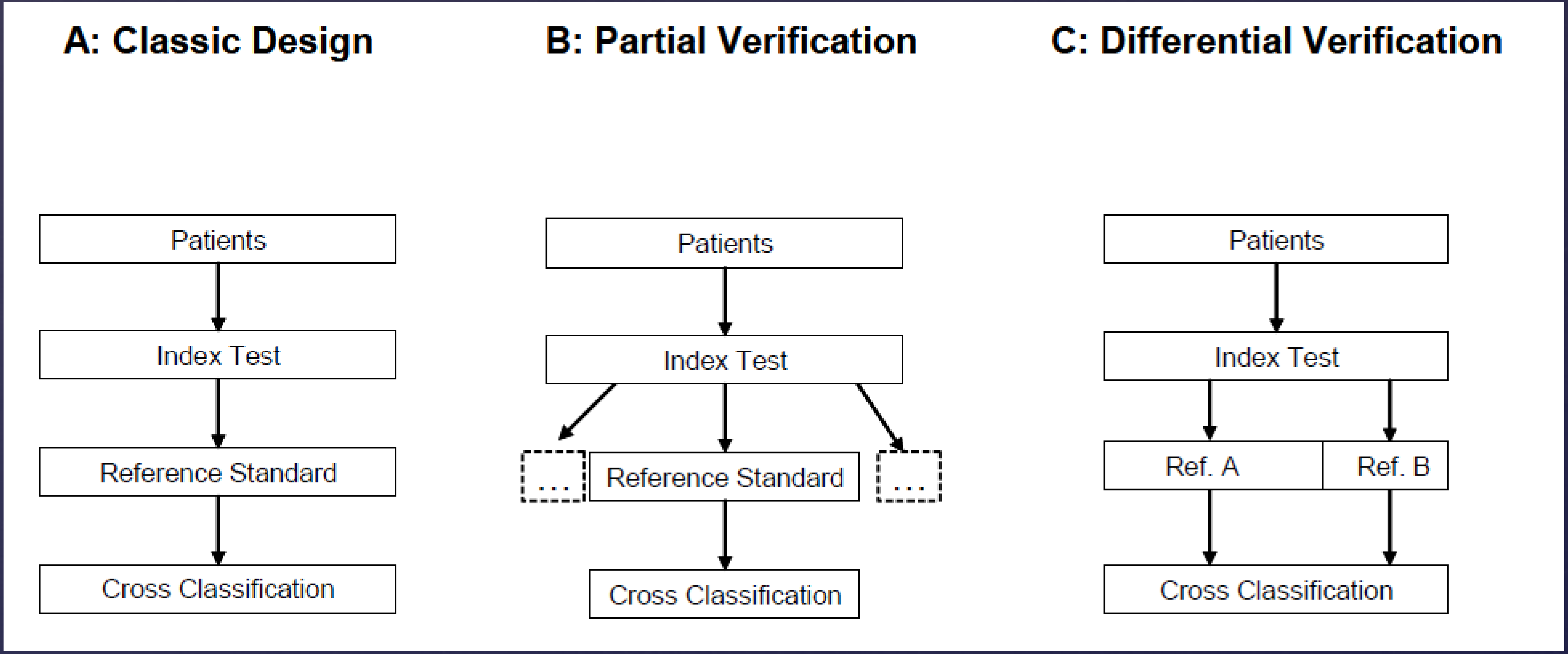
(Glueck *et al.*, 2009)

Comparing Two Continuous Screening Tests



(Glueck *et al.*, 2009)

Types of Bias



(Rutjes *et al.*, 2007)

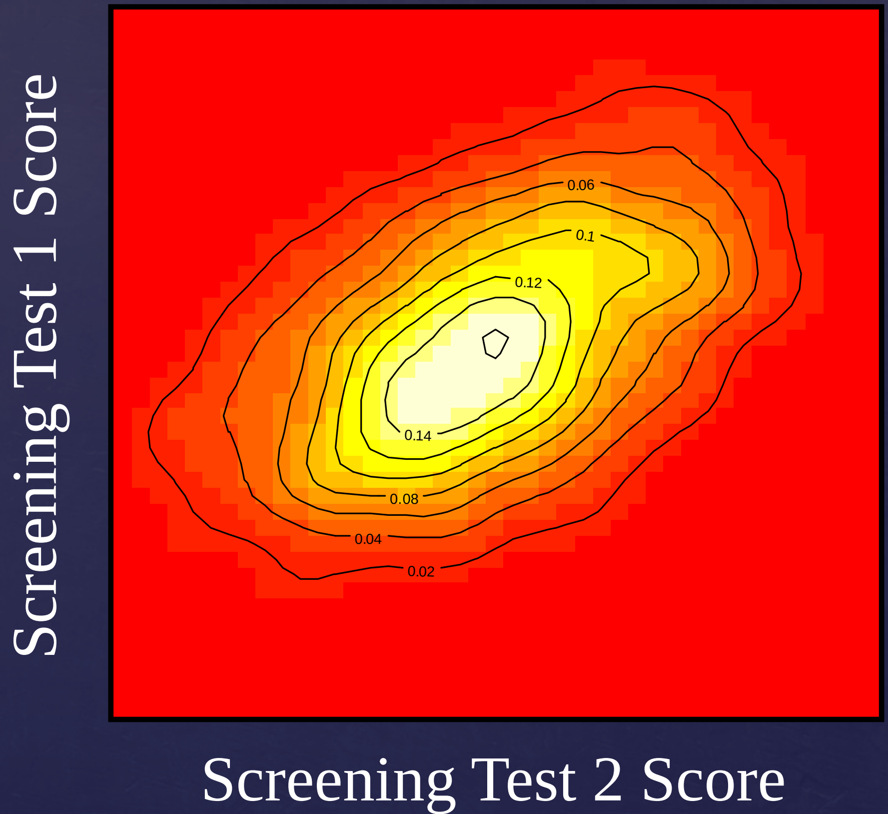
Bias Correction Algorithm

1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.

2. Use the maximum likelihood estimates and the sampling fractions in each partition to calculate weighted estimates.

Bias Correction Algorithm

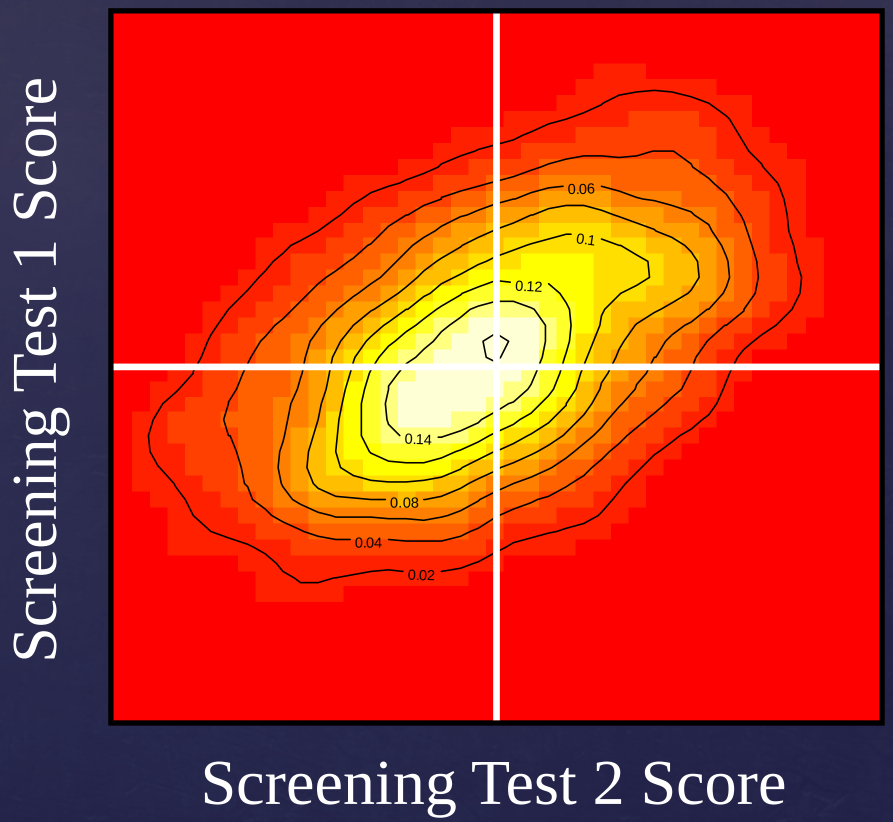
1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.



(Nath, 1971)

Bias Correction Algorithm

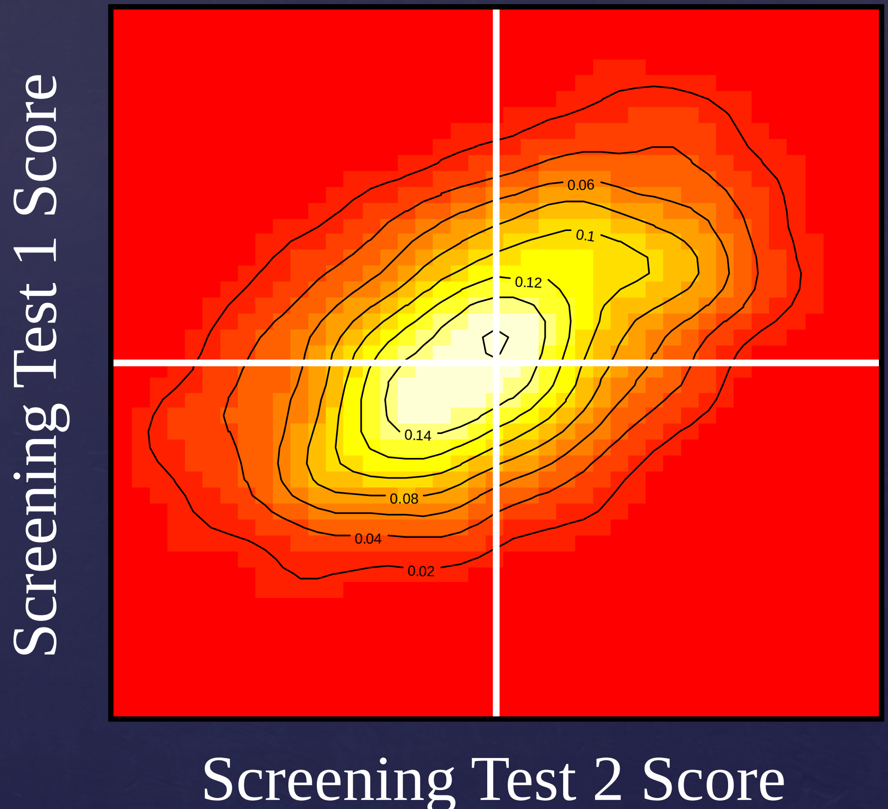
1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.



(Nath, 1971)

Bias Correction Algorithm

1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.



μ_1

μ_2

$\sigma^2{}_1$

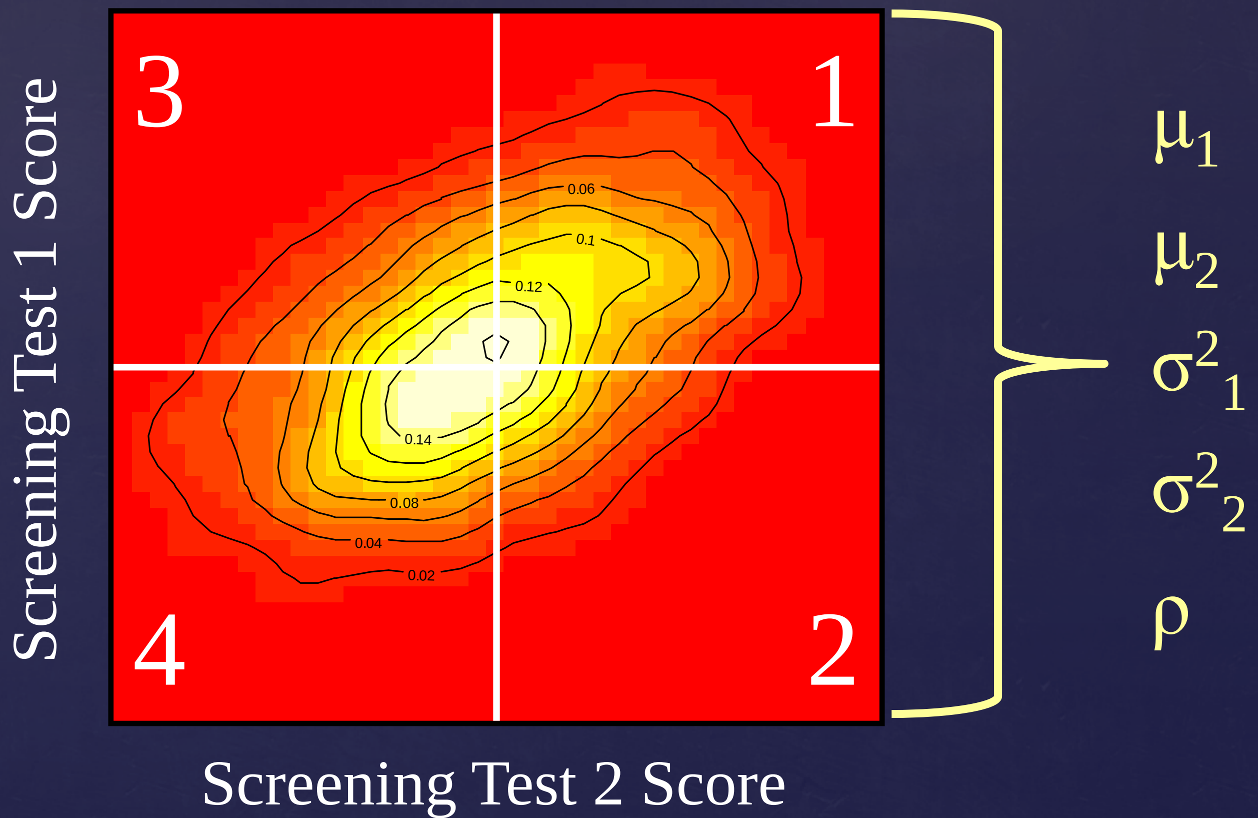
$\sigma^2{}_2$

ρ

(Nath, 1971)

Bias Correction Algorithm

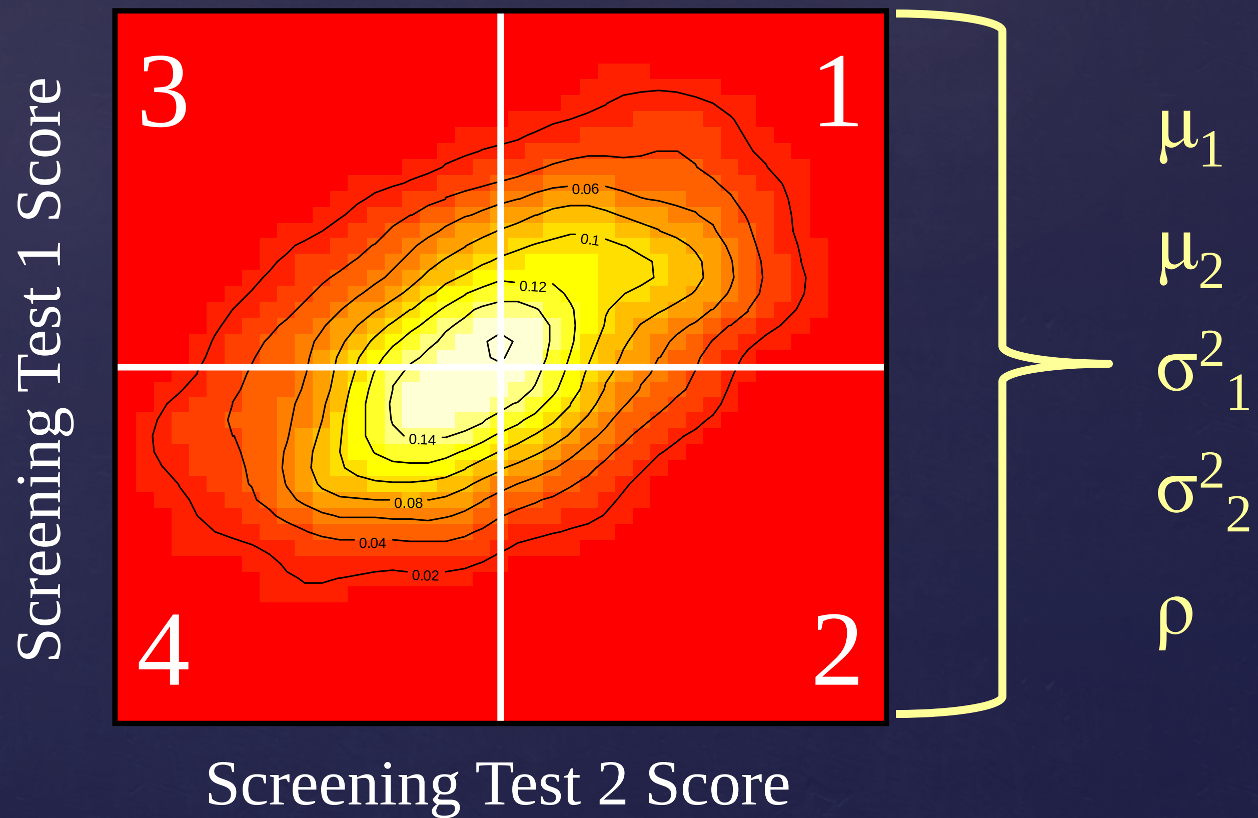
1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.



(Nath, 1971)

Bias Correction Algorithm

1. Find the maximum likelihood estimates of the parameters of the bivariate Gaussian distribution of test scores for the cases.



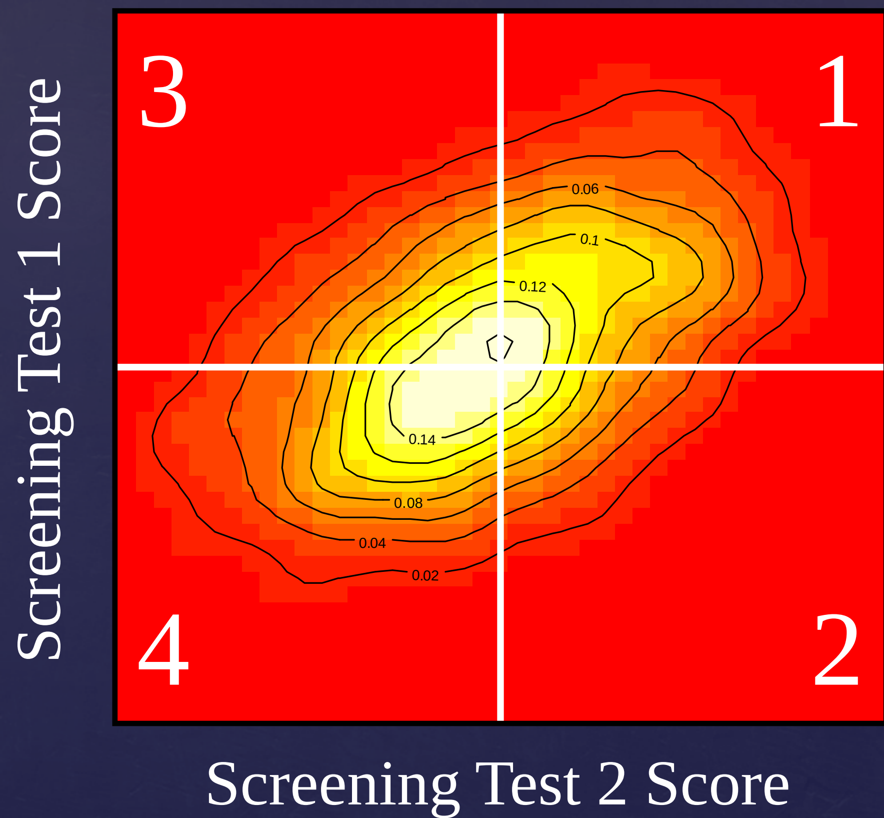
(Nath, 1971)

Bias Correction Algorithm

2. Use the maximum likelihood estimates and the sampling fractions in each partition to calculate weighted estimates.

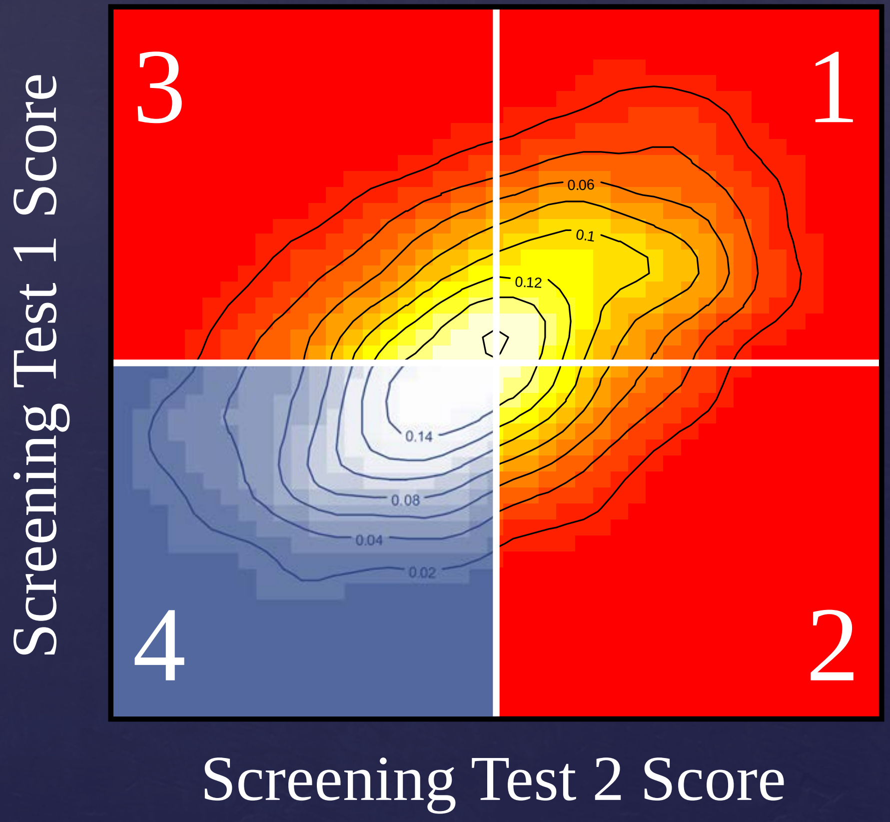
Bias Correction Algorithm

2. Use the maximum likelihood estimates and the sampling fractions in each partition to calculate weighted estimates.



Bias Correction Algorithm

2. Use the maximum likelihood estimates and the sampling fractions in each partition to calculate weighted estimates.



Weighted Estimates

$$\widehat{\lambda} = 1 - \Phi\left(\frac{a_1 - \widehat{\mu}_{11,N}}{\widehat{\sigma}_{11,N}}, \frac{a_2 - \widehat{\mu}_{21,N}}{\widehat{\sigma}_{21,N}}, \widehat{\rho}_{1,N}\right)$$

Weighted Estimates

$$\widehat{\lambda} = 1 - \Phi\left(\frac{a_1 - \widehat{\mu}_{11,N}}{\widehat{\sigma}_{11,N}}, \frac{a_2 - \widehat{\mu}_{21,N}}{\widehat{\sigma}_{21,N}}, \widehat{\rho}_{1,N}\right)$$

Weighted Estimates

$$\widehat{\lambda} = 1 - \Phi\left(\frac{a_1 - \widehat{\mu}_{11,N}}{\widehat{\sigma}_{11,N}}, \frac{a_2 - \widehat{\mu}_{21,N}}{\widehat{\sigma}_{21,N}}, \widehat{\rho}_{1,N}\right)$$

Weighted Estimates

$$\widehat{\lambda} = 1 - \Phi\left(\frac{a_1 - \widehat{\mu}_{11,N}}{\widehat{\sigma}_{11,N}}, \frac{a_2 - \widehat{\mu}_{21,N}}{\widehat{\sigma}_{21,N}}, \widehat{\rho}_{1,N}\right)$$

Weighted Estimates

$$\widehat{\mu}_{11,W} = \widehat{\lambda}\bar{X}_{11,A} + \left(1-\widehat{\lambda}\right)\bar{X}_{11,B}$$

$$\widehat{\mu}_{21,W} = \widehat{\lambda}\bar{X}_{21,A} + \left(1-\widehat{\lambda}\right)\bar{X}_{21,B}$$

$$\widehat{\sigma}^2_{11,W} = \mathcal{G}_1 + \mathcal{H}_1 - \widehat{\mu}^2_{11,W}$$

$$\widehat{\sigma}^2_{21,W} = \mathcal{G}_2 + \mathcal{H}_2 - \widehat{\mu}^2_{21,W}$$

$$\widehat{\rho}_{1,W} = \widehat{\sigma}^2_{11,W}\widehat{\sigma}^2_{21,W}\left(\mathcal{P} + \mathcal{Q} - \widehat{\mu}_{11,W}\widehat{\mu}_{21,W}\right)$$

(Equation 3.3.1, p. 81, Kish, 1995; Proposition 5.2, p. 348, Ross, 2009)

Weighted Estimates

$$\widehat{\mu}_{11,W} = \widehat{\lambda}\overline{X}_{11,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{11,B}$$

$$\widehat{\mu}_{21,W} = \widehat{\lambda}\overline{X}_{21,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{21,B}$$

$$\widehat{\sigma}^2_{11,W} = \mathcal{G}_1 + \mathcal{H}_1 - \widehat{\mu}^2_{11,W}$$

$$\widehat{\sigma}^2_{21,W} = \mathcal{G}_2 + \mathcal{H}_2 - \widehat{\mu}^2_{21,W}$$

$$\widehat{\rho}_{1,W} = \widehat{\sigma}^2_{11,W}\widehat{\sigma}^2_{21,W}\left(\mathcal{P} + \mathcal{Q} - \widehat{\mu}_{11,W}\widehat{\mu}_{21,W}\right)$$

(Equation 3.3.1, p. 81, Kish, 1995; Proposition 5.2, p. 348, Ross, 2009)

Weighted Estimates

$$\widehat{\mu}_{11,W} = \widehat{\lambda}\overline{X}_{11,A} + \left(1-\widehat{\lambda}\right)\overline{X}_{11,B}$$

$$\widehat{\mu}_{21,W} = \widehat{\lambda}\overline{X}_{21,A} + \left(1-\widehat{\lambda}\right)\overline{X}_{21,B}$$

$$\widehat{\sigma}^2_{11,W} = \mathcal{G}_1 + \mathcal{H}_1 - \widehat{\mu}^2_{11,W}$$

$$\widehat{\sigma}^2_{21,W} = \mathcal{G}_2 + \mathcal{H}_2 - \widehat{\mu}^2_{21,W}$$

$$\widehat{\rho}_{1,W} = \widehat{\sigma}^2_{11,W}\widehat{\sigma}^2_{21,W}\left(\mathcal{P} + \mathcal{Q} - \widehat{\mu}_{11,W}\widehat{\mu}_{21,W}\right)$$

(Equation 3.3.1, p. 81, Kish, 1995; Proposition 5.2, p. 348, Ross, 2009)

Weighted Estimates

$$\widehat{\mu}_{11,W} = \widehat{\lambda}\overline{X}_{11,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{11,B}$$

$$\widehat{\mu}_{21,W} = \widehat{\lambda}\overline{X}_{21,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{21,B}$$

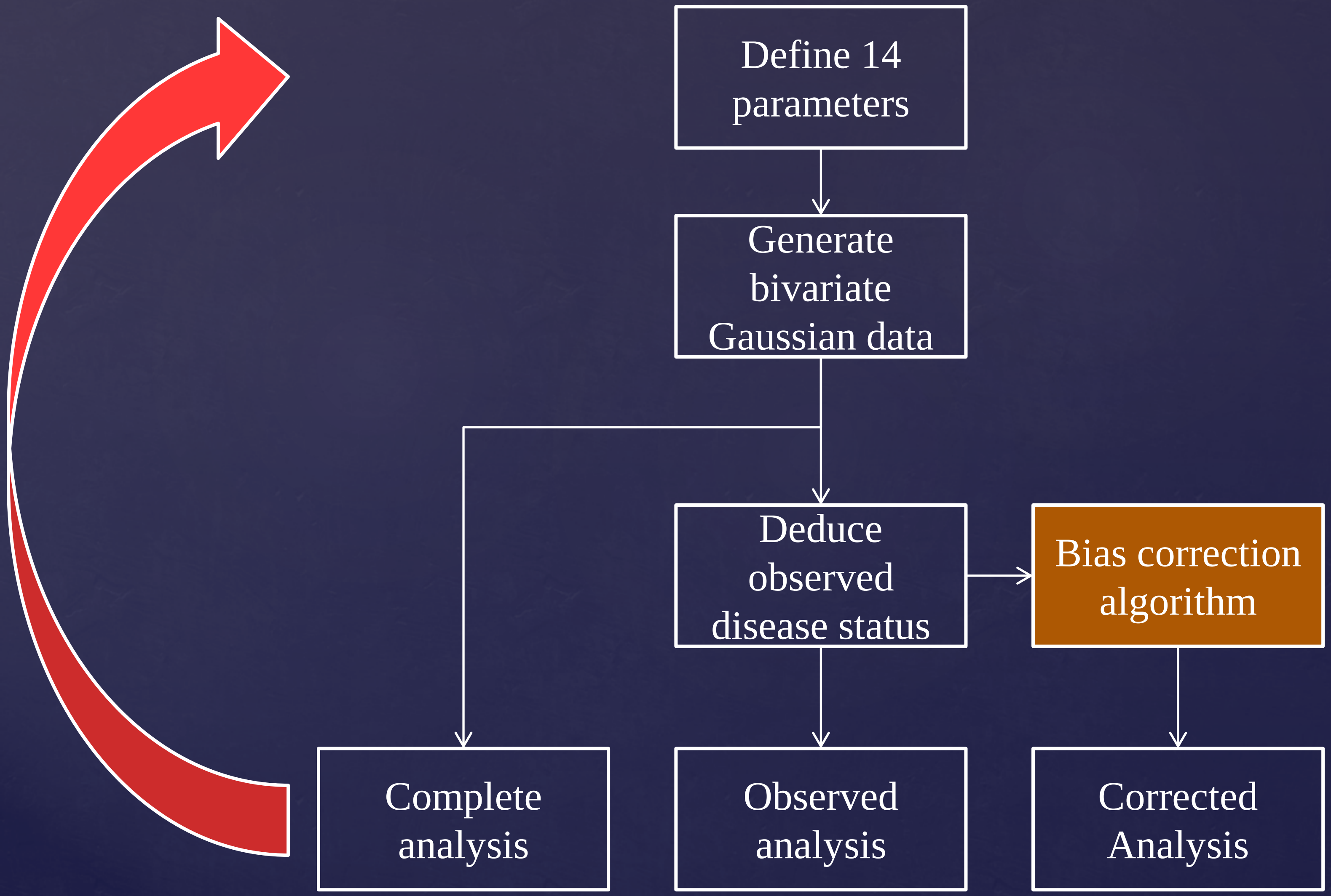
$$\widehat{\sigma}^2_{11,W} = \mathcal{G}_1 + \mathcal{H}_1 - \widehat{\mu}^2_{11,W}$$

$$\widehat{\sigma}^2_{21,W} = \mathcal{G}_2 + \mathcal{H}_2 - \widehat{\mu}^2_{21,W}$$

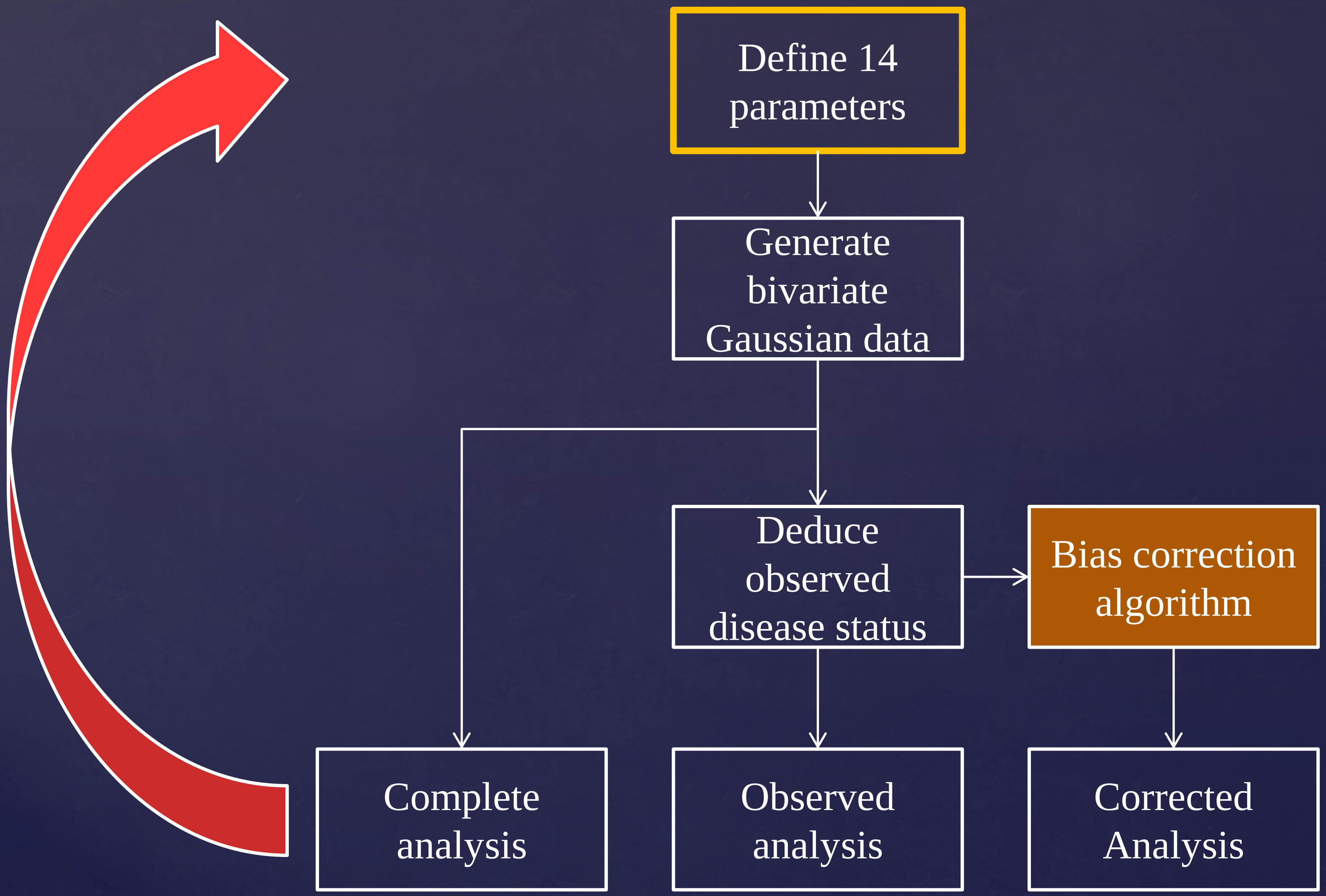
$$\widehat{\rho}_{1,W} = \widehat{\sigma}^2_{11,W}\widehat{\sigma}^2_{21,W}\left(\mathcal{P} + \mathcal{Q} - \widehat{\mu}_{11,W}\widehat{\mu}_{21,W}\right)$$

(Equation 3.3.1, p. 81, Kish, 1995; Proposition 5.2, p. 348, Ross, 2009)

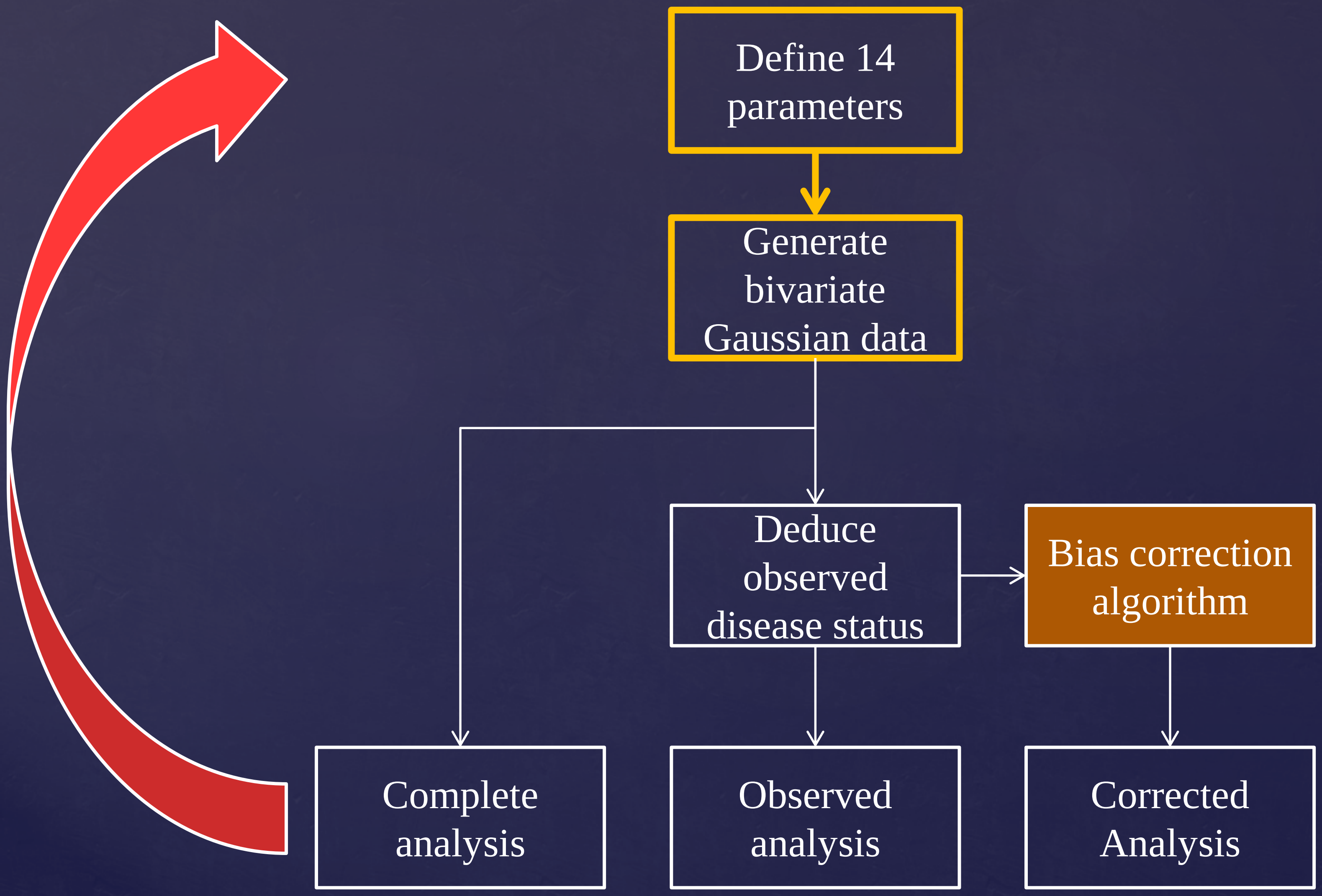
Design for Simulation Studies



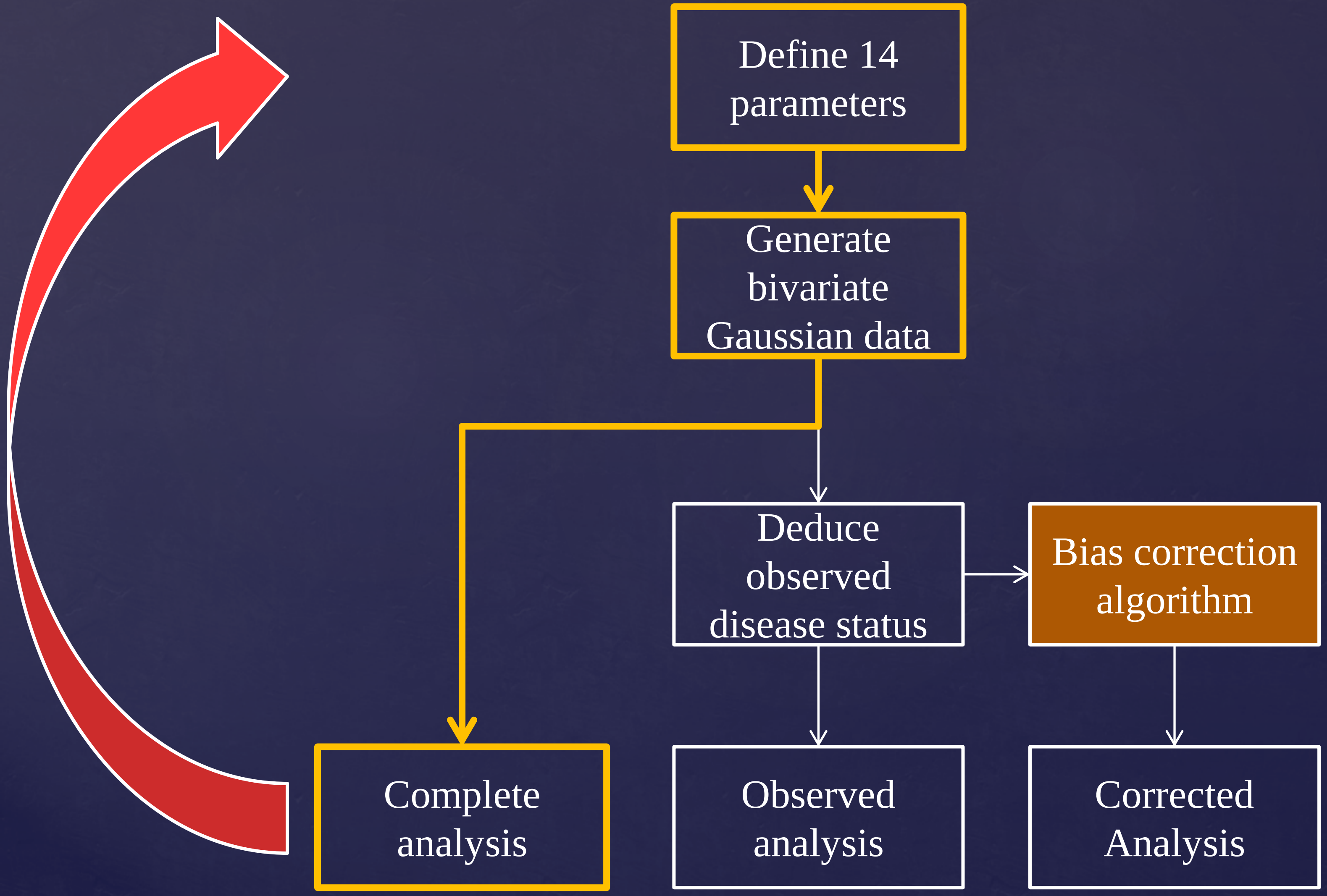
Design for Simulation Studies



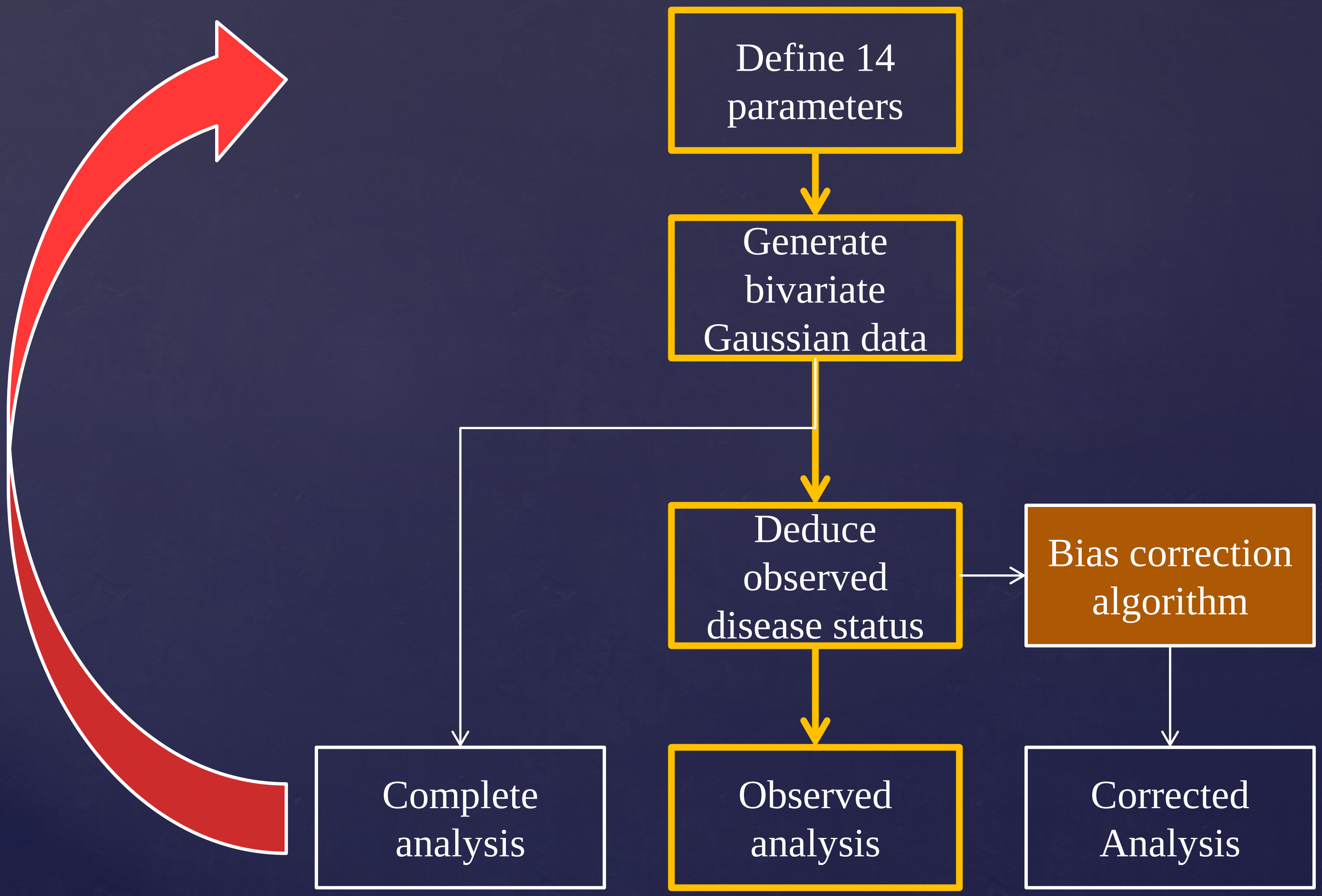
Design for Simulation Studies



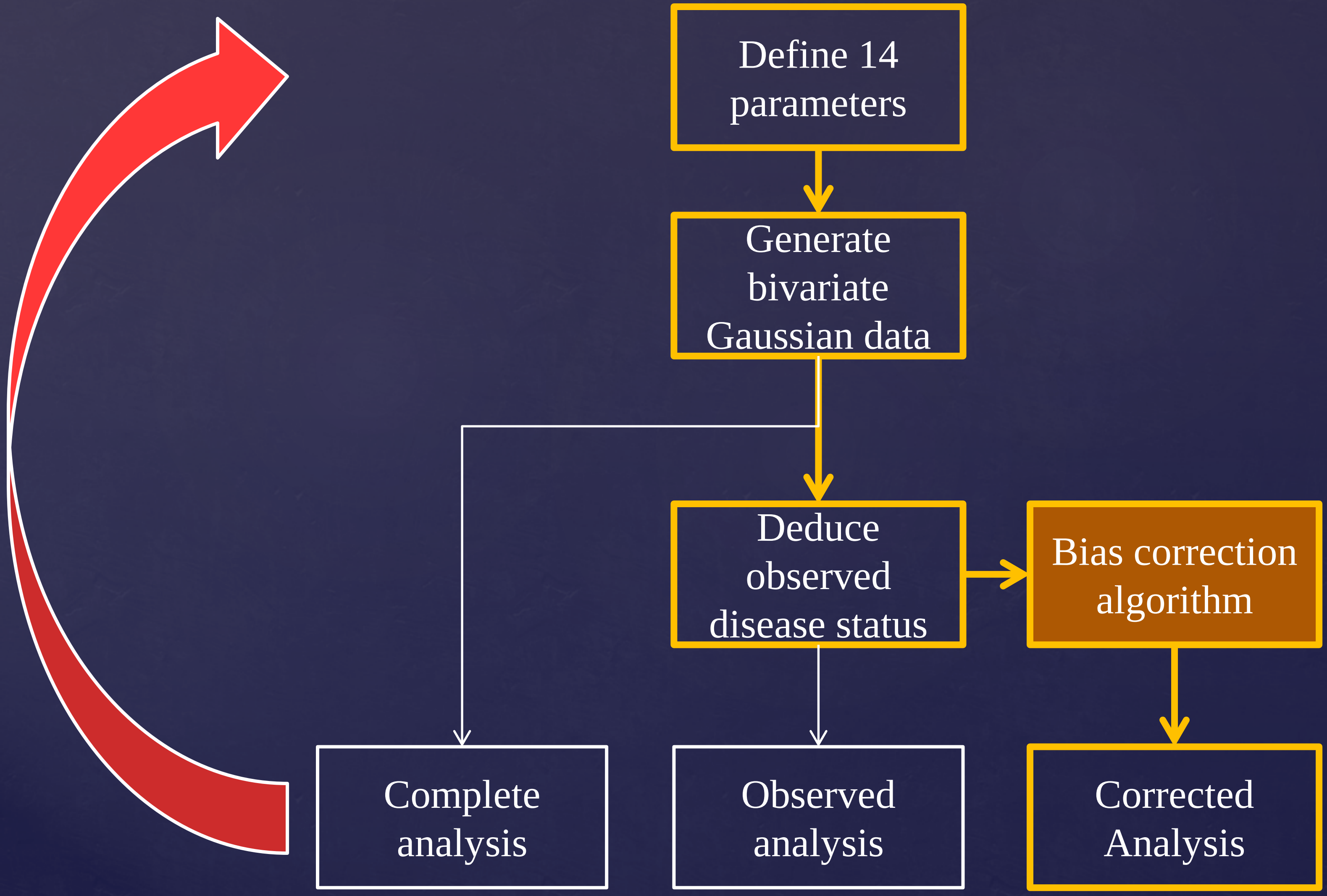
Design for Simulation Studies



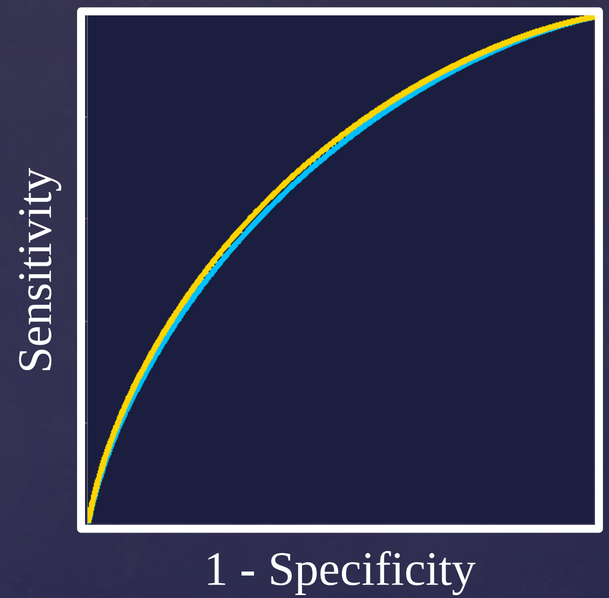
Design for Simulation Studies

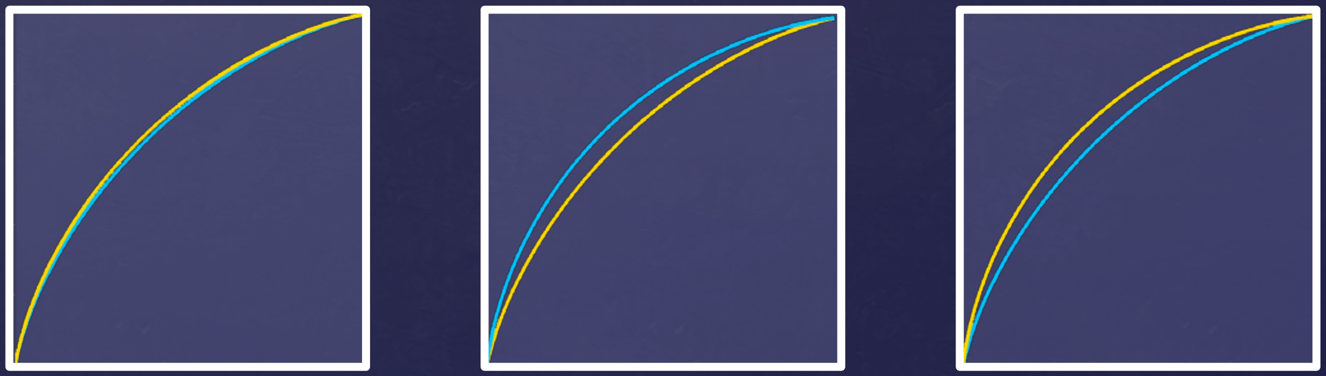


Design for Simulation Studies

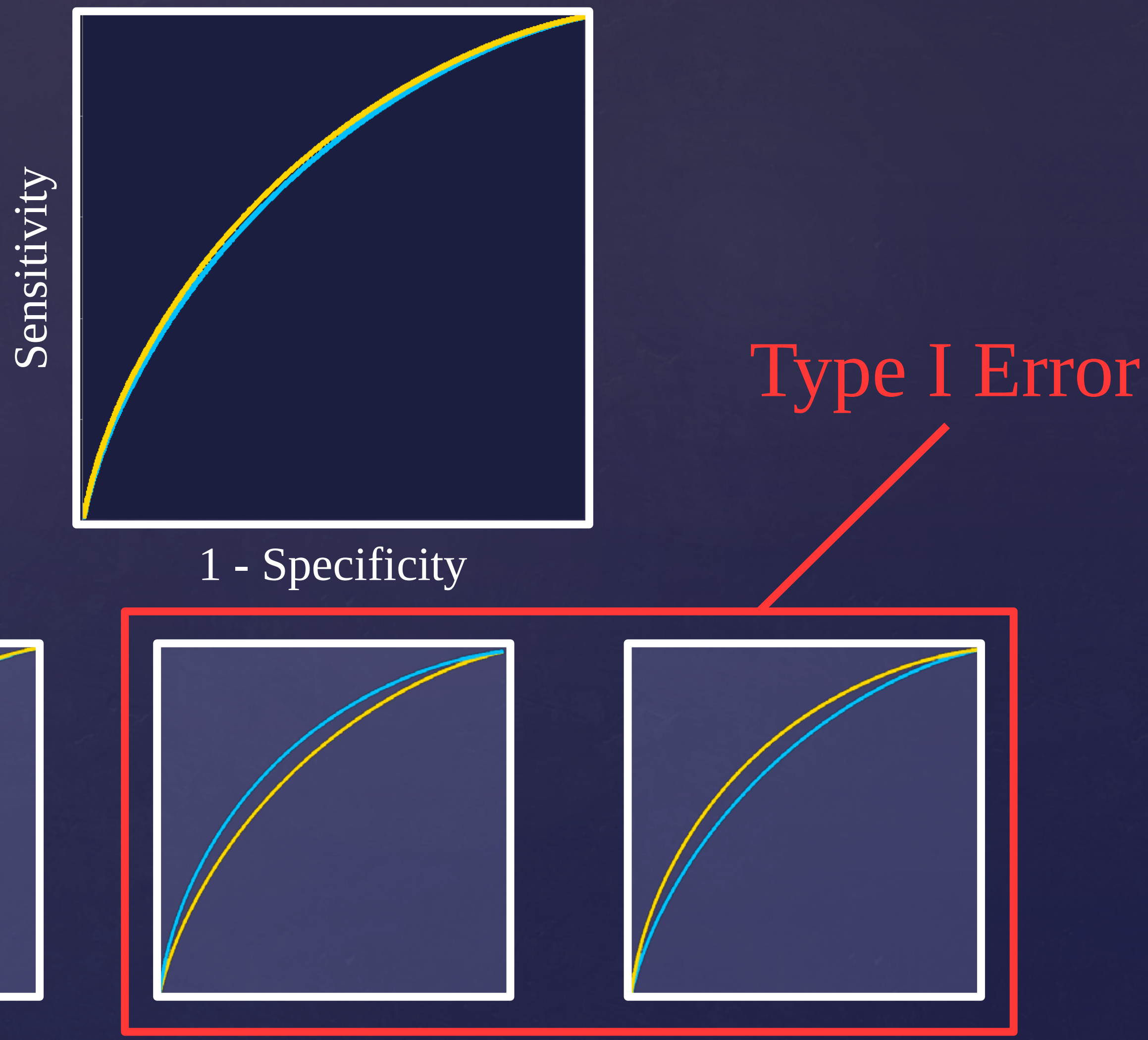


Evaluation of the Method

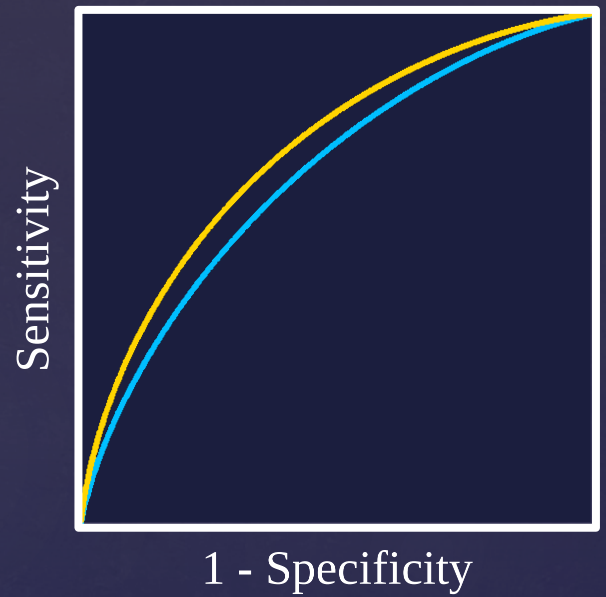


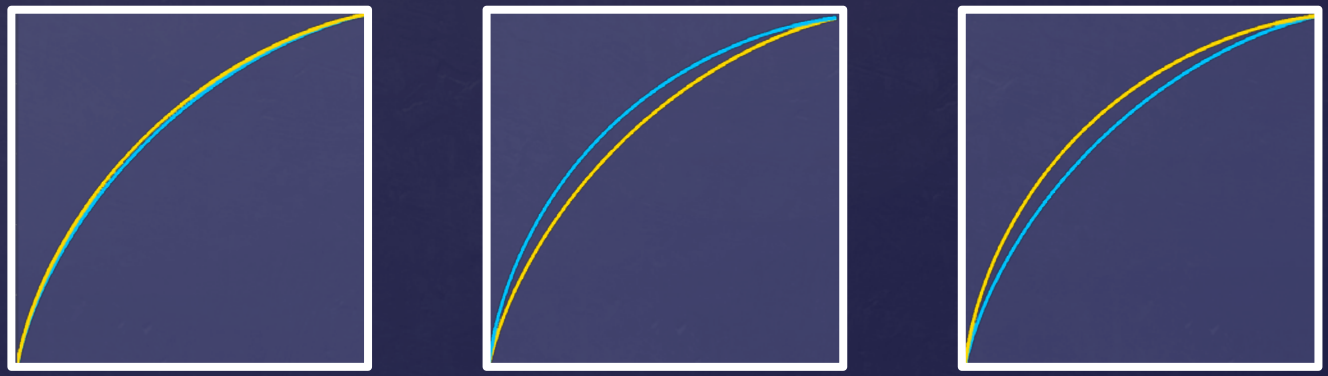


Evaluation of the Method

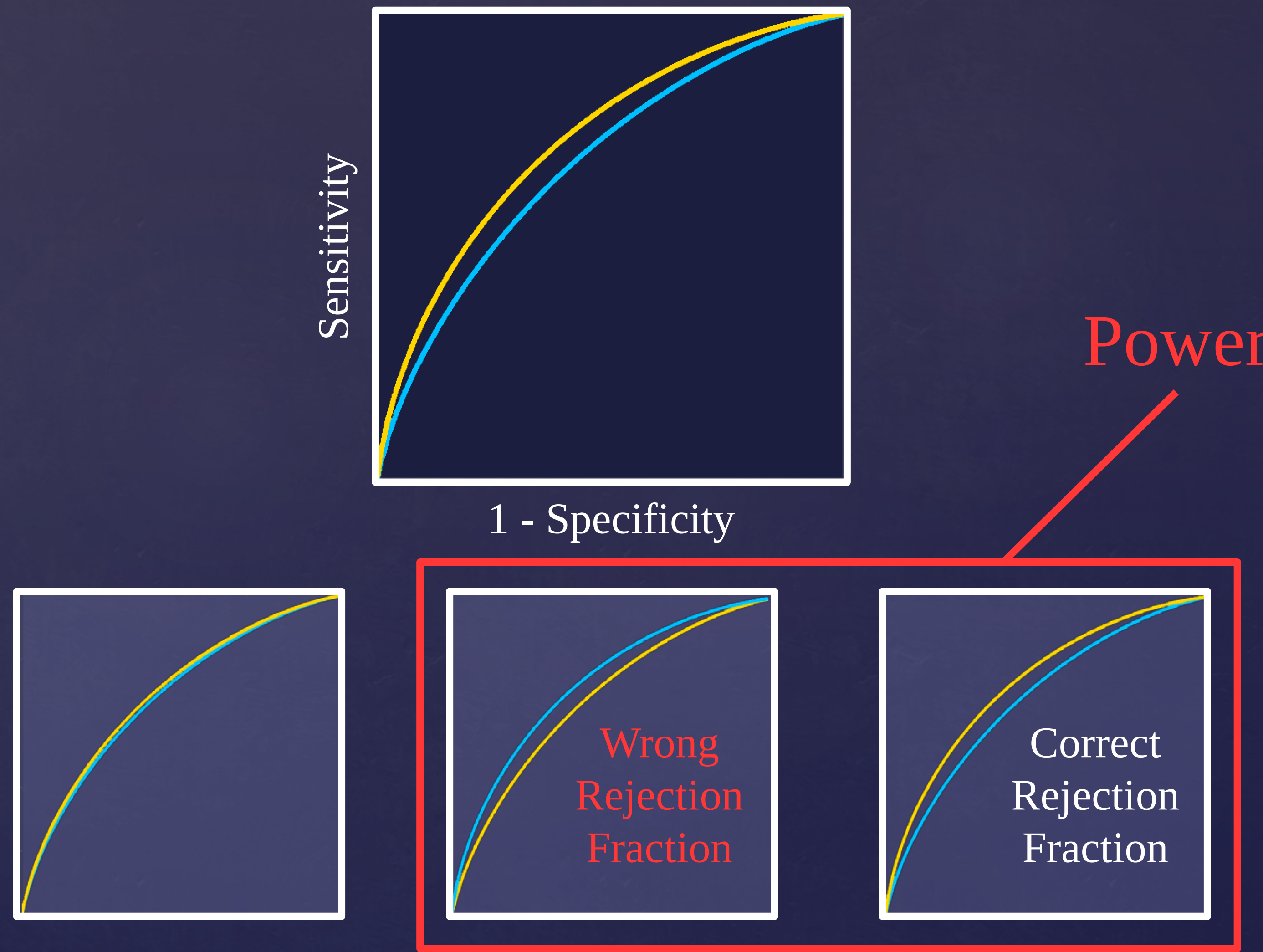


Evaluation of the Method

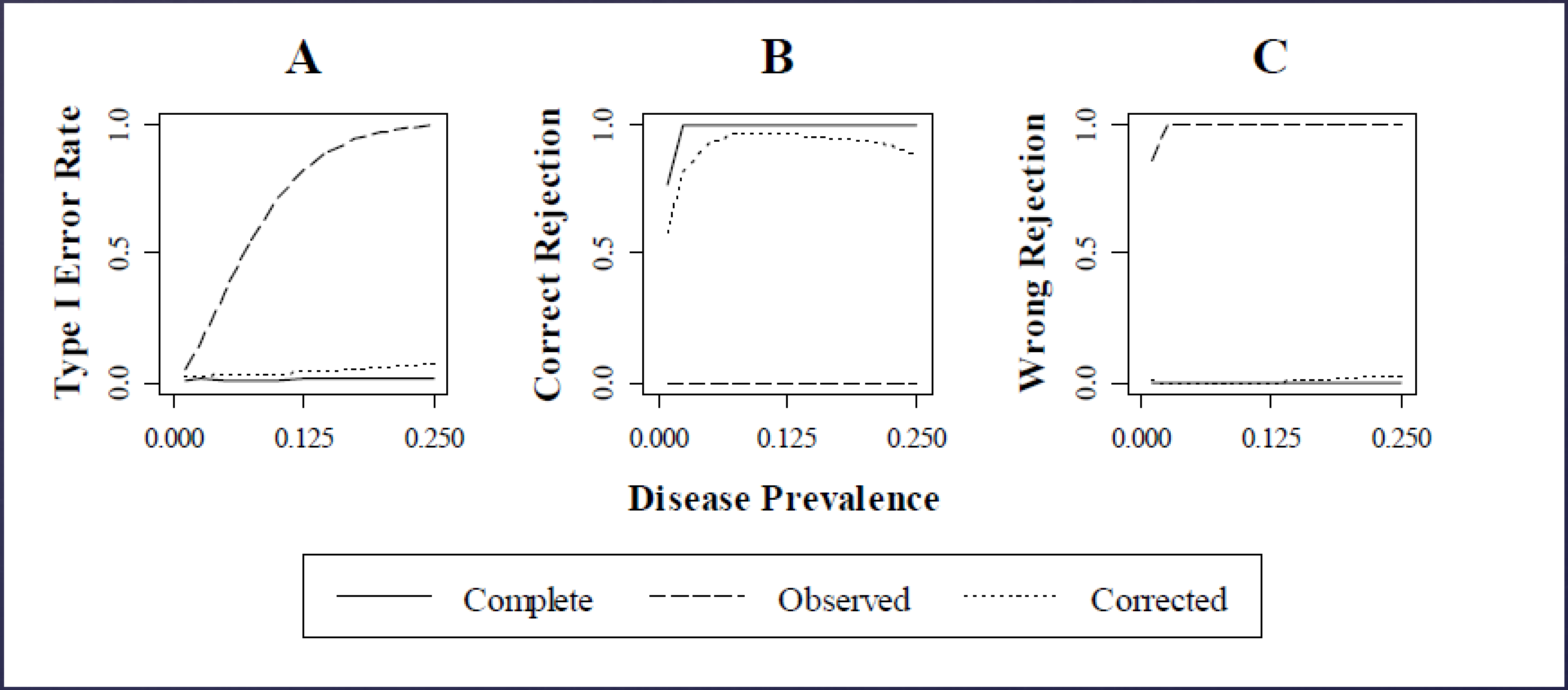




Evaluation of the Method



Disease Prevalence Simulation Study



Percent Identified Simulation Study

Bias?	Percent Identified (Test 1 / Test 2)	Complete Type I Error	Observed Type I Error	Corrected Type I Error
Yes	15/50	0.01	0.89	0.36
	15/80	0.02	0.95	0.25
	50/80	0.01	0.23	0.12
No	15/15	0.01	0.02	0.23
	50/50	0.01	0.02	0.12
	80/80	0.02	0.02	0.18

Percent Identified Simulation Study

Bias?	Percent Identified (Test 1 / Test 2)	Complete Type I Error	Observed Type I Error	Corrected Type I Error
Yes	15/50	0.01	0.89	0.36
	15/80	0.02	0.95	0.25
	50/80	0.01	0.23	0.12
No	15/15	0.01	0.02	0.23
	50/50	0.01	0.02	0.12
	80/80	0.02	0.02	0.18

Percent Identified Simulation Study

Bias?	Percent Identified (Test 1 / Test 2)	Complete Type I Error	Observed Type I Error	Corrected Type I Error
Yes	15/50	0.01	0.89	0.36
	15/80	0.02	0.95	0.25
	50/80	0.01	0.23	0.12
No	15/15	0.01	0.02	0.23
	50/50	0.01	0.02	0.12
	80/80	0.02	0.02	0.18

Recommendation

- Study investigators should conduct a simulation of their study using both the standard analysis and the bias correction method.
- Study investigators should choose the analysis plan that has a nominal Type I error rate and the highest power for the correct decision.

Oral Cancer Screening Demonstration

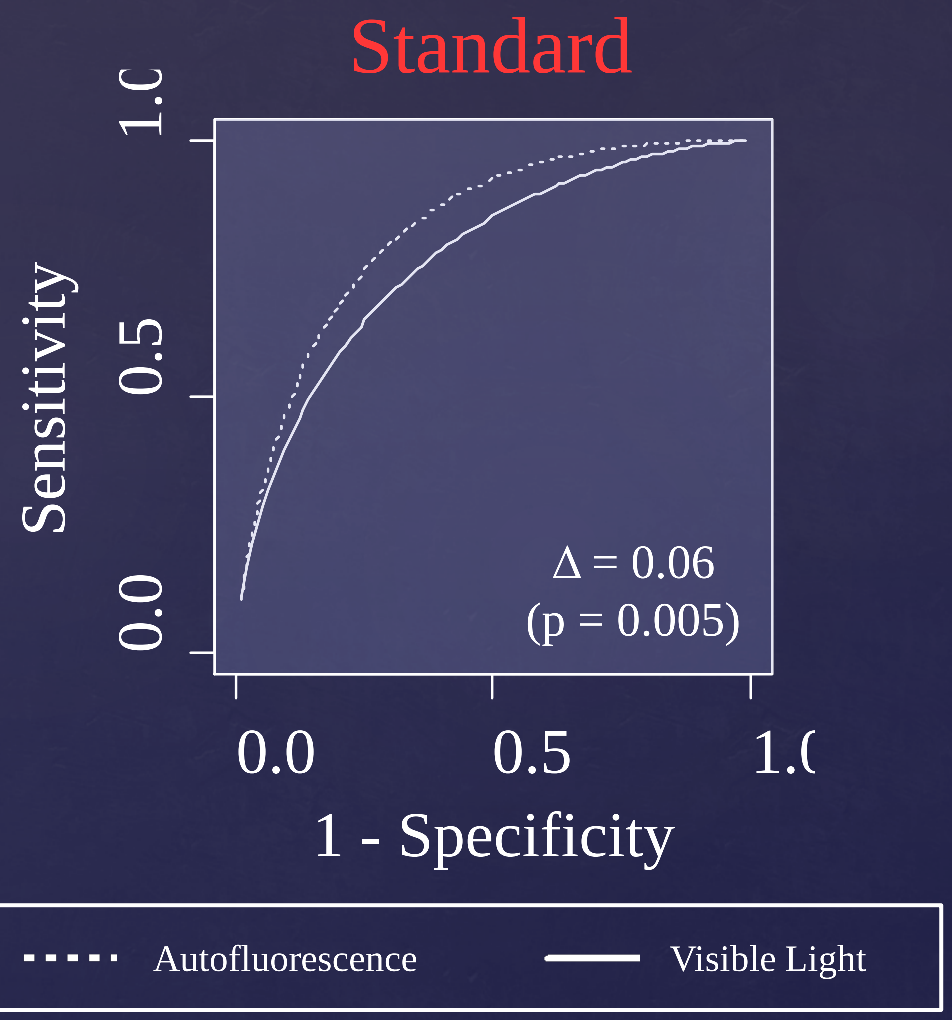
VISIBLE LIGHT

No visible lesion

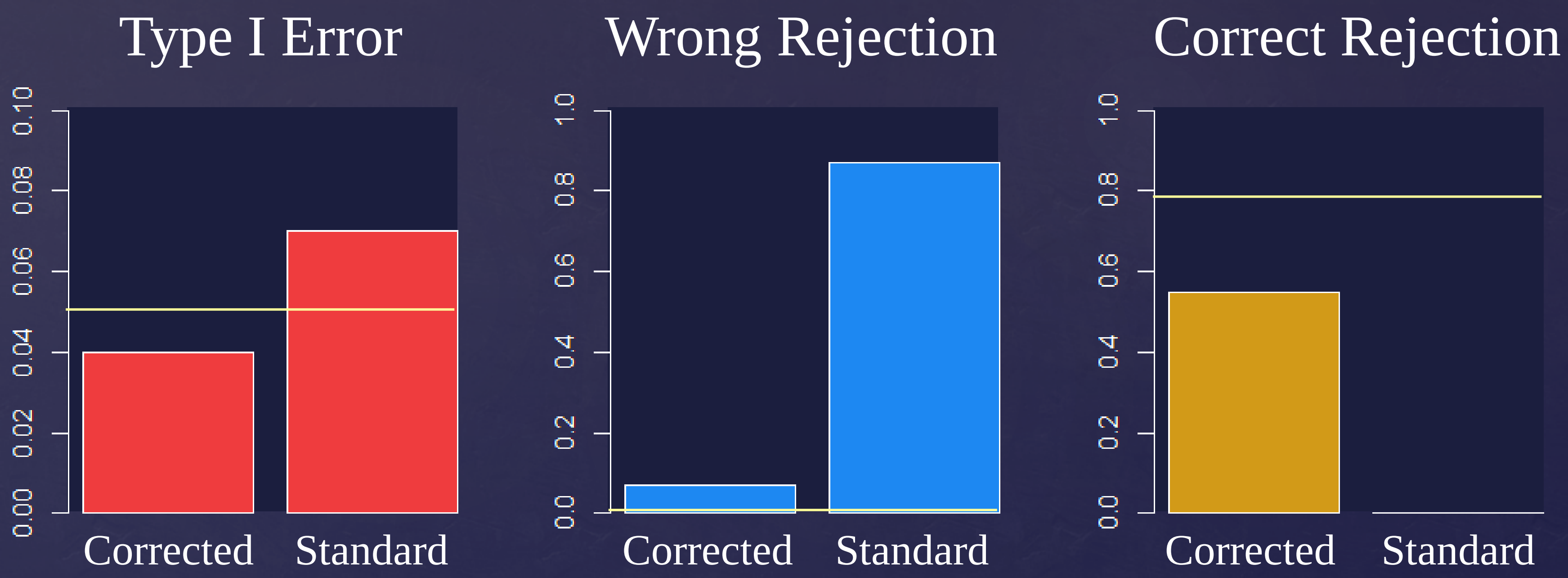
AUTOFLUORESCENCE

Dark region confirmed to be carcinoma in situ

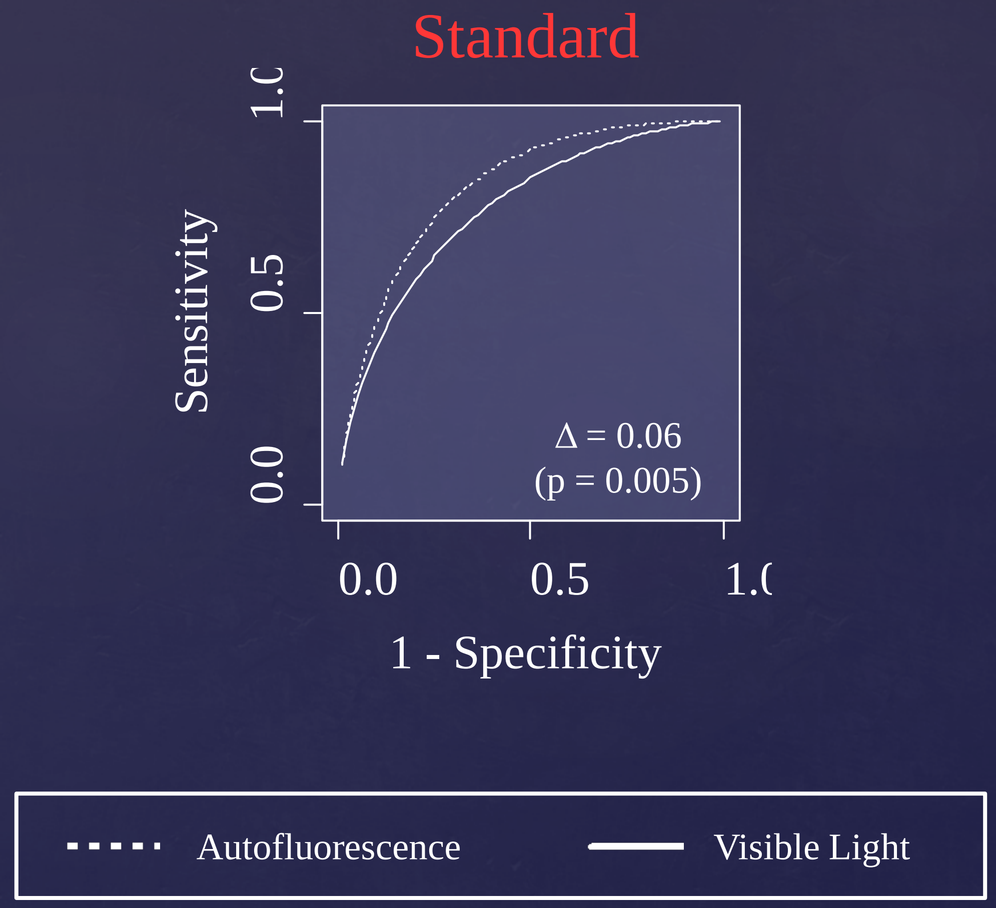
Oral Cancer Screening Analysis



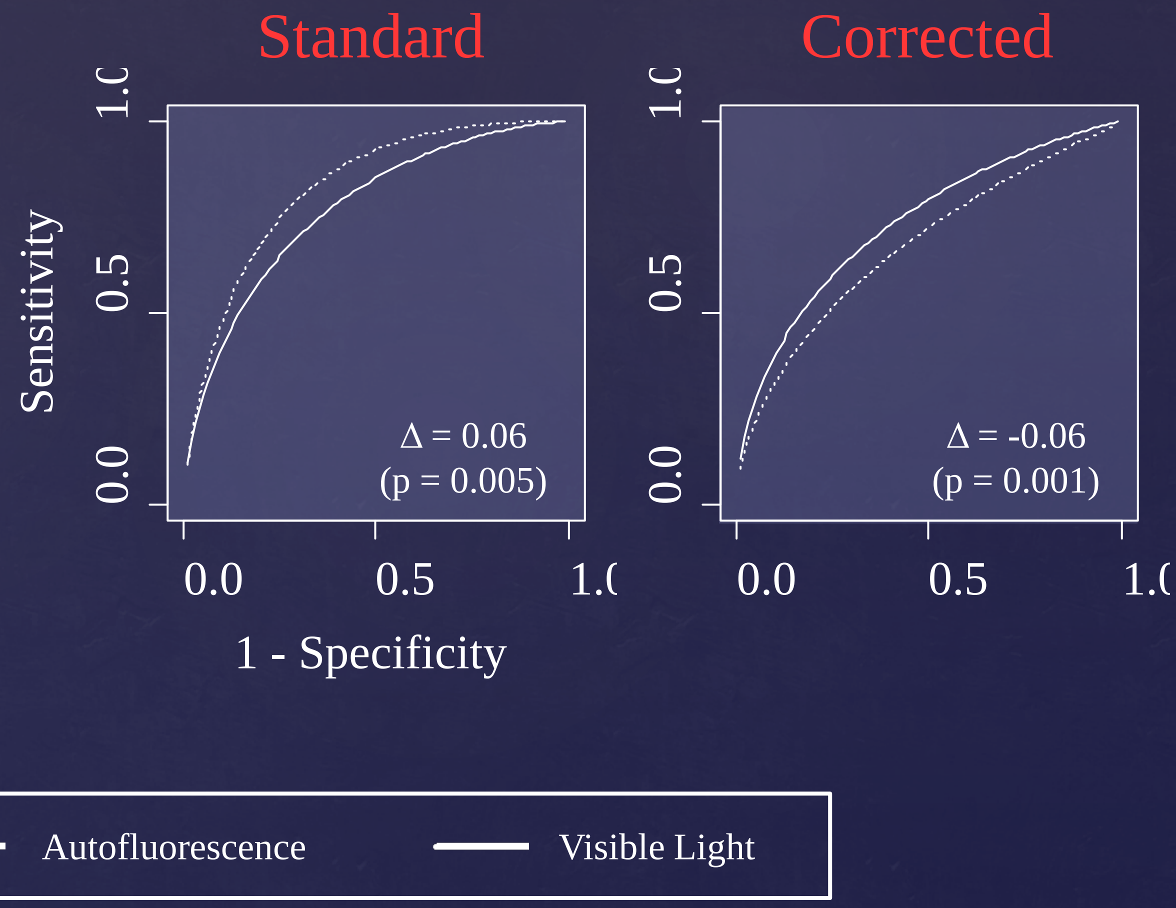
Decision Errors Simulation



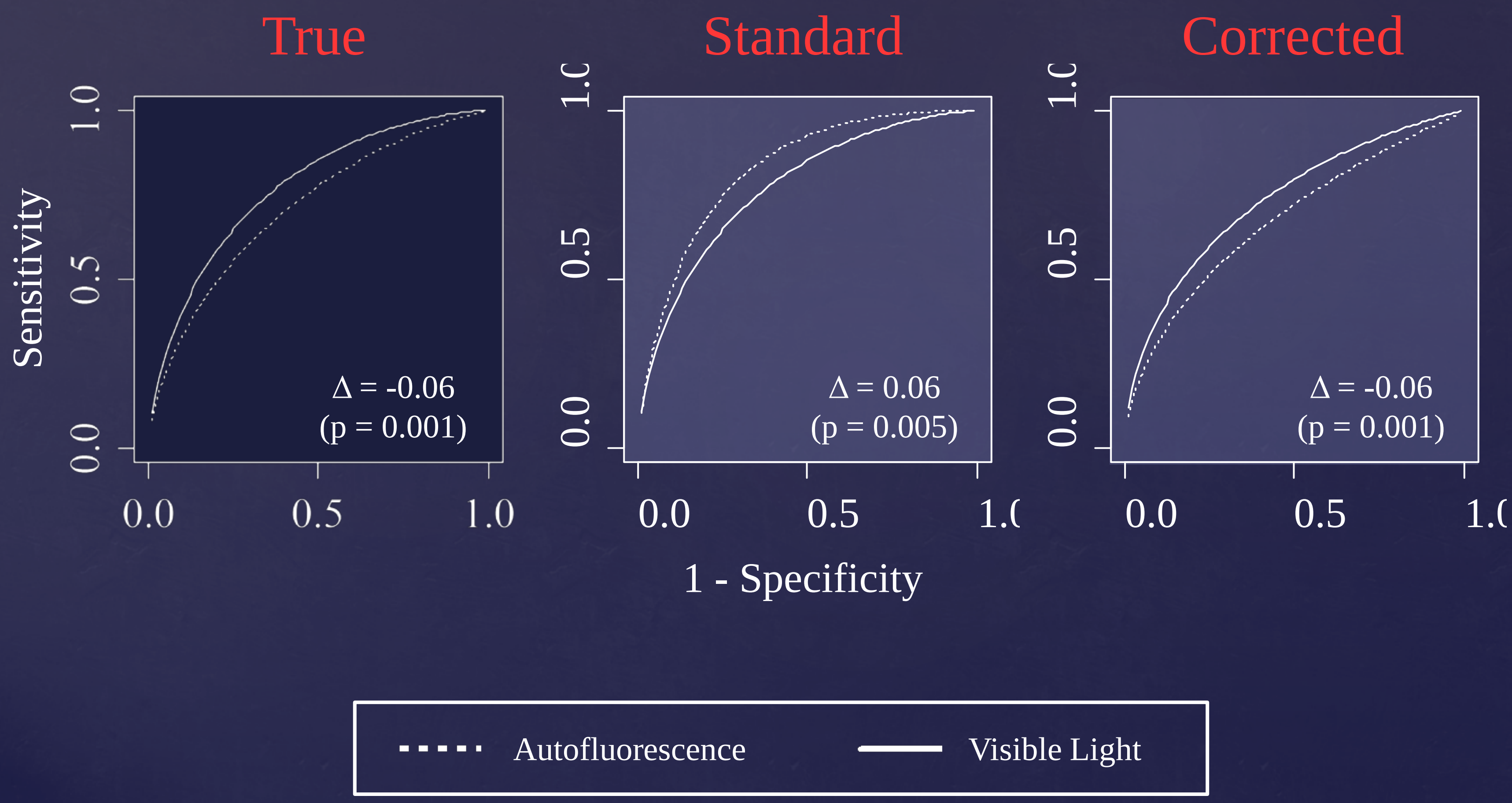
Oral Cancer Screening Analysis



Oral Cancer Screening Analysis



Oral Cancer Screening Analysis



References

Glueck, D. H., Lamb, M. M., O'Donnell, C. I., Ringham, B. M., Brinton, J. T., Muller, K. E., Lewin, J. M., et al. (2009). Bias in trials comparing paired continuous tests can cause researchers to choose the wrong screening modality. BMC Medical Research Methodology, 9, 4. doi:10.1186/1471-2288-9-4Kish, L. (1995). Survey Sampling. Wiley-Interscience.

Kish, L. (1995). Survey Sampling. Wiley-Interscience.

Lewin, J. M., D'Orsi, C. J., Hendrick, R. E., Moss, L. J., Isaacs, P. K., Karellas, A., & Cutter, G. R. (2002). Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. AJR. American Journal of Roentgenology, 179(3), 671-677.

Nath, G. B. (1971). Estimation in Truncated Bivariate Normal Distributions. Journal of the Royal Statistical Society. Series C (Applied Statistics), 20(3), 313-319. doi:10.2307/2346762

Obuchowski, N. A., & McClish, D. K. (1997). Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. Statistics in Medicine, 16(13), 1529-1542.

Ross, S. (2009). First Course in Probability, A (8th ed.). Prentice Hall.

Rutjes A.W.S., Reitsma J.B., Coomarasamy A., Khan K.S., Bossuyt P.M.M. (2007). Evaluation of diagnostic tests when there is no gold standard. A review of methods. Health Technol. Assess., 11(50).

Zhou, X.-H., McClish, D. K., & Obuchowski, N. A. (2002). Statistical Methods in Diagnostic Medicine (1st ed.). Wiley-Interscience.

Literature Review

- Re-weighting, imputation, and Bayesian approaches have been proposed to reduce the effect of partial verification bias
- Maximum likelihood methods and latent class models have been proposed to estimate diagnostic accuracy in the presence of imperfect reference standard bias
- A method using general estimating equations can correct for missing disease status, but does not account for misclassification of disease status.
- We have not found any methods that reduce the effect of paired screening trial bias.

Nath Algorithm

$$\widehat{\mu}_1 = \overline{x} - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\widehat{\sigma}_1$$

$$\widehat{\mu}_2 = \overline{y} - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)\widehat{\sigma}_2$$

$$\widehat{\sigma}_1 = s_1\left[\left(1 + \widehat{\xi}\widehat{P}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\eta}\widehat{Q}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\sigma}_1 = s_2\left[\left(1 + \widehat{\eta}\widehat{Q}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\xi}\widehat{P}\right) - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\rho} = rs_1s_2(\widehat{\sigma}_1\widehat{\sigma}_2)^{-1}\Big[\left(1 + \widehat{\xi}\widehat{P} + \widehat{\eta}\widehat{Q} + \widehat{A}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)/\widehat{\rho}\Big]^{-1}$$

(Nath, 1971)

Nath Algorithm

$$\widehat{\mu}_1 = \overline{x} - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\widehat{\sigma}_1$$

$$\widehat{\mu}_2 = \overline{y} - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)\widehat{\sigma}_2$$

1 $$\widehat{\sigma}_1 = s_1\left[\left(1 + \widehat{\xi}\widehat{P}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\eta}\widehat{Q}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)^2\right]^{-\frac{1}{2}}$$

1 $$\widehat{\sigma}_1 = s_2\left[\left(1 + \widehat{\eta}\widehat{Q}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\xi}\widehat{P}\right) - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\rho} = rs_1s_2(\widehat{\sigma}_1\widehat{\sigma}_2)^{-1}\left[\left(1 + \widehat{\xi}\widehat{P} + \widehat{\eta}\widehat{Q} + \widehat{A}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)/\widehat{\rho}\right]^{-1}$$

(Nath, 1971)

Nath Algorithm

2 $$\widehat{\mu}_1 = \overline{x} - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\widehat{\sigma}_1$$

2 $$\widehat{\mu}_2 = \overline{y} - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)\widehat{\sigma}_2$$

$$\widehat{\sigma}_1 = s_1\left[\left(1 + \widehat{\xi}\widehat{P}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\eta}\widehat{Q}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\sigma}_1 = s_2\left[\left(1 + \widehat{\eta}\widehat{Q}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\xi}\widehat{P}\right) - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\rho} = rs_1s_2(\widehat{\sigma}_1\widehat{\sigma}_2)^{-1}\left[\left(1 + \widehat{\xi}\widehat{P} + \widehat{\eta}\widehat{Q} + \widehat{A}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)/\widehat{\rho}\right]^{-1}$$

(Nath, 1971)

Nath Algorithm

$$\widehat{\mu}_1 = \overline{x} - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\widehat{\sigma}_1$$

$$\widehat{\mu}_2 = \overline{y} - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)\widehat{\sigma}_2$$

$$\widehat{\sigma}_1 = s_1\left[\left(1 + \widehat{\xi}\widehat{P}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\eta}\widehat{Q}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)^2\right]^{-\frac{1}{2}}$$

$$\widehat{\sigma}_1 = s_2\left[\left(1 + \widehat{\eta}\widehat{Q}\right) + \widehat{\rho}^2\left(\widehat{A} + \widehat{\xi}\widehat{P}\right) - \left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)^2\right]^{-\frac{1}{2}}$$

3

$$\widehat{\rho} = rs_1s_2(\widehat{\sigma}_1\widehat{\sigma}_2)^{-1}\left[\left(1 + \widehat{\xi}\widehat{P} + \widehat{\eta}\widehat{Q} + \widehat{A}\right) - \left(\widehat{P} + \widehat{\rho}\widehat{Q}\right)\left(\widehat{Q} + \widehat{\rho}\widehat{P}\right)/\widehat{\rho}\right]^{-1}$$

(Nath, 1971)

We derived expressions for the weighted parameter estimates using the conditional covariance formula (Proposition 5.2, p. 348, Ross, 2009) and the definition of the weighted mean (Equation 3.2.1, p. 77, Kish, 1995). We define the estimates as follows:

$$\widehat{\mu}_{11,W} = \widehat{\lambda}\overline{X}_{11,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{11,B}, \quad (3)$$

$$\widehat{\mu}_{21,W} = \widehat{\lambda}\overline{X}_{21,A} + \left(1 - \widehat{\lambda}\right)\overline{X}_{21,B}, \quad (4)$$

$$\widehat{\sigma}^2_{11,W} = \mathcal{G}_1 + \mathcal{H}_1 - \widehat{\mu}^2_{11,W}, \quad (5)$$

$$\widehat{\sigma}^2_{21,W} = \mathcal{G}_2 + \mathcal{H}_2 - \widehat{\mu}^2_{21,W}, \quad (6)$$

and

$$\widehat{\rho}_{1,W} = \widehat{\sigma}^2_{11,W}\widehat{\sigma}^2_{21,W}\left(\mathcal{P} + \mathcal{Q} - \widehat{\mu}_{11,W}\widehat{\mu}_{21,W}\right) \quad (7)$$

where

$$\mathcal{G}_j = \widehat{\lambda}\left(\overline{X}^2_{j1,A} + S^2_{j1,A}\right),$$
$$\mathcal{H}_j = \left(1 - \widehat{\lambda}\right)\left(\overline{X}^2_{j1,B} + S^2_{j1,B}\right),$$
$$\mathcal{P} = \widehat{\lambda}\overline{X}_{11,A}\overline{X}_{21,A} + \widehat{\lambda}S_{11,A}S_{21,A}r_{1,A},$$

and

$$\mathcal{Q} = \left(1 - \widehat{\lambda}\right)\overline{X}_{11,B}\overline{X}_{21,B} + \left(1 - \widehat{\lambda}\right)S_{11,B}S_{21,B}r_{1,B}.$$