Multi-reader diagnostic imaging case studies: components of variance

Brandon D. Gallas Ph.D.
FDA/CDRH/OSEL/DIAM

Abstract
In this talk we will introduce the framework for analyzing multi-reader diagnostic imaging case studies. This framework is the launching point for the advanced methods discussed in the other talks of this session. The core of the framework is the accounting for reader variability in addition to case variability in the analysis of diagnostic imaging performance metrics. There are several methods (and software) for representing and estimating the constituent components of variance, and we will summarize a few. The methods are often referred to as multi-reader multi-case (MRMC) variance analysis methods. We will demonstrate MRMC analysis of several real studies and discuss the sizing of future studies.

Keywords
MRMC, AUC, variance, imaging

25 minutes for presentation + 10 for questions

Multi-reader diagnostic imaging case studies: components of variance

Brandon D. Gallas, PhD

FDA/CDRH

Office of Science and Engineering Laboratories

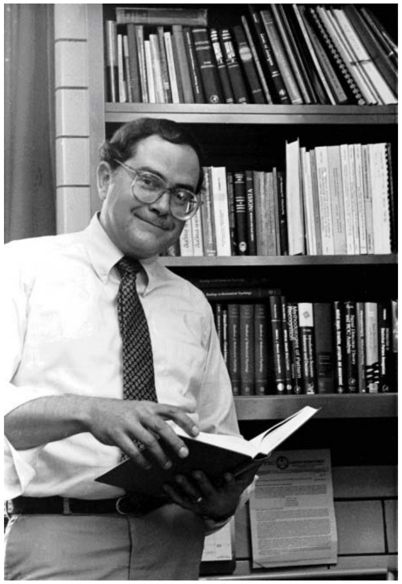
Division of Imaging and Applied Mathematics

Outline

- Diagnostic Performance
  - Area under ROC curve
- Variance representations
- Sizing a study
  - Using Components of Variance
  - Investigating Study designs Practical vs. Statistical Efficiency

Variance Representations
DBM 3-way ANOVA

– Charlie Metz, University of Chicago

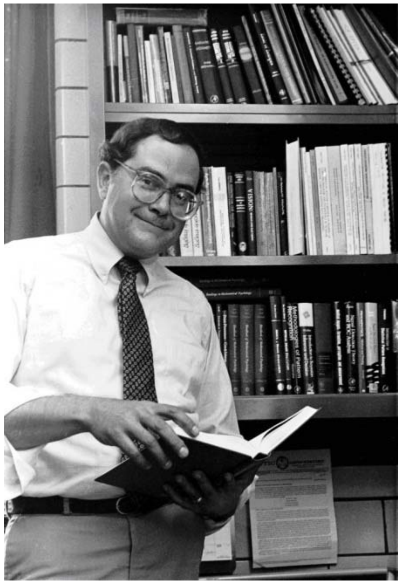


Numerous contributions:
radiological imaging
nuclear medicine
ROC analysis
medical-decision making
computer-aided diagnosis
…
Scientist
Educator
Mentor

- September 11, 1942 - July 4, 2012

Variance Representations
DBM 3-way ANOVA

– Charlie Metz, University of Chicago



<u>Memorial service:</u>
11 a.m. Monday, Aug. 27
Rockefeller Memorial Chapel, followed by a reception in Ida Noyes Hall's Cloister Club.

- September 11, 1942 - July 4, 2012

Diagnostic Performance

- Sensitivity
  - Success rate on diseased cases
- Specificity
  - Success rate on non-diseased cases
- ROC and Area Under ROC curve
  - Tradeoff between Sensitivity and Specificity
  - Interaction btw diseased and non-diseased

Diagnostic Performance

Observer Data

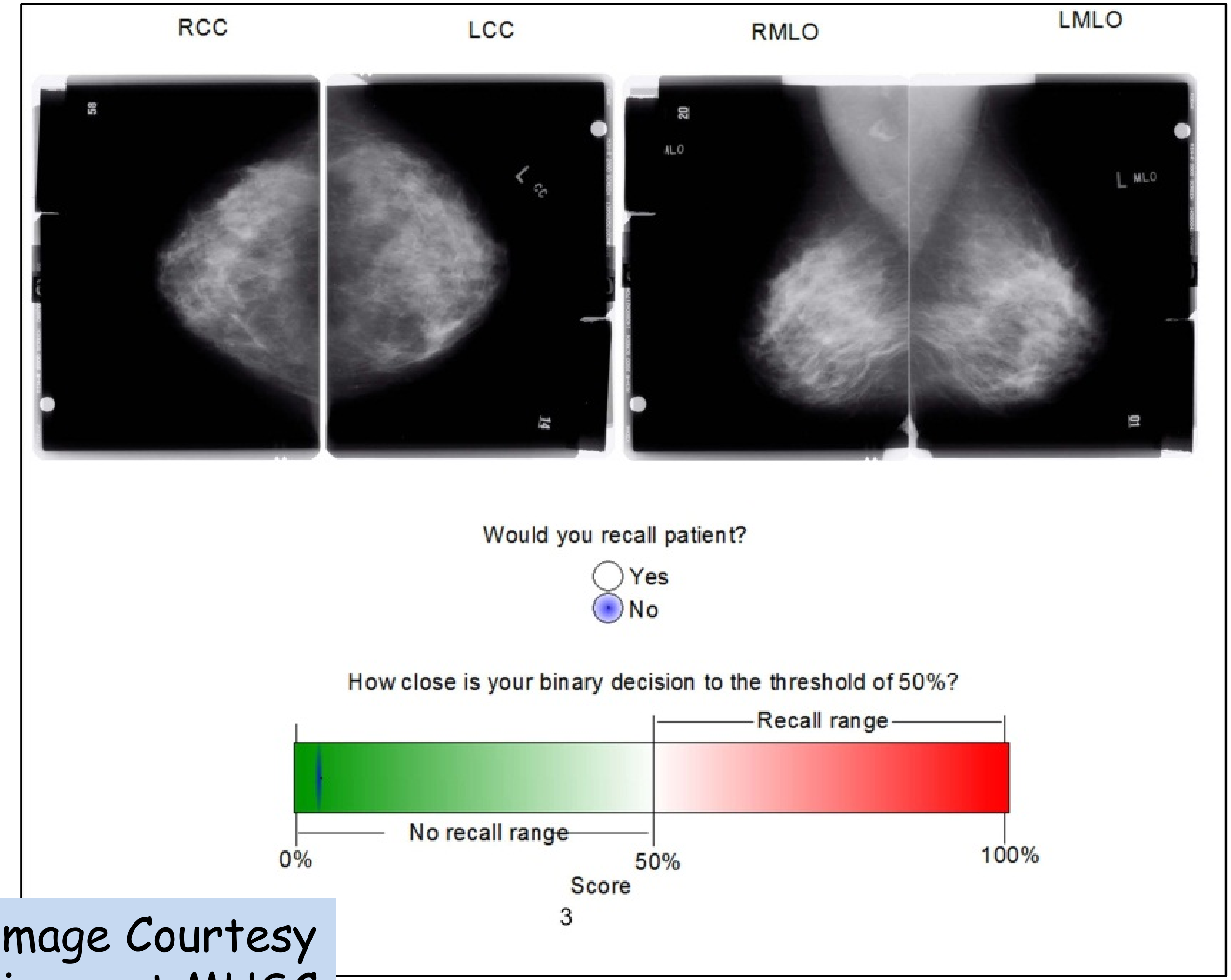


Image Courtesy
Cole and Pisano at MUSC

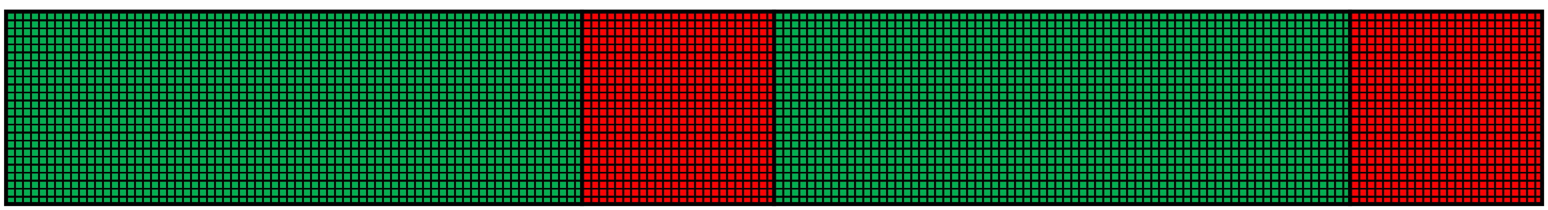
Diagnostic Performance

Observer Data

- Fully-crossed study
  - All readers read all cases
  - Readers and cases are paired across modalities

Data Array:
Rows = readers
Cols = cases

Modality 1 | Modality 2



Non-Diseased Cases | Diseased Cases | Non-Diseased Cases | Diseased Cases

Time = 4:00

- Pause
- Previous Time = 5:13

Diagnostic Performance
Reader-averaged AUC

- Focus: Reader-averaged AUC
  - Summary of ROC
  - Reader average of ability to separate
  - Equivalent to probability

Diagnostic Performance
Reader-averaged AUC

- Nonparametric estimator
  - U-stats, Mann-Whitney, Wilcoxon, Trapezoid

$$\widehat{\mathrm{AUC}}_m = \sum_{r=1}^{N_R} \frac{1}{N_R} \sum_{j=1}^{N_1} \frac{1}{N_1} \sum_{i=1}^{N_0} \frac{1}{N_0} s_{mijr}$$

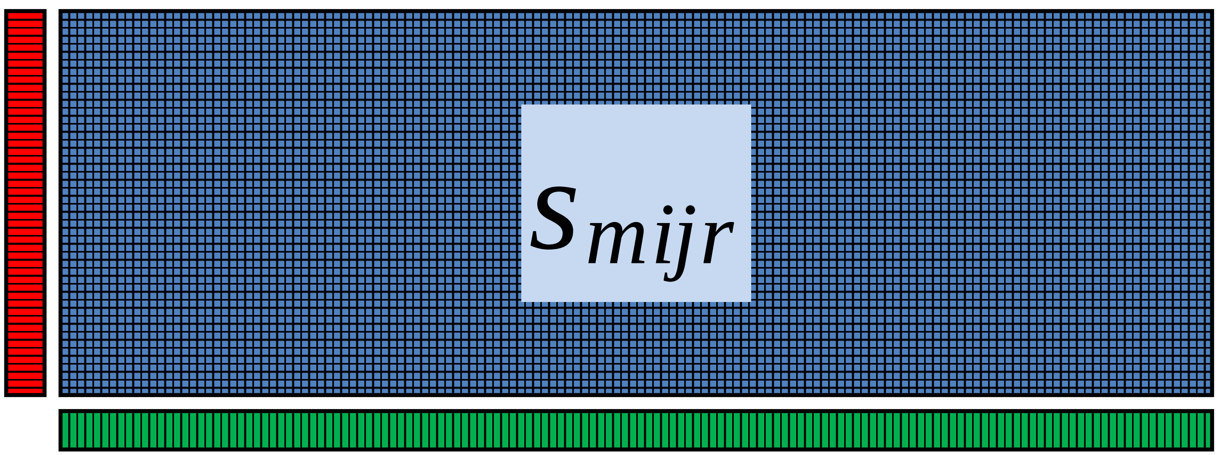
$$s_{mijr} = s(y_{mjr} - x_{mir}) = \begin{cases} 1 & y_{mjr} - x_{mir} > 0 \\ 1/2 & y_{mjr} - x_{mir} = 0 \\ 0 & y_{mjr} - x_{mir} < 0 \end{cases},$$

m: Modality
i: Non-diseased case
j: Diseased case
r: Reader

Diagnostic Performance

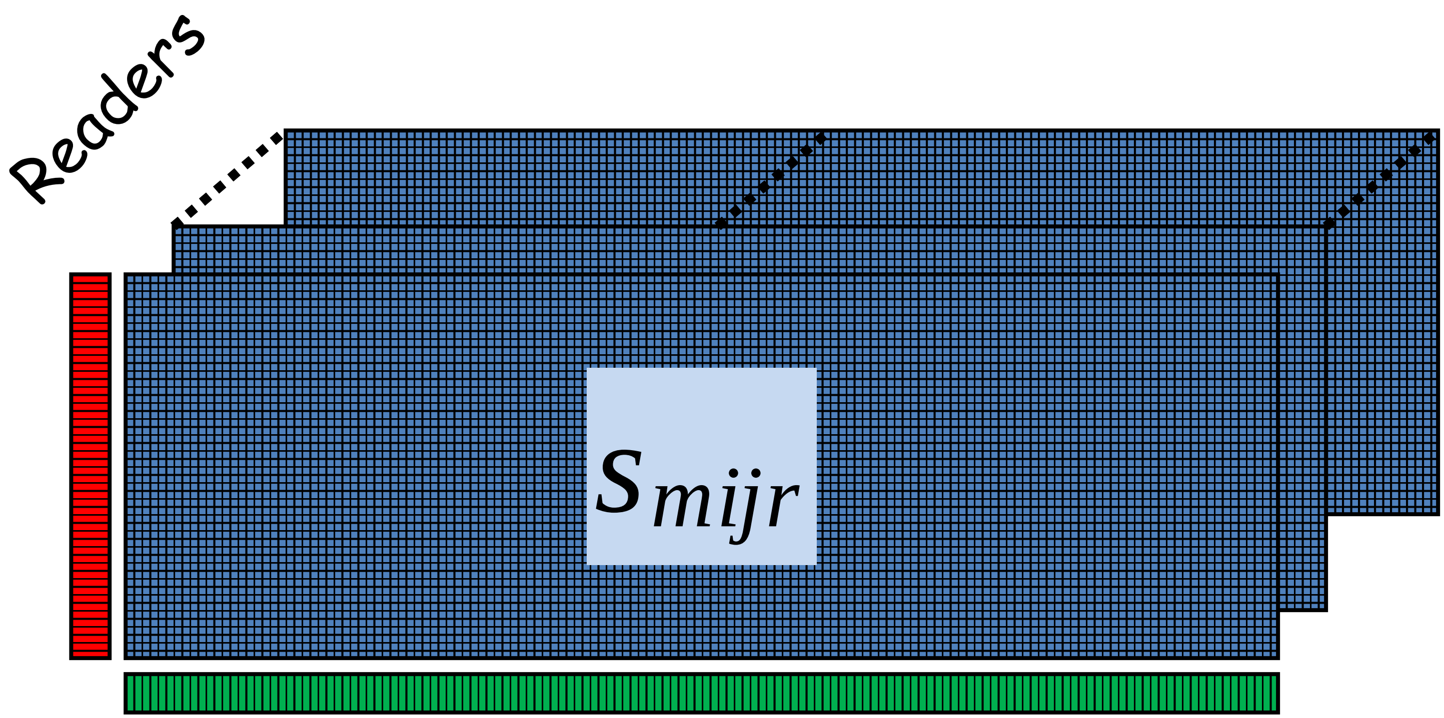
Success Matrix

- One modality *m*
- One reader *r*
- Row *i*: Diseased case
- Col *j*: Non-diseased case



Diagnostic Performance
Success Matrix

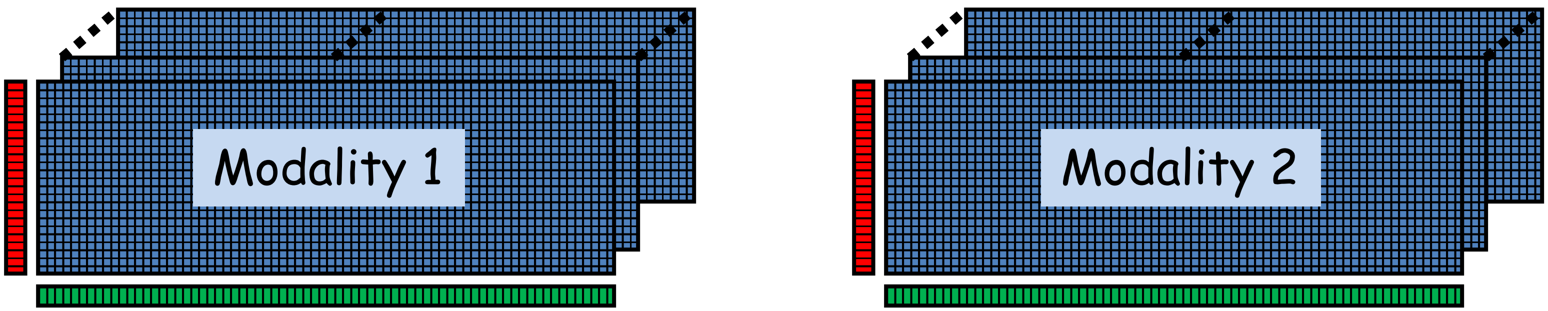
- One modality
- Multiple readers
- Row *i*: Diseased case
- Col *j*: Non-diseased case



Diagnostic Performance

Success Matrix

- Multiple modalities
- Multiple readers
- Row *i*: Diseased case
- Col *j*: Non-diseased case



Diagnostic Performance
Nonparametric AUCs

- Average elements of Success Matrix

Estimates

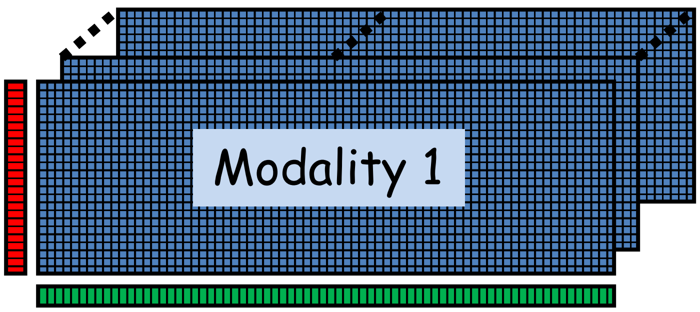
$$\bar{s}_{m \bullet \bullet r} = \widehat{\mathrm{AUC}}_{mr}$$

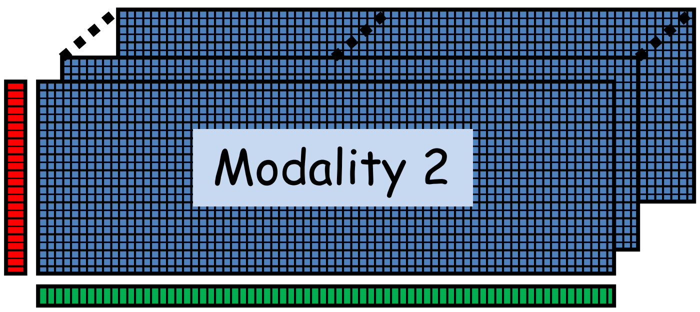
$$\bar{s}_{m \bullet \bullet \bullet} = \widehat{\mathrm{AUC}}_{m}$$

population quantities

$$E(s_{mijr} \mid mr) = \mathrm{AUC}_{mr}$$

$$E(s_{mijr} \mid m) = \mathrm{AUC}_{m}$$





Time = 6:00

- Pause
- Time=2:30, Total=7:45

Variance Representations

- Main Random Effects
  - case variability
    difficulty
  - reader variability
    skill
  - reader/case interaction
    training
    readers reading same cases

- **MRMC:** Multi-Reader Multi-Case Variance Analysis

Variance Representations

- Main Random Effects
  - case variability
    Non-disease + Disease + Interaction
  - reader variability
  - reader/case interaction
    Non-disease + Disease + Interaction

Variance Representations
U-statistic result

- Single Modality
  - Gallas et al. (2009)

Non-diseased cases

Diseased cases

Interaction

$$\mathrm{var}\left(\widehat{\mathrm{AUC}}\right) = \frac{\sigma_0^2}{N_0} + \frac{\sigma_1^2}{N_1} + \frac{\sigma_{01}^2}{N_0 N_1} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{0R}^2}{N_0 N_R} + \frac{\sigma_{1R}^2}{N_0 N_R} + \frac{\sigma_{01R}^2}{N_0 N_1 N_R}$$

No modeling, MVUE

7components

Variance Representations
U-statistic result

- Two Modalities
  - Gallas et al. (2009)

Non-diseased cases

Diseased cases

Interaction

$$\mathrm{var}\left(\widehat{\mathrm{AUC}}_1 - \widehat{\mathrm{AUC}}_2\right) = \frac{\sigma_0^2}{N_0} + \frac{\sigma_1^2}{N_1} + \frac{\sigma_{01}^2}{N_0 N_1} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{0R}^2}{N_0 N_R} + \frac{\sigma_{1R}^2}{N_0 N_R} + \frac{\sigma_{01R}^2}{N_0 N_1 N_R}$$

No modeling, MVUE

Of course, different interpretation for these components of variance from before.

Variance Representations
U-statistic result

- Two Modalities
  - Gallas et al. (2009)

Non-diseased cases

Diseased cases

Interaction

$$\operatorname{var}\left(\widehat{\mathrm{AUC}}_1 - \widehat{\mathrm{AUC}}_2\right) = \frac{\sigma_0^2}{N_0} + \frac{\sigma_1^2}{N_1} + \frac{\sigma_{01}^2}{N_0 N_1} + \frac{\sigma_R^2}{N_R} + \frac{\sigma_{0R}^2}{N_0 N_R} + \frac{\sigma_{1R}^2}{N_0 N_R} + \frac{\sigma_{01R}^2}{N_0 N_1 N_R}$$

7 components modality 1
7 components modality 2
7 covariances

Variance Representations
MLE, Ideal Bootstrap, Method of Moments

- Two Modalities
  - Gallas et al. (2009)

Non-diseased cases

Diseased cases

Interaction

$$\mathrm{var}\left(\widehat{\mathrm{AUC}}_1 - \widehat{\mathrm{AUC}}_2\right) = \frac{\sigma_0^2}{N_0} + \frac{\sigma_1^2}{N_1} + \frac{\sigma_{01}^2}{N_0 N_1}$$

$$+ \frac{\sigma_R^2}{N_R}$$

$$+ \frac{\sigma_{0R}^2}{N_0 N_R} + \frac{\sigma_{1R}^2}{N_0 N_R} + \frac{\sigma_{01R}^2}{N_0 N_1 N_R}$$

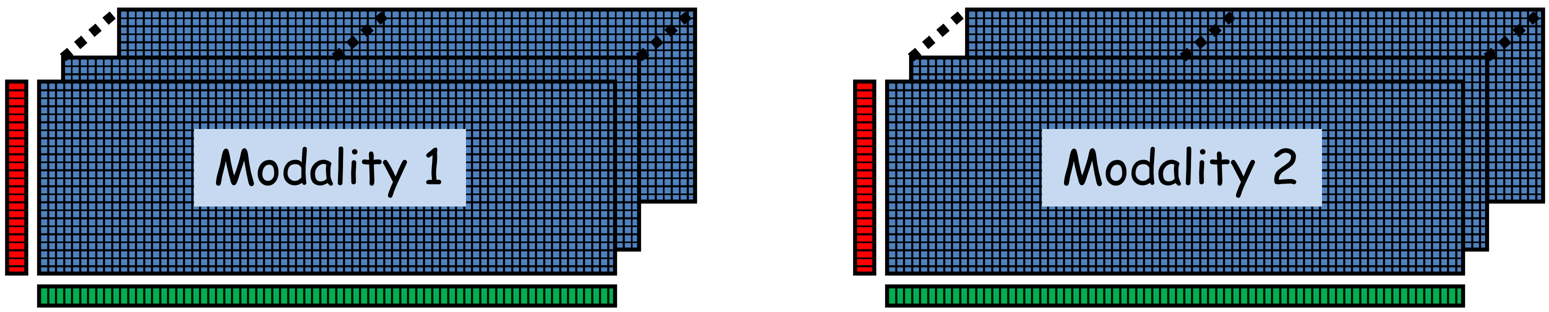
MLE: Assume samples are entire distribution.

Introduce bias for positive variances (components)

Variance Representations

U-statistic result = 4-way ANOVA

- 4 effects:
  - Multiple modalities (fixed)
  - Multiple readers (random)
  - Rows: Diseased cases (random)
  - Columns: Non-diseased cases (random)

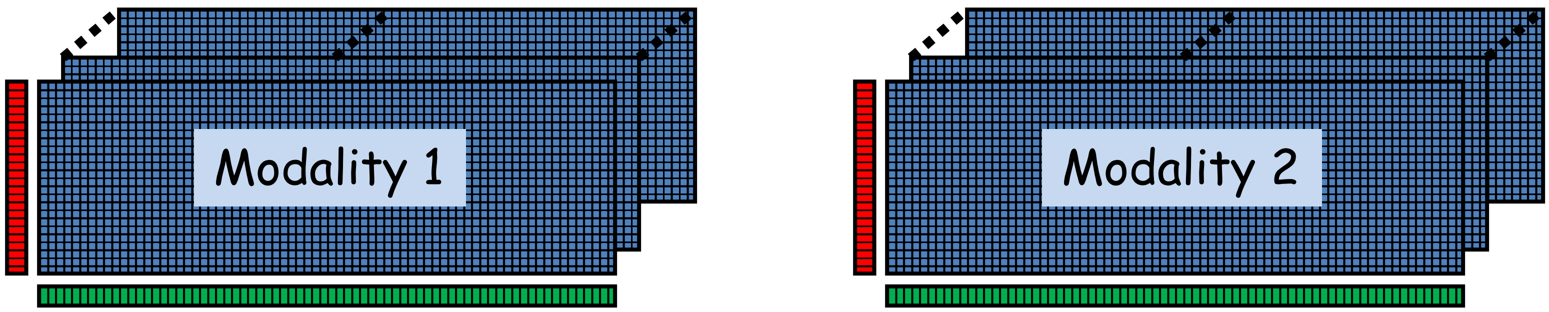


Variance Representations

U-statistic result = 4-way ANOVA

- Caveats:
  - Observed data are success elements

    Not strictly independent observations
  - Can specify variances per modality or pool



Time = 9:30

- Pause
- Time=5:15, Total=13:00

Variance Representations
ANOVA

- DBM: Dorfman, Berbaum and Metz (1992)
  - 3-way ANOVA: modality, readers, cases
  - Jackknife pseudovalues
- OR: Obuchowski & Rockette (1995)
  - 2-way ANOVA: modality, reader
  - Correlated errors

- Marginal-Mean ANOVA
  - Mm-ANOVA
  - (Hillis to be submitted)
  - Bridge linking both variance representations
  - Hypothetical 3-way ANOVA no pseudovalues
  - Estimation based on OR

Variance Representations
3-way ANOVA

- Single Modality: DBM & mm-ANOVA

$$\mathrm{var}\left(\widehat{\mathrm{AUC}}\right) = \frac{\sigma^2_{\text{cases}} + \sigma^2_{\text{mod}\times\text{cases}}}{N_{\text{cases}}} + \frac{\sigma^2_{\text{readers}} + \sigma^2_{\text{mod}\times\text{readers}}}{N_{\text{readers}}} + \frac{\sigma^2_{\text{cases}\times\text{readers}} + \sigma^2_{\text{mod}\times\text{cases}\times\text{readers}}}{N_{\text{cases}} N_{\text{readers}}}$$

Can also estimate these components from U-statistics or MLE

Diseased and non-diseased effects are pooled

Variance Representations

3-way ANOVA: DBM & mm-ANOVA

- Two Modalities: DBM & mm-ANOVA

$$\operatorname{var}\left(\widehat{AUC_1} - \widehat{AUC_2}\right) = \frac{2\sigma^2_{\text{mod}\times\text{cases}}}{N_{\text{cases}}} + \frac{2\sigma^2_{\text{mod}\times\text{readers}}}{N_{\text{readers}}} + \frac{2\sigma^2_{\text{mod}\times\text{cases}\times\text{readers}}}{N_{\text{cases}}N_{\text{readers}}}$$

Can also estimate these components from U-statistics or MLE

Diseased and non-diseased effects are pooled

Variance Representations

2-way ANOVA & Correlated Errors
OR & mm-ANOVA

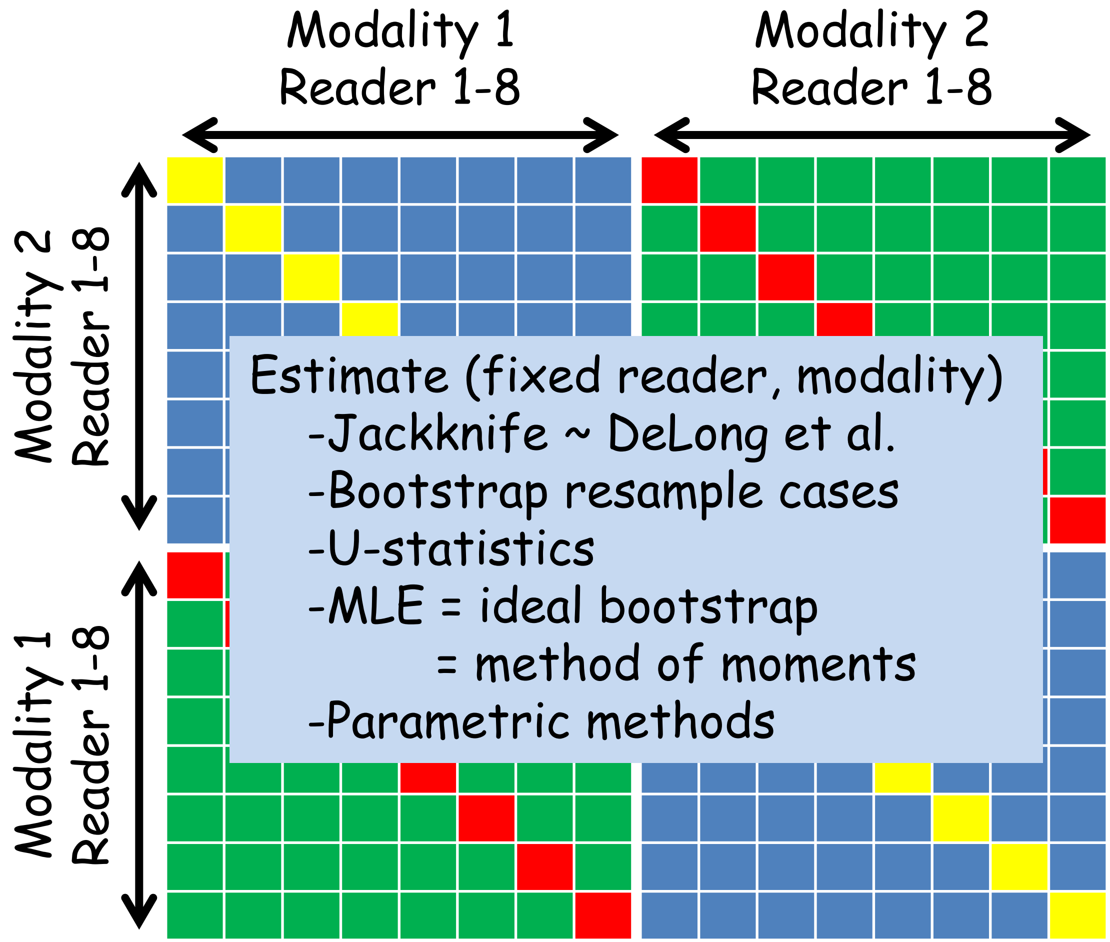
$$\mathrm{cov}\left(\widehat{\mathrm{AUC}}_{mr}, \widehat{\mathrm{AUC}}_{mr} \,|mr\right)$$

Var_ε =
Same modality, same reader

Cov1 =
Diff modality, same reader

Cov2 =
Same modality, Diff reader

Cov3 =
Diff. modality, Diff. reader



Estimate (fixed reader, modality)
- -Jackknife ~ DeLong et al.
- -Bootstrap resample cases
- -U-statistics
- -MLE = ideal bootstrap
  = method of moments
- -Parametric methods

Variance Representations
Hillis Marginal-Mean (3-way) ANOVA

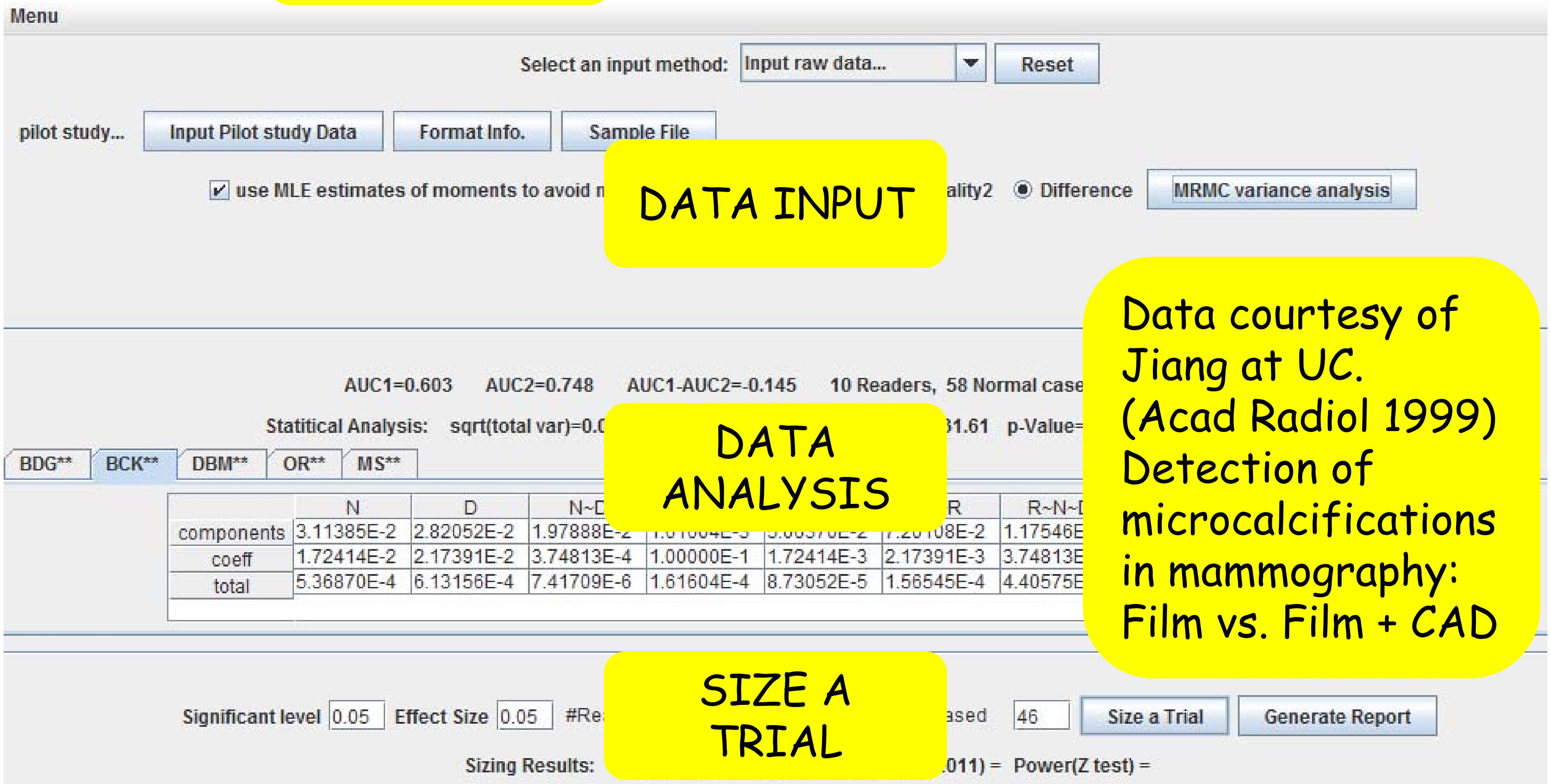
- If use U-stats (or MLE) to estimate fixed-reader covariances

- … then equivalence across variance representations.

Variance in Reader Studies: iMRMC

- http://...rmc.html (coming to googlecode)

Main author: Xin He, FDA



DATA INPUT

DATA ANALYSIS

SIZE A TRIAL

Data courtesy of Jiang at UC. (Acad Radiol 1999) Detection of microcalcifications in mammography: Film vs. Film + CAD

Variance in Reader Studies:
iMRMC, U-Stats

- http://js.cx/~xin/mrmc.html

AUC1=0.603 AUC2=0.748 AUC1-AUC2=-0.145

10 Readers, 58 Normal cases, 46 Disease cases.

Statitical Analysis: sqrt(total var)=0.001073 tStat= 4.41 df(Hillis 2008)= 321.68 p-Value= 0.0000 Conf. Int.=(-0.21, -0.08)

BDG BCK DBM OR MS

	N	D	N~D	R	N~R	D~R	R~N~D
components	2.58076E-2	2.05323E-2	6.99835E-3	-9.18608E-4	5.42972E-2	7.94478E-2	1.35851E-1
coeff	1.72414E-2	2.17391E-2	3.74813E-4	1.00000E-1	1.72414E-3	2.17391E-3	3.74813E-5
total	4.44958E-4	4.46354E-4	2.62307E-6	-9.18608E-5	9.36158E-5	1.72713E-4	5.09186E-6

Var=0.001073

Variance in Reader Studies: iMRMC, MLE

- http://js.cx/~xin/mrmc.html

AUC1=0.603   AUC2=0.748   AUC1-AUC2=-0.145   10 Readers, 58 Normal cases, 46 Disease cases.

Statitical Analysis: sqrt(total var)=0.001567   tStat= 3.65   df(Hillis 2008)= 131.61   p-Value= 0.0004 Conf. Int.=(-0.22, -0.07)

BDG**   BCK**   DBM**   OR**   MS**

	N	D	N~D	R	N~R	D~R	R~N~D
components	3.11385E-2	2.82052E-2	1.97888E-	1.61604E-3	06370E-2	7.20108E-2	1.17546E-1
coeff	1.72414E-2	2.17391E-2	3.74813E-	1.00000E-1	72414E-3	2.17391E-3	3.74813E-5
total	5.36870E-4	6.13156E-4	7.41709E-	1.61604E-4	73052E-5	1.56545E-4	4.40575E-6

Var=0.001567

Variance in Reader Studies: iMRMC, DBM

- http://js.cx/~xin/mrmc.html

AUC1=0.603 AUC2=0.748 AUC1-AUC2=-0.145 10 Readers, 58 Normal cases, 46 Disease cases.

Statitical Analysis: sqrt(total var)=0.001567 tStat= 3.65 df(Hillis 2008)= 131.61 p-Value= 0.0004 Conf. Int.=(-0.22, -0.07)

BDG** BCK** DBM** OR** MS**

	R	C	R~C	T~R	T~C	T~R~C
componen	7.75679E-4	.77109E-2	2.02402E-2	8.08018E-4	6.01870E-2	1.29093E-1
coeff	0.00000E0	.00000E0	0.00000E0	2.00000E-1	1.92308E-2	1.92308E-3
total	0.00000E0	.00000E0	0.00000E0	1.61604E-4	1.15744E-3	2.48256E-4

Var=0.001567

Variance in Reader Studies: iMRMC, OR & mm-ANOVA

- http://js.cx/~xin/mrmc.html

AUC1=0.603 AUC2=0.748 AUC1-AUC2=-0.145 10 Readers, 58 Normal cases, 46 Disease cases.

Statitical Analysis: sqrt(total var)=0.001567 tStat= 3.65 df(Hillis 2008)= 131.61 p-Value= 0.0004 Conf. Int.=(-0.22, -0.07)

BDG** BCK** DBM** OR** MS**

	R	TR	COV1	COV2	COV3	ERROR
componen	7.75679E-4	.08018E-4	9.41837E-4	1.32594E-3	7.47220E-4	2.76184E-3
coeff	0.00000E0	.00000E-1	-2.00000E-1	1.80000E0	-1.80000E0	2.00000E-1
total	0.00000E0	.61604E-4	-1.88367E-4	2.38669E-3	-1.34500E-3	5.52368E-4

Var=0.001567

Time = 16:15

- Pause
- Time=9:00, Total=22:00

Variance Representations

Methodology Evaluation Tool

- Roe and Metz (1997)
  ROC simulation model
  - Multiple modalities (fixed effect)
  - Multiple readers
  - Multiple cases
  - TRUTH!

Except for truth looks like 3-way ANOVA

$$X_{ijkt} = \mu_t + \tau_{it} + R_{jt} + C_{kt} + (\tau R)_{ijt} + (\tau C)_{ikt} + (RC)_{jkt} + (\tau RC)_{ijkt} + E_{ijkt}$$

Variance Representations

Size a Trial

- Two Modalities

Non-diseased cases

Diseased cases

Interaction

$$\mathrm{var}\left(\widehat{\mathrm{AUC}}_1 - \widehat{\mathrm{AUC}}_2\right) = \boxed{\frac{\sigma_0^2}{N_0}} + \boxed{\frac{\sigma_1^2}{N_1}} + \frac{\sigma_{01}^2}{N_0 N_1}$$

$$+ \frac{\sigma_R^2}{N_R}$$

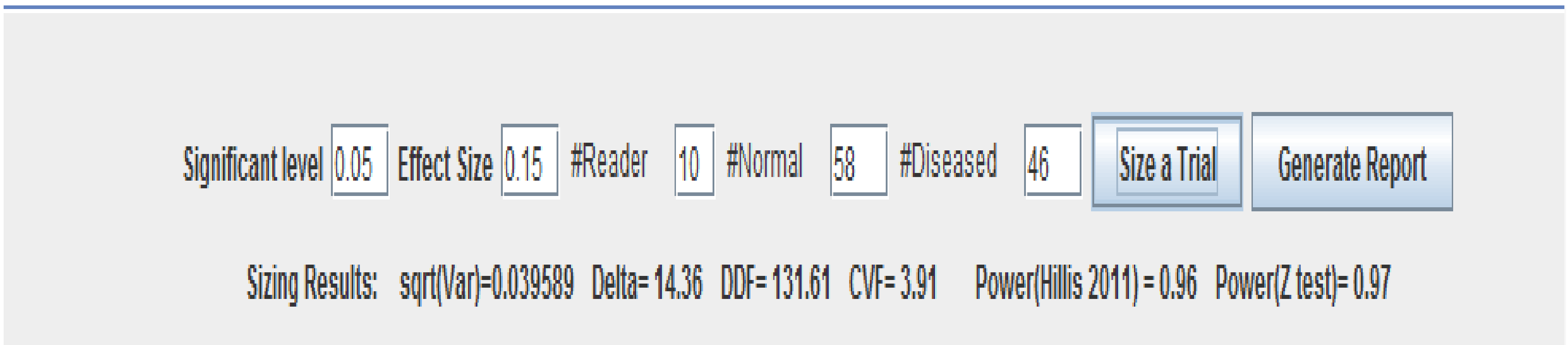
$$+ \boxed{\frac{\sigma_{0R}^2}{N_0 N_R}} + \boxed{\frac{\sigma_{1R}^2}{N_0 N_R}} + \frac{\sigma_{01R}^2}{N_0 N_1 N_R}$$

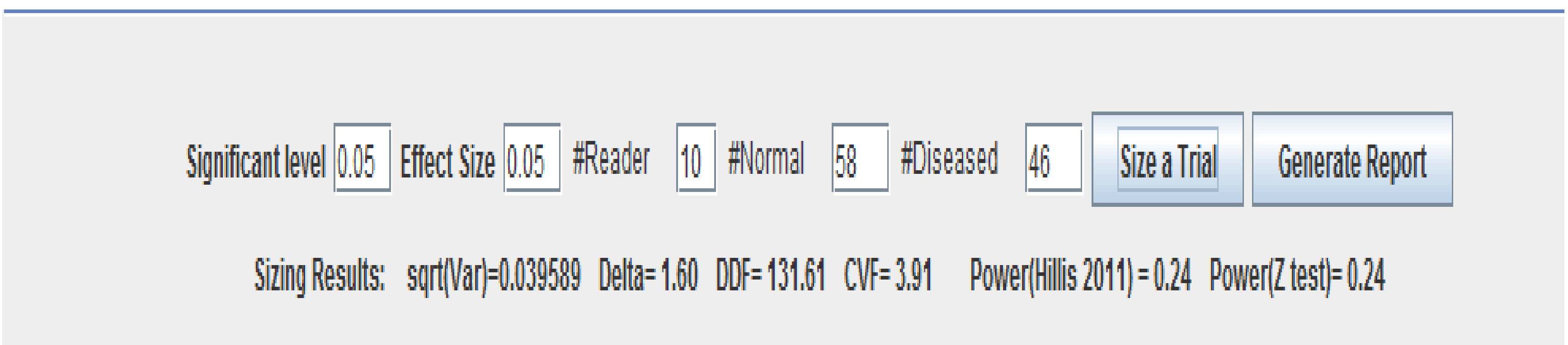
Pick N0, N1, NR, see the variance.

C.I. & Hypothesis test:
- -Gaussian model
- -d.o.f. from Hillis

Variance in Reader Studies: Size a Trial

- http://js.cx/~xin/mrmc.html





Variance in Reader Studies:
Size a Trial

- NIH/ASCCP sub-study of ALTS [2-3]
  - Atypical Squamous Cells of Undetermined Significance (ASCUS) Low-Grade Squamous Intraepithelial Lesion (LSIL) Triage Study
  - Colposcopy

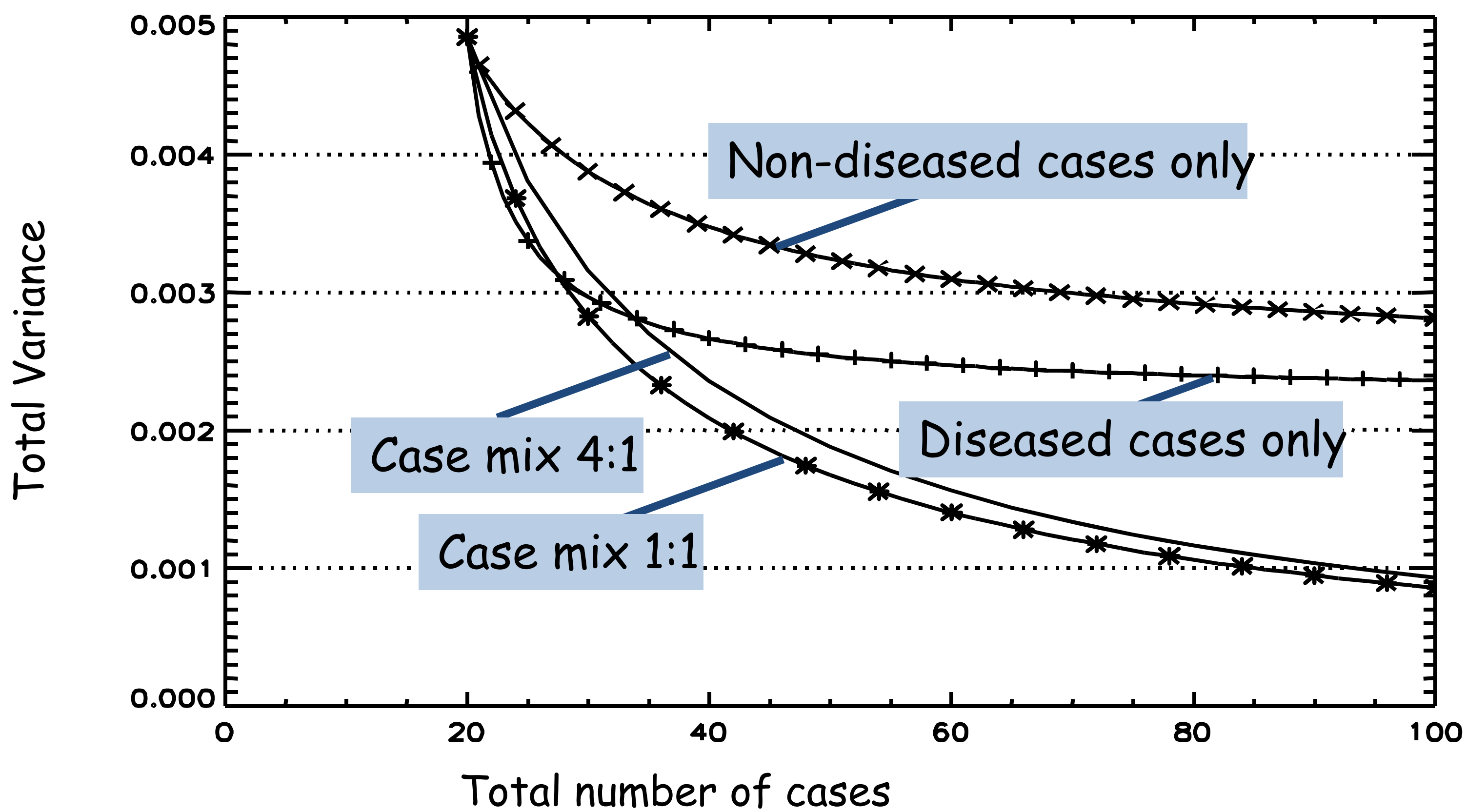
- 1,000 women enrolled; 939 with evaluable Cervigrams™

- 21 colposcopists
- 20 patients (16 normal and 4 diseased) had Cervigrams™ read by every reader (420 readings)

- Overall diagnosis for patient

4:1 sampling
-> 25% study prevalence

Variance in Reader Studies:
Size a Trial



Reduce variance by 50%.
1:1 sampling: add 15 cases.
4:1 sampling: add 20 cases

Plot courtesy of Hsu, NCI.

Variance in Reader Studies: Size a Trial

- Which components of variance should I use?
  - Pilot study
  - Find data from a completed study that is
    - Relevant to imaging modality
    - Relevant to viewing conditions
    - Relevant to task/disease
- iMRMC: working on annotated database

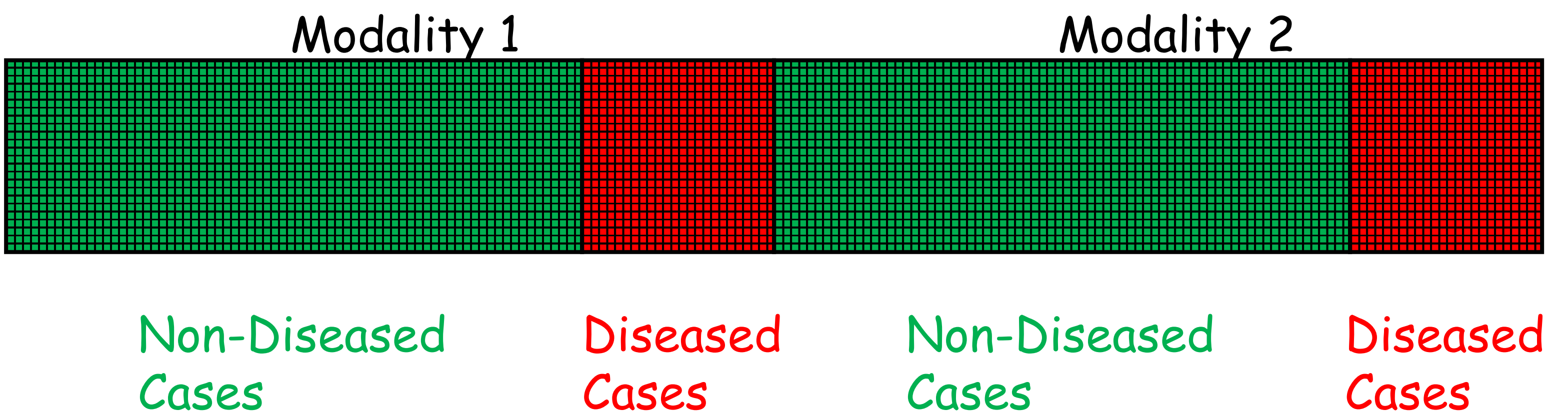
Time = 21:44

- Pause
- Time=5:30, Total=27:30

Study Designs

- Fully-crossed study
  - All readers read all cases
  - Readers and cases are paired across modalities

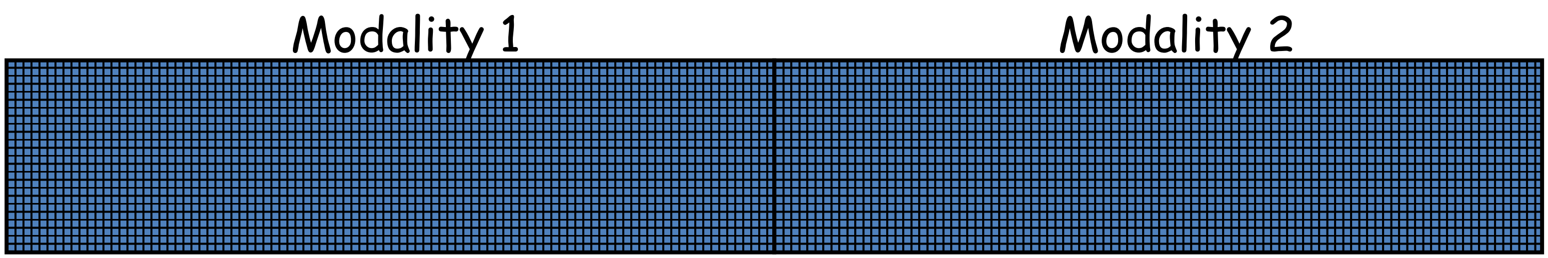
Data Array:
Rows = readers
Cols = cases



Study Designs

- Fully-crossed study
  - All readers read all cases
  - Readers and cases are paired across modalities

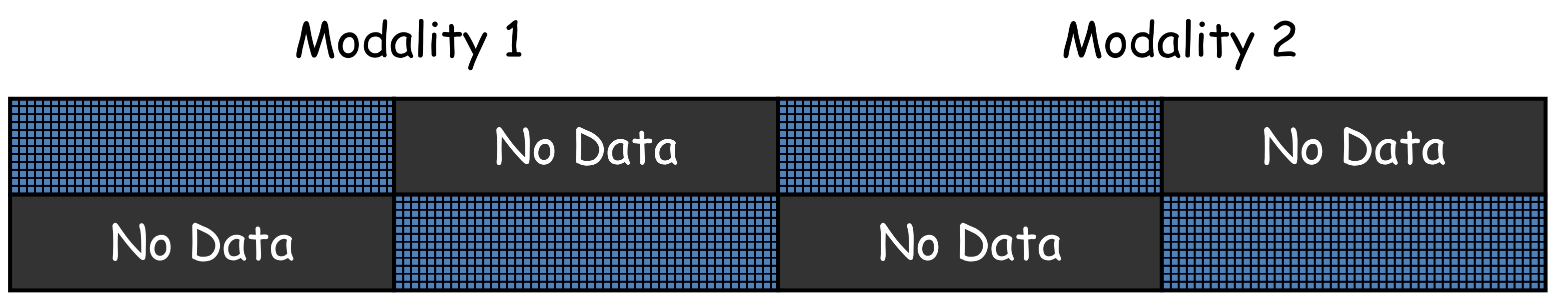
Remove truth labels to unclutter study design concepts.
Diseased and non-diseased cases need to be treated separately.



Data Array:
Rows = readers
Cols = cases

Study Designs

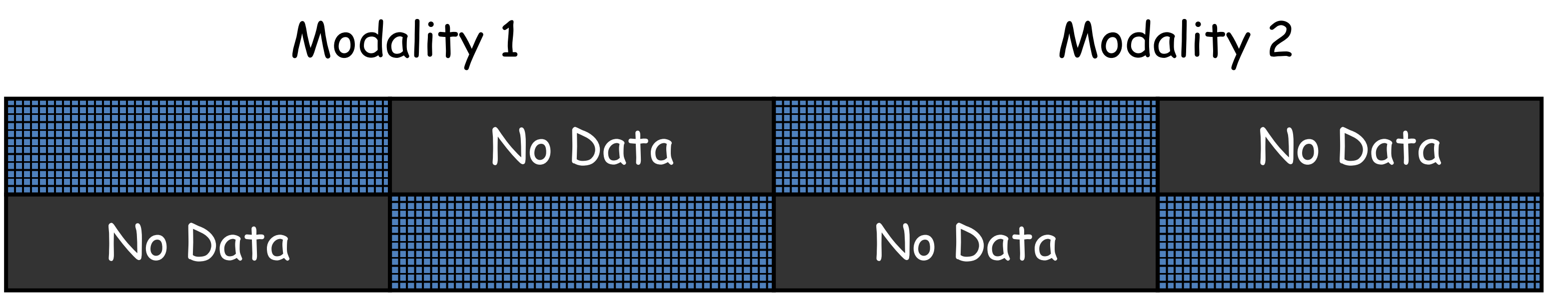
- Fully-crossed study is burdensome
  - All readers read all cases
  - Readers and cases are paired across modalities
- Split-plot study
  - Readers and cases split into 2 groups
  - Data is fully-crossed within a group



Data Array:
Rows = readers
Cols = cases

Study Designs

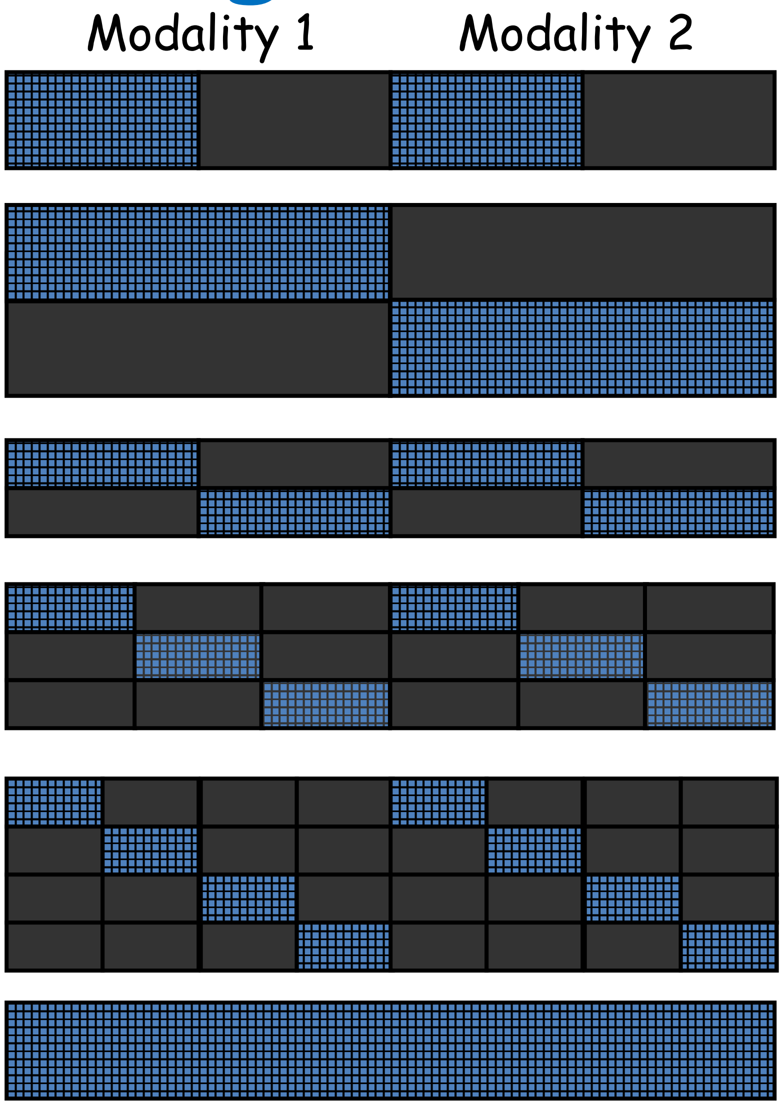
- Fully-crossed is burdensome
  - A lot of reads per reader
  - A lot of reads total
- Split-plot study may save time (and money)
  - Half the reads per reader
  - Half the reads total



Data Array:
Rows = readers
Cols = cases

Study Designs

- Fully-Crossed A
- Readers Unpaired Across Modalities
- 2-Groups
- 3-Groups
- 4-Groups
- Fully-Crossed B



Study Designs
Efficiency

- "U-statistic" approach that decouples variance components from study design
  - *Gallas and Brown (2008)*

- Roe and Metz simulation
  - given description of scores, know the components of variance (numerical integration)

- Model parameters ($\Delta\mu$ = 1.53)

Var_r	Var_c	Var_rc	Var_tr	Var_tc	Var_trc
0.011	0.100	0.200	0.030	0.100	0.200

Study Designs: Efficiency

Study Design	Groups	Readers		Cases		Reads		Statistical Efficiency
	G	$\frac{J}{G}$	J	$\frac{N_1}{G} + \frac{N_0}{G}$	N_{Total} $N_0 + N_1$	per reader	total	$\frac{\text{var(2-groups)}}{\text{var(alt. design)}}$
Full-A	1	6	6	30+30	60	120	720	0.83
Unpaired Readers	1	6	12*	60+60	120	120	1440	0.90
2-groups	2	3	6	30+30	120	120	720	1.00
3-groups	3	3	9	20+20	120	80	720	1.20
4-groups	4	3	12	15+15	120	60	720	1.33
Full-B	1	6	6	60+60	120	240	1440	1.16

Resources: Tried to control
- total # reads
- total # cases
- # reads per reader

Study Designs:
Efficiency

Study Design	Groups	Readers		Cases		Scores		Statistical Efficiency
	G	$\frac{J}{G}$	J	$\frac{N_1}{G} + \frac{N_0}{G}$	N_{Total} $N_0 + N_1$	per reader	total	$\frac{\text{var(2-groups)}}{\text{var(alt. design)}}$
Full-A	1	6	6	30+30	60	120	720	0.83
Unpaired Readers	1	6	12*	60+60	120	120	1440	0.90
2-groups	2	3	6	30+30	120	120	720	1.00
3-groups	3	3	9	20+20	120	80	720	1.20
4-groups	4	3	12	15+15	120	60	720	1.33
Full-B	1	6	6	60+60	120	240	1440	1.16

Take-away 1. It is possible (and fairly easy) to compare study designs.

Study Designs:
Efficiency

Study Design	Groups	Readers		Cases		Scores		Statistical Efficiency
	G	$\frac{J}{G}$	J	$\frac{N_1}{G}+\frac{N_0}{G}$	N_{Total} N_0+N_1	per reader	total	$\frac{\text{var(2-groups)}}{\text{var(alt. design)}}$
Full-A	1	6	6	30+30	60	120	720	0.83
Unpaired Readers	1	6	12*	60+60	120	120	$\boxed{1440}$	$\boxed{0.90}$
2-groups	2	3	6	30+30	120	120	720	1.00
3-groups	3	3	9	20+20	120	80	720	1.20
4-groups	4	3	12	15+15	120	60	720	1.33
Full-B	1	6	6	60+60	120	240	1440	1.16

Take-away 2. Pay a price when you don't pair readers across modalities.

Study Designs:
Efficiency

Study Design	Groups	Readers		Cases		Scores		Statistical Efficiency
	G	$\frac{J}{G}$	J	$\frac{N_1}{G}+\frac{N_0}{G}$	$\frac{N_{Total}}{N_0+N_1}$	per reader	total	$\frac{\text{var(2-groups)}}{\text{var(alt. design)}}$
Full-A	1	6	6	30+30	60	120	720	0.83
Unpaired Readers	1	6	12*	60+60	120	120	1440	0.90
2-groups	2	3	6	30+30	120	120	720	1.00
3-groups	3	3	9	20+20	120	80	720	1.20
4-groups	4	3	12	15+15	120	60	720	1.33
Full-B	1	6	6	60+60	120	240	1440	1.16

Take-away 3. There is a moderate hit to efficiency when you split the experiment into two groups.

Study Designs:
Efficiency

Study Design	Groups	Readers		Cases		Scores		Statistical Efficiency
	G	$\frac{J}{G}$	J	$\frac{N_1}{G}+\frac{N_0}{G}$	N_{Total} N_0+N_1	per reader	total	$\frac{\text{var(2-groups)}}{\text{var(alt. design)}}$
Full-A	1	6	6	30+30	60	120	720	0.83
Unpaired Readers	1	6	12*	60+60	120	120	1440	0.90
2-groups	2	3	6	30+30	120	120	720	1.00
3-groups	3	3	9	20+20	120	80	720	1.20
4-groups	4	3	12	15+15	120	60	720	1.33
Full-B	1	6	6	60+60	120	240	1440	1.16

Take-away 4. You can be more efficient by splitting more. (need more readers, should avoid splitting below 25 cases per truth per reader)

Time = 25:50

- Time=3:45, Total=31:15

Conclusions

- Diagnostic Performance
  - Area under ROC curve
- Variance representations
- Sizing a study
  - Using Components of Variance
  - Investigating Study designs
    Practical vs. Statistical Efficiency

Future Work

- Include arbitrary study design in iMRMC
- Generalize methods to concordance

Variance in Reader Studies: Methods & Software

- General Regression,Tosteson and Begg (1988)
- The jackknife/ANOVA, Dorfman, Berbaum and Metz (1992)
  - http://metz-roc.uchicago.edu/MetzROC
- ANOVA and correlation model, Obuchowski (1995)
  - http://www.bio.ri.ccf.org/html/rocanalysis.html
- Ordinal Regression, Toledano and Gatsonis (1995)
- Bootstrap, Beiden, Wagner, and Campbell (2000)
- U-statistics, Gallas
  - http://js.cx/~xin/index

Radiologist Variability Example

Beam et al., Arch Intern Med 1996

- 108 US Radiologists
- 79 mammograms:
  - 34 normal/benign
  - 45 breast cancer

- Fully-crossed data
  - Every radiologist read every case

Radiologist Variability Example

Beam et al., Arch Intern Med 1996

- Measurement Scale
  BIRADS: Breast Imaging-Reporting and Data System
  (Ordinal)
  - 1, negative
  - 2, no evidence of malignancy
  - 3, probably benign findings; short-interval follow-up
  - 4, suspicious abnormality; biopsy should be considered
  - 5, high probability of cancer; biopsy recommended

Radiologist Variability Example

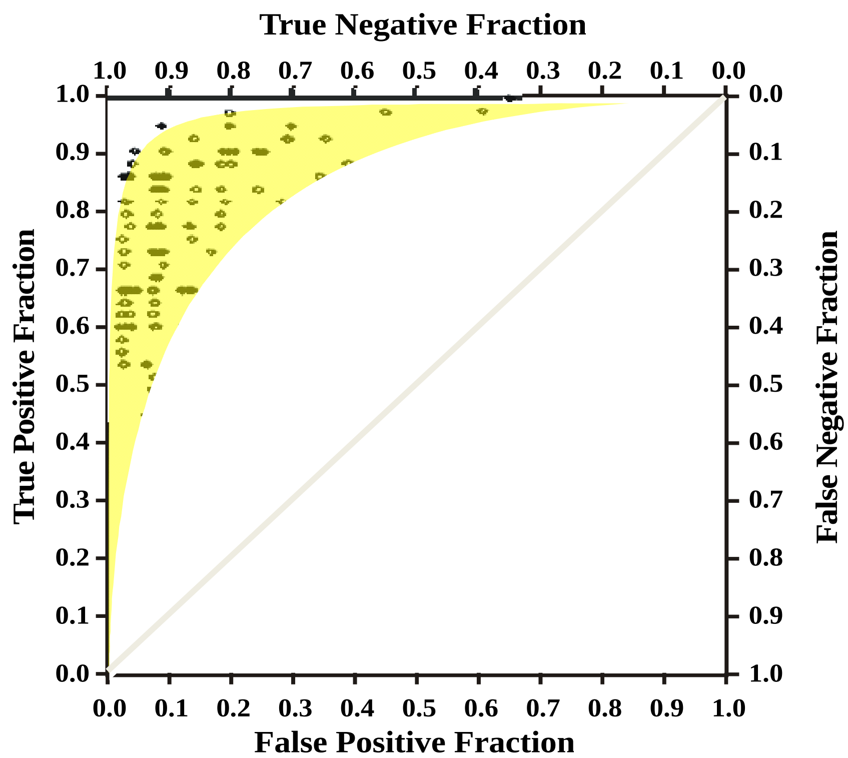
Beam et al., Arch Intern Med 1996

- Measurement Scale BIRADS (Ordinal)
  - 1, negative
  - 2, no evidence of malignancy
  - 3, probably benign findings short-interval follow-up

Threshold

  - 4, suspicious abnormality biopsy should be considered
  - 5, high probability of cancer
  - biopsy recommended

Beam et al., Arch Intern Med 1996



Radiologist Variability Example

Beam et al., Arch Intern Med 1996

	Mean	(Min, Max) Range	Variance Confidence Interval
Sensitivity N=45	80%	(47,100) 53	?
Specificity N=34	90%	(35,99) 63	?
AUC N_{normal}=34 N_{cancer}=45	85%	(74,94) 21	?