PATHOLOGY INFORMATICS SUMMIT 2014
May 14, 2014, Pittsburgh, PA

Reader Studies for Digital Pathology: Software for Simulation, Analysis, and Sizing

Weijie Chen, Adam Wunderlich,
Nicholas Petrick, Brandon D. Gallas

Division of Imaging and Applied Mathematics
Office of Science and Engineering Laboratories
CDRH, FDA

Outline

- Introduction to multi-reader multi-case (MRMC) reader studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory reader studies
- Discussion and summary
- Software freely available: http://code.google.com/p/imrmc/wiki/iMRMC_Binary

Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Digital Pathology: Promises and Challenges

- **Whole Slide Imaging (WSI)**
  - Digitization of glass slides
  - Image viewing on computer monitors
- **Promises**
  - Integration with digital management of tissue samples, reports, etc.
  - Computerized image analysis
  - Telepathology, consulting, education
- **Challenges**
  - Not approved for a general indication
  - Need for consensus on the evaluation methodology

Evaluation of Digital Pathology

- **Technical assessment**
  - Draft guidance under development
- **Clinical effectiveness**
  - Diagnostic performance of pathologists using a device to make diagnosis on patients
  - Comparison between two modalities:
    - *Whole Slide Imaging (WSI)* vs. *Optical microscope (OM)*
    - *WSI (setting 1)* vs. *WSI (setting 2)*
  - A particular paradigm: reader study

vs.

Reader Studies: Overview

- **Multi-reader Multi-case study**
  - Readers: pathologists reading the images or slides; representative of the pathologist population.
  - Cases: patients representative of the patient population.
- **Reading mode**
  - Fully crossed: every reader reads both modalities for every case
  - Alternative designs (to be discussed)
- **Reference standard**
  - E.g., expert pathologist
- **Scoring**
  - Ordinal scale on 0-100 scale (laboratory based)
  - Binary assessment (diseased or not) or converted from clinical report to binary: whether or not in agreement with the reference standard (clinical)

Reader Studies: Data and Endpoint

Modality 1

	Reader 1	…	Reader N_R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_C	0	1	…	1
	PC_{11}	…	PC_{1N_R}	$\widehat{P}_1$

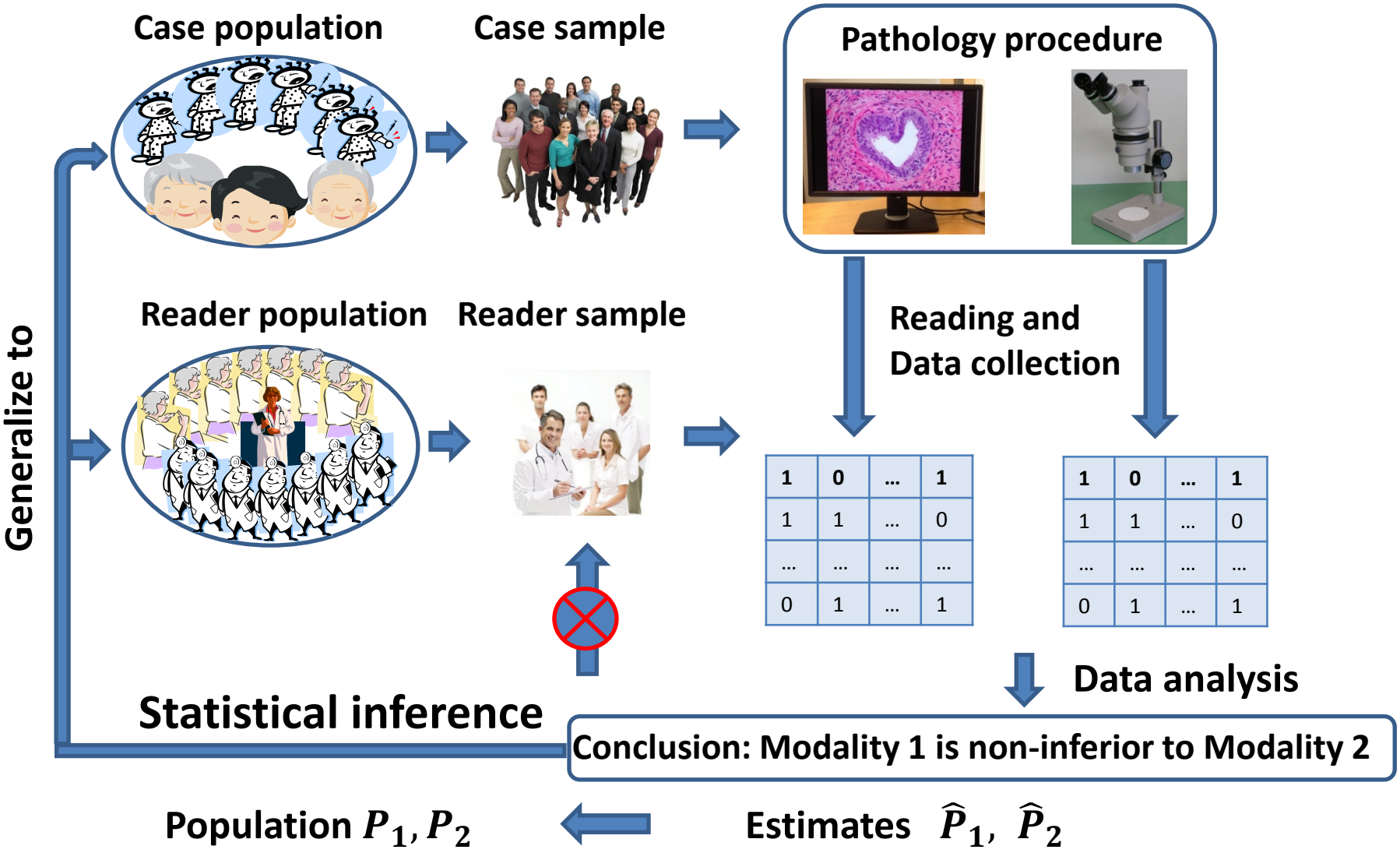
Modality 2

	Reader 1	…	Reader N_R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_C	0	1	…	1
	PC_{21}	…	PC_{2N_R}	$\widehat{P}_2$

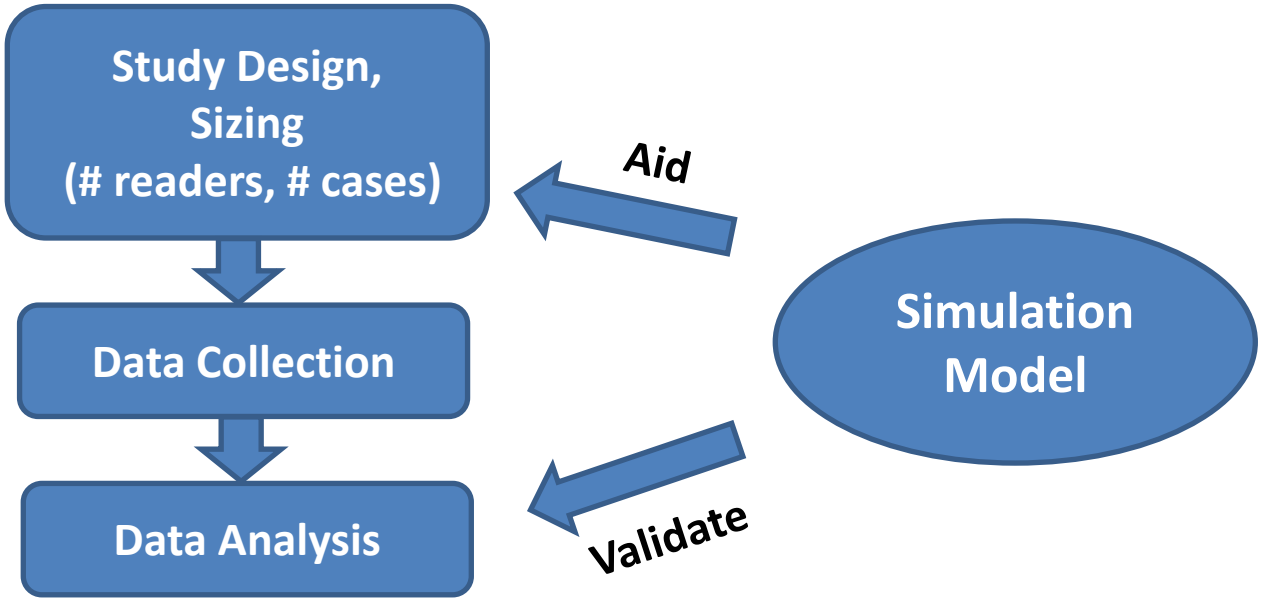
- **Probability of agreement** (a.k.a. percent correct, percentage agreement)

$$\widehat{P}_i = \frac{1}{N_R N_C} \sum_{j=1}^{N_R} \sum_{k=1}^{N_C} Y_{ijk}$$

Reader Studies: Role of Data Analysis



Objectives



- Simulation model to generate binary MRMC study data
- Use of our simulation model for validating a data analysis method
- Use of our simulation model for sizing a study
- Demonstration on a real dataset

MRMC data characteristics

- **Correlations in the MRMC data…**

Modality 1

	Reader 1	…		Reader N_R
Case 1	**1**	**0**	**…**	**1**
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1

Modality 2

	Reader 1	…		Reader N_R
Case 1	**1**	**0**	**…**	**1**
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1

Between modalities: the same reader and the same cases

Modality 1: glass

	Reader 1	…	Reader R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N	0	1	…	1

Modality 2: WSI

	Reader 1	…	Reader R
Case 1	1	1	…	1
	0	1	…	0
…	1	0	…	0
	…	…	…	…
Case N	0	1	…	1

- Correlation between two modalities for the same reader reading the same cases

Between readers: the same cases from the same modality

Modality 1: glass

	Reader 1		…	Reader R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N	0	1	…	1

Modality 2: WSI

	Reader 1		…	Reader R
Case 1	1	1	…	1
	0	1	…	0
…	1	0	…	0
	…	…	…	…
Case N	0	1	…	1

- Correlation between readers reading the same cases for the same modality

Between readers: the same cases from different modalities

Modality 1: glass

	Reader 1	…	Reader R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N	0	1	…	1

Modality 2: WSI

	Reader 1	…	Reader R
Case 1	1	1	…	1
	0	1	…	0
…	1	0	…	0
	…	…	…	…
Case N	0	1	…	1

- Correlation between readers reading the same cases for different modalities

Between cases: the same readers for the same modality

Modality 1: glass

	Reader 1	…	Reader R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N	0	1	…	1

Modality 2: WSI

	Reader 1	…	Reader R
Case 1	1	1	…	1
	0	1	…	0
…	1	0	…	0
	…	…	…	…
Case N	0	1	…	1

- Correlation between different cases read by the same set of readers for the same modality

Between cases: the same readers for different modalities

Modality 1: glass

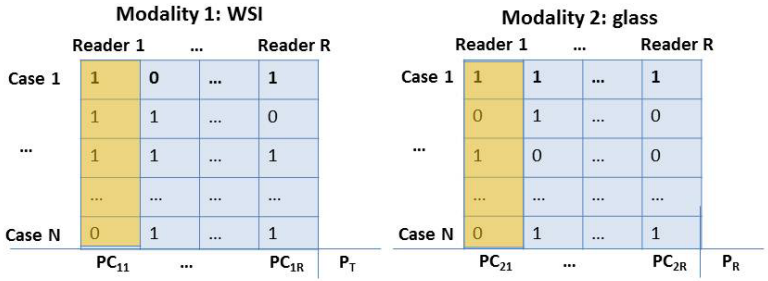
	Reader 1	…	Reader R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N	0	1	…	1

Modality 2: WSI

	Reader 1	…	Reader R
Case 1	1	1	…	1
	0	1	…	0
…	1	0	…	0
	…	…	…	…
Case N	0	1	…	1

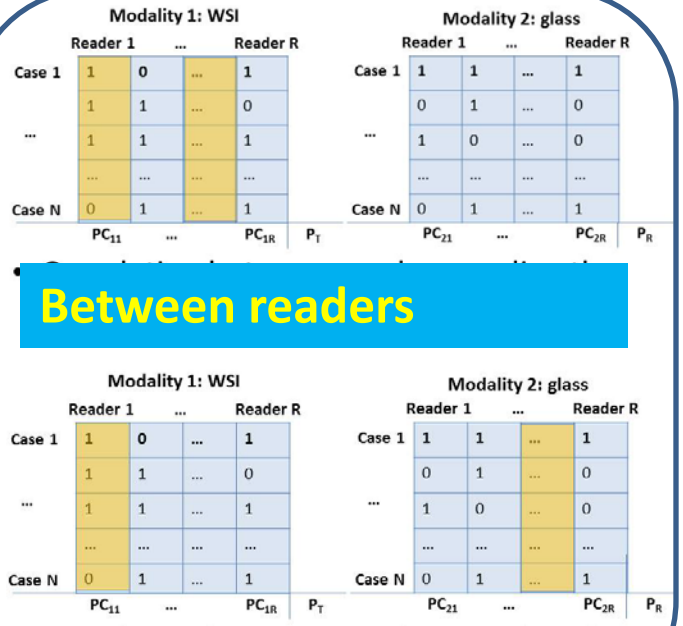
- Correlation between different cases read by the same set of readers for different modalities

A Summary View



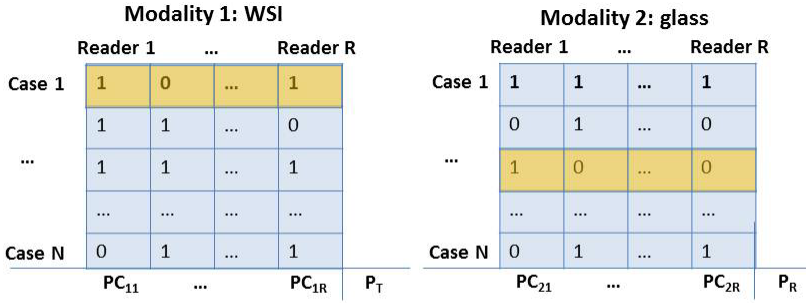
- Corr ing the s

Between modalities



Between readers

- Correlation between readers reading the same cases for different modalities



- Correlation between different ca een different cases read by the same reader for the same m for different modalities

Between cases

Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Simulation

- What characteristics of the data do we need to capture?
  - Correlations in the MRMC data
  - Probability of agreement for each modality

Modality 1

	Reader 1	…	Reader N_R
Case 1	**1**	**0**	…	**1**
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1

Modality 2

	Reader 1	…	Reader N_R
Case 1	**1**	**0**	…	**1**
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1

Simulation: Threshold Model

- **Latent continuous variable**
  For modality i, reader j, and case k,

$$X_{ijk} = \tau_i + R_j + C_k + (\tau R)_{ij} + (\tau C)_{ik} + (RC)_{jk} + e_{ijk}$$

  modality τ_i is a fixed effect

  reader R_j is a random effect, $\sim N(0, {\sigma_R}^2)$

  case C_k is a random effect, $\sim N(0, {\sigma_C}^2)$

  $(\tau R)_{ij}$ is a random effect, $\sim N(0, {\sigma_{\tau R}}^2)$

  $(\tau C)_{ik}$ is a random effect, $\sim N(0, {\sigma_{\tau C}}^2)$

  $(RC)_{jk}$ is a random effect, $\sim N(0, {\sigma_{RC}}^2)$

  e_{ijk} is a random error, $\sim N(0, {\sigma_e}^2)$

- **Apply a threshold to dichotomize**: $Y_{ijk} = I(X_{ijk} > 0)$

Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Analysis Methods: Literature Review

- **(Incomplete) Literature for MRMC ROC studies**
  - Dorfman-Berbaum-Metz (DBM) (1992): jackknife, three-way mixed-effect ANOVA model
  - Obuchowski-Rockette (OR) (1995): two-way mixed-effect ANOVA model
  - Hillis (2007, 2011): unified the DBM and OR methods, refined, call it ORH
  - Beiden-Wagner-Campbell (BWC) (2000): use the bootstrap to estimate the variance components in DBM
  - Gallas (2006): U-statistic, nonparametric
- **Methods can be adapted to binary data**

Analysis Methods: Hypotheses

- **To establish Modality 1 is non-inferior to Modality 2**

Modality 1

	Reader 1	…	Reader N_R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1
	PC_{11}	…	PC_{1N_R}	$\widehat{P}_1$

Modality 2

	Reader 1	…	Reader N_R
Case 1	1	0	…	1
	1	1	…	0
…	1	1	…	1
	…	…	…	…
Case N_c	0	1	…	1
	PC_{21}	…	PC_{2N_R}	$\widehat{P}_2$

Null hypothesis H_0: $P_1 - P_2 \leq -\delta$

Alternative hypothesis H_1: $P_1 - P_2 > -\delta$

δ: non-inferiority margin

Calculate the 95% CI of $\widehat{P}_1 - \widehat{P}_2$ as [LB, UB]
Claim success of LB > $-\delta$

Analysis Methods: ORH and Gallas

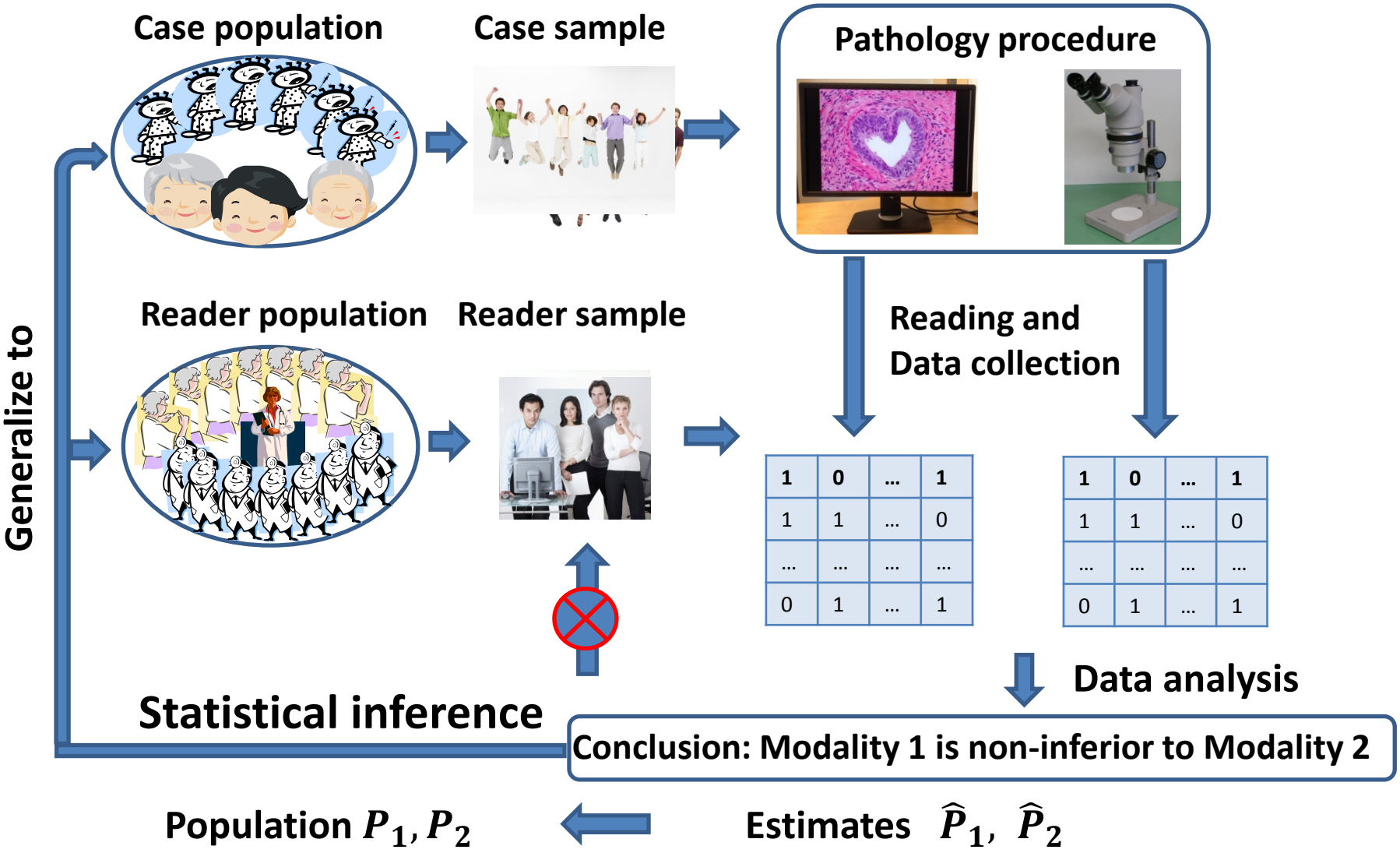
- Both methods use a *t* statistic to get the 95% CI of $\widehat{\boldsymbol{P}}_{\boldsymbol{1}} - \widehat{\boldsymbol{P}}_{\boldsymbol{2}}$

$$\boldsymbol{t} = \frac{\widehat{\boldsymbol{P}}_{\boldsymbol{1}} - \widehat{\boldsymbol{P}}_{\boldsymbol{2}}}{\sqrt{\widehat{\boldsymbol{Var}}(\widehat{\boldsymbol{P}}_{\boldsymbol{1}} - \widehat{\boldsymbol{P}}_{\boldsymbol{2}})}}$$

- In both methods, the variance accounts for
  - Randomness of the case sample (case variability)
  - Randomness of the reader sample (reader variability)

$$\boldsymbol{VAR_{Total}} = \frac{\boldsymbol{VAR_{Case}}}{\boldsymbol{N_C}} + \frac{\boldsymbol{VAR_{Reader}}}{\boldsymbol{N_R}} + \frac{\boldsymbol{VAR_{error}}}{\boldsymbol{N_C N_R}}$$

Reader Studies: Sources of Variability



Analysis Methods: ORH and Gallas

- Variance allows the conclusions:
  - to generalize to both the case (patient) population and the reader (pathologist) population
  - not conditioned on the particular reader sample or the patient sample being studied
- The two methods differ in
  - Variance estimation
    - ORH: based on ANOVA model
    - Gallas: non-parametric, U-statistic based
  - Degree of freedom of the *t* statistic (see references)

Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Validation of Analysis Methods

- Specify population parameters in the simulation model; specify sample size (# readers, # cases)
- Repeat (e.g., 10,000 times)
    - Draw a dataset using the simulation model with the specified parameters;
    - Apply the analysis method, calculate the 95% CI of $\widehat{\boldsymbol{P}}_{\boldsymbol{1}} - \widehat{\boldsymbol{P}}_{\boldsymbol{2}}$;
    - Check if the estimated 95% CI covers the true modality difference of the performance $\boldsymbol{P}_{\boldsymbol{1}} - \boldsymbol{P}_{\boldsymbol{2}}$
  - Coverage probability should be 95% for a valid method
  - The procedure can be repeated by specifying different parameters

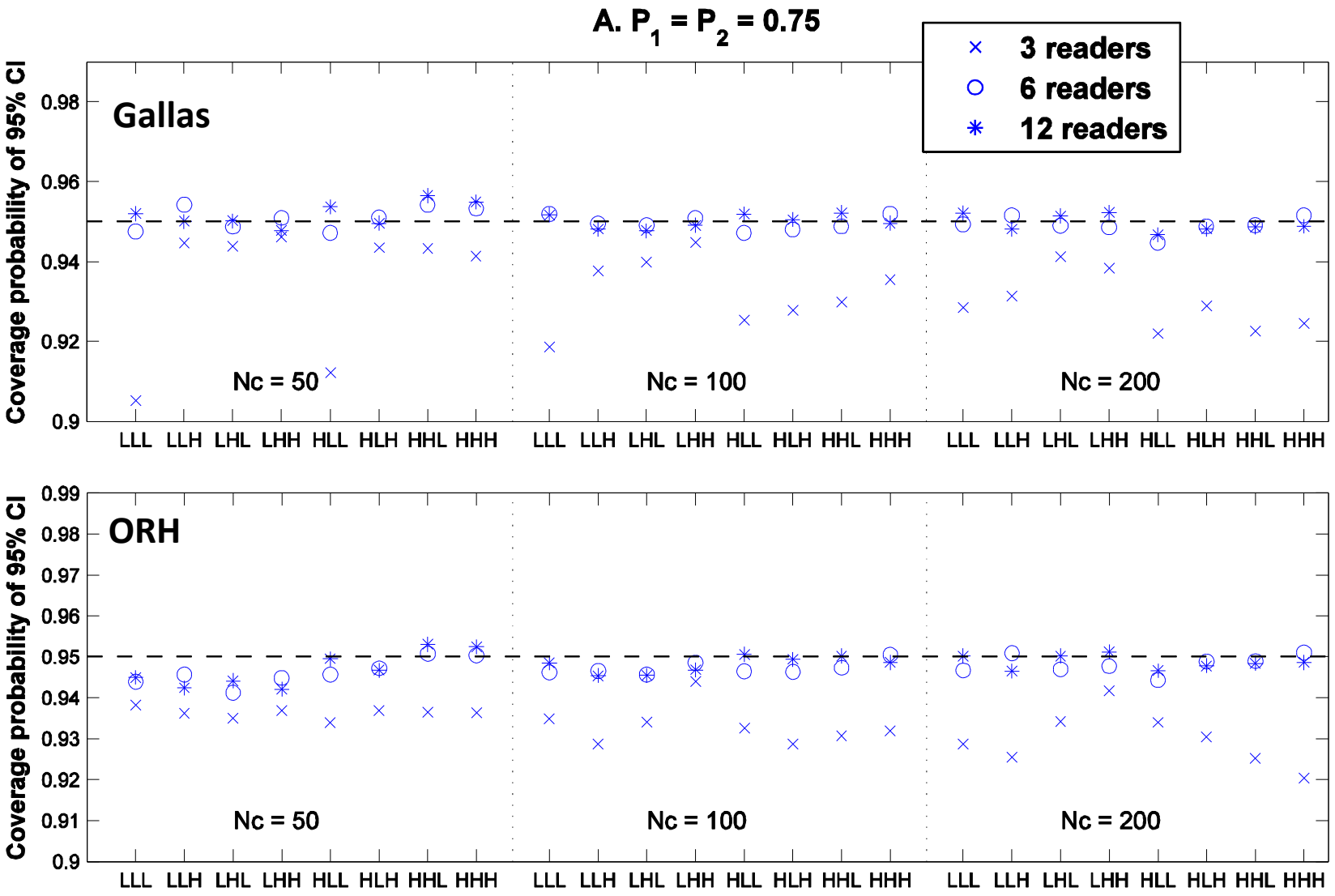
Parameters chosen for validation

- **Expected performance for each modality**: $P_1 = P_2 = 0.75, 0.85, 0.95$
- **Number of readers**: $N_R = 3, 6, 12$
- **Number of cases**: $N_c = 50, 100, 200$
- **Correlation parameters**

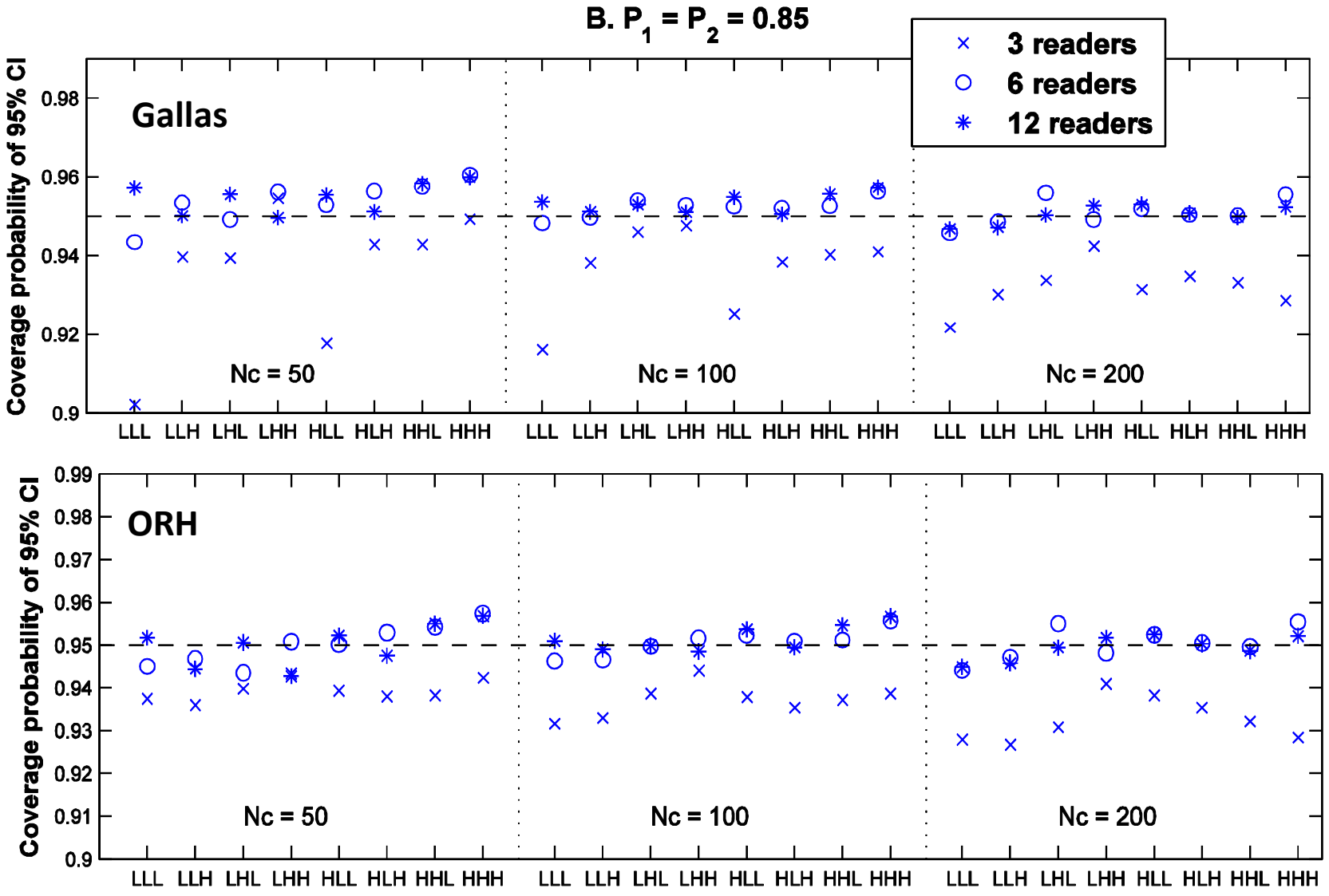
Structure	ρ_c	ρ_{tc}	ρ_r	ρ_{tr}	ρ_t
LLL	0.006	0.005	0.240	0.200	0.300
LLH	0.008	0.005	0.320	0.200	0.300
LHL	0.006	0.005	0.500	0.400	0.500
LHH	0.008	0.005	0.600	0.400	0.500
HLL	0.040	0.030	0.240	0.200	0.300
HLH	0.050	0.030	0.320	0.200	0.300
HHL	0.040	0.030	0.500	0.400	0.500
HHH	0.050	0.030	0.600	0.400	0.500

L: relatively low
H: relatively high

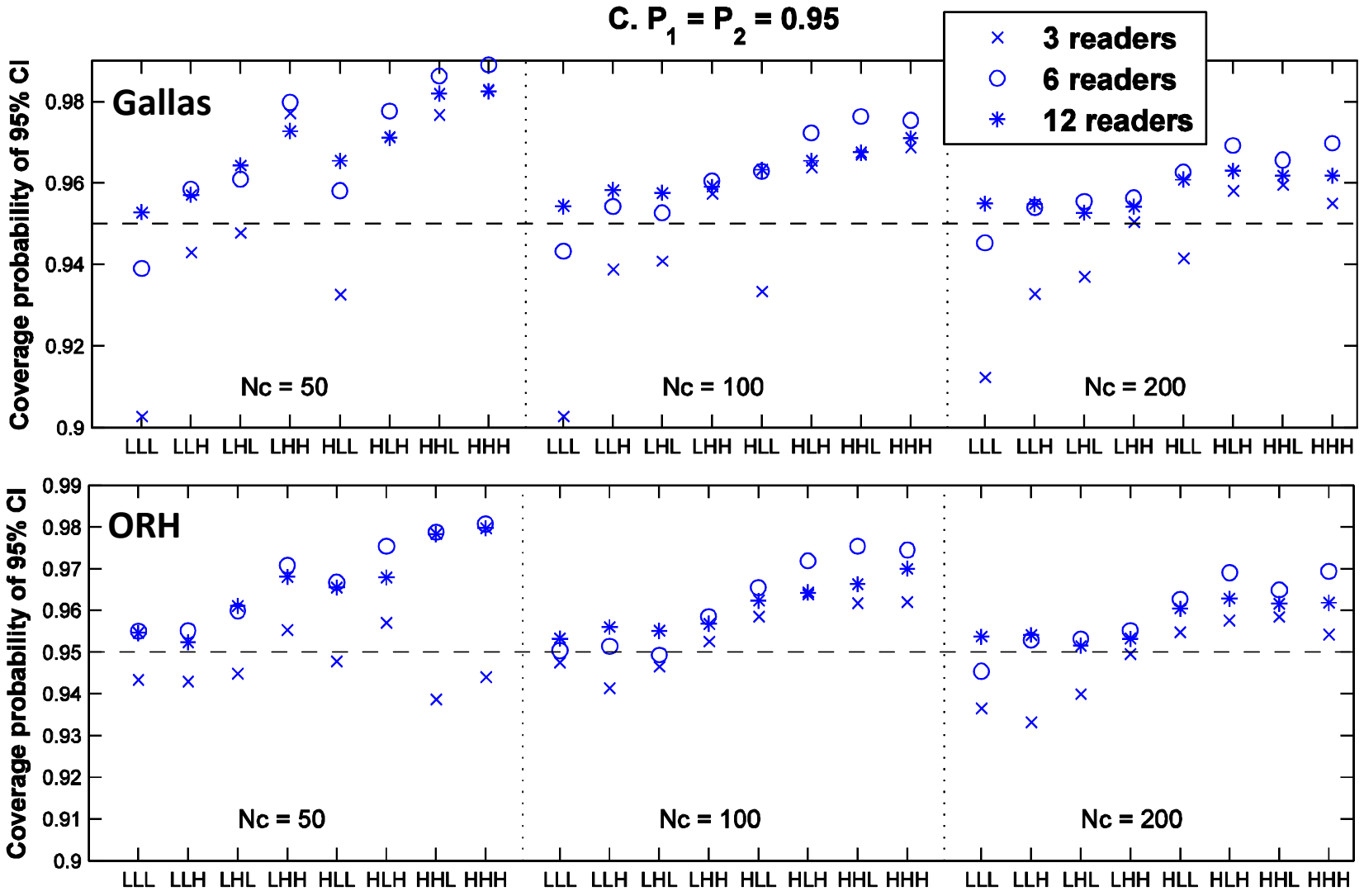
Validation Results



Validation Results



Validation Results



Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Simulation Method for Sizing

- Measure the correlation and probability of agreement parameters in a pilot study.
- Generate a large number of datasets for specified # readers and # cases.
- For each dataset, apply an analysis method to see if the alternative hypothesis is established for a given non-inferiority margin (e.g., WSI is non-inferior to OM).
- Power = Fraction of success.
- Iteratively change # readers and # cases until desired power is achieved.

Sizing: a demo

- Expected performance for each modality: $P_1 = P_2 = 0.80, 0.90$;
- Number of readers: $N_r = 6, 12$;
- Non-inferiority margin: $\delta = 0.03, 0.05$;
- Correlation structure: LHH and LHL
- Results: number of cases needed to achieve 80% power

		$P_1 = P_2 = 0.80$		$P_1 = P_2 = 0.90$	
		$N_r = 6$	$N_r = 12$	$N_r = 6$	$N_r = 12$
$\delta = 0.03$	LHH	>10,000	3,200	>10,000	610
	LHL	1483	493	444	244
$\delta = 0.05$	LHH	889	305	216	150
	LHL	240	150	115	81

Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

A real example: laboratory MRMC studies

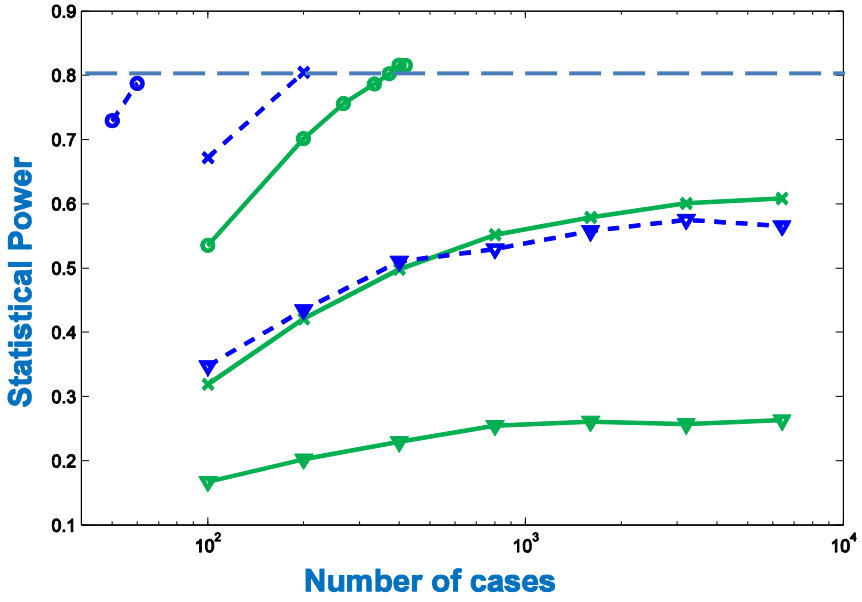
- Task: differentiate individual cells
  - Mitotic Figures or not
  - Plasma Cells or not
- Modalities: WSI with display parameter γ = 1.0 vs. γ = 1.8 (manufacturer default)
- Reference standard: one expert pathologist reading optical microscope
- Pilot study: 5 readers, 50 cases
- Goal: WSI with setting γ = 1.0 is non-inferior to γ = 1.8
- Brandon Gallas, 8:00-9:00 am 05/15, Ballroom 1

Task 1: Mitotic figures
Gamma = 1.0 vs. Gamma=1.8

- **Analysis result**

	R1	R2	R3	R4	R5	Average	Difference (SD)	95% CI
γ = 1.0	0.74	0.76	0.78	0.78	0.84	0.780	0.004 (0.0346)	[-0.08, 0.09]
γ = 1.8	0.84	0.70	0.82	0.74	0.78	0.776		

- **Sizing**

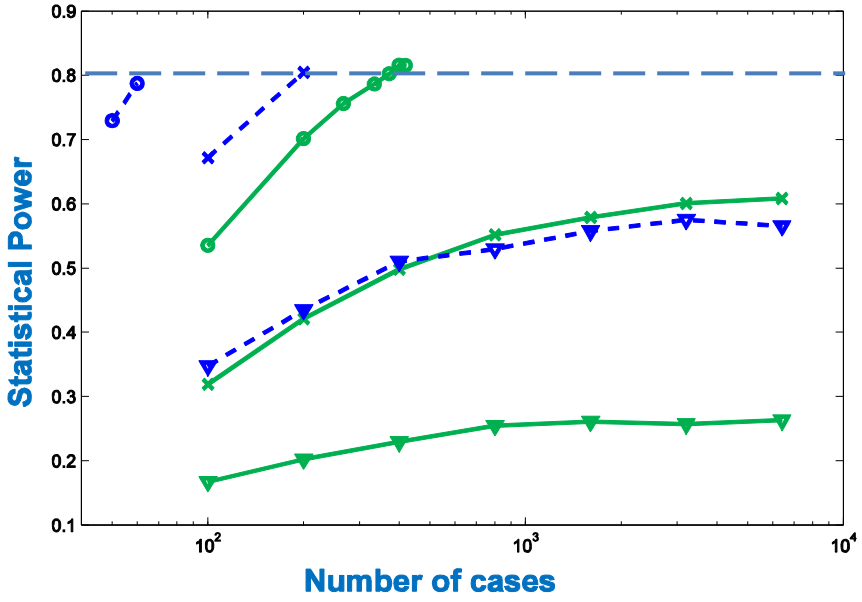


	# cases to achieve 80% power
5 readers, NIM = 0.03	Never
10 readers, NIM = 0.03	Never
20 readers, NIM = 0.03	372
5 readers, NIM = 0.05	Never
10 readers, NIM = 0.05	200
20 readers, NIM = 0.05	60

Task 1: Mitotic figures
Gamma = 1.0 vs. Gamma=1.8

$$VAR_{Total} = \frac{VAR_{Case}}{N_C} + \frac{VAR_{Reader}}{N_R} + \frac{VAR_{error}}{N_C N_R}$$

- **Sizing**



	# cases to achieve 80% power
5 readers, NIM = 0.03	Never
10 readers, NIM = 0.03	Never
20 readers, NIM = 0.03	372
5 readers, NIM = 0.05	Never
10 readers, NIM = 0.05	200
20 readers, NIM = 0.05	60

Task 2: Plasma cells
Gamma = 1.0 vs. Gamma=1.8

- **Analysis result**

	R1	R2	R3	R4	R5	Average	Difference (SD)	95% CI
γ = 1.0	0.84	0.74	0.60	0.76	0.94	0.776	0.032 (0.0162)	[-0.01, 0.07]
γ = 1.8	0.76	0.72	0.62	0.72	0.90	0.744		

- **Sizing**: not needed; already a success.

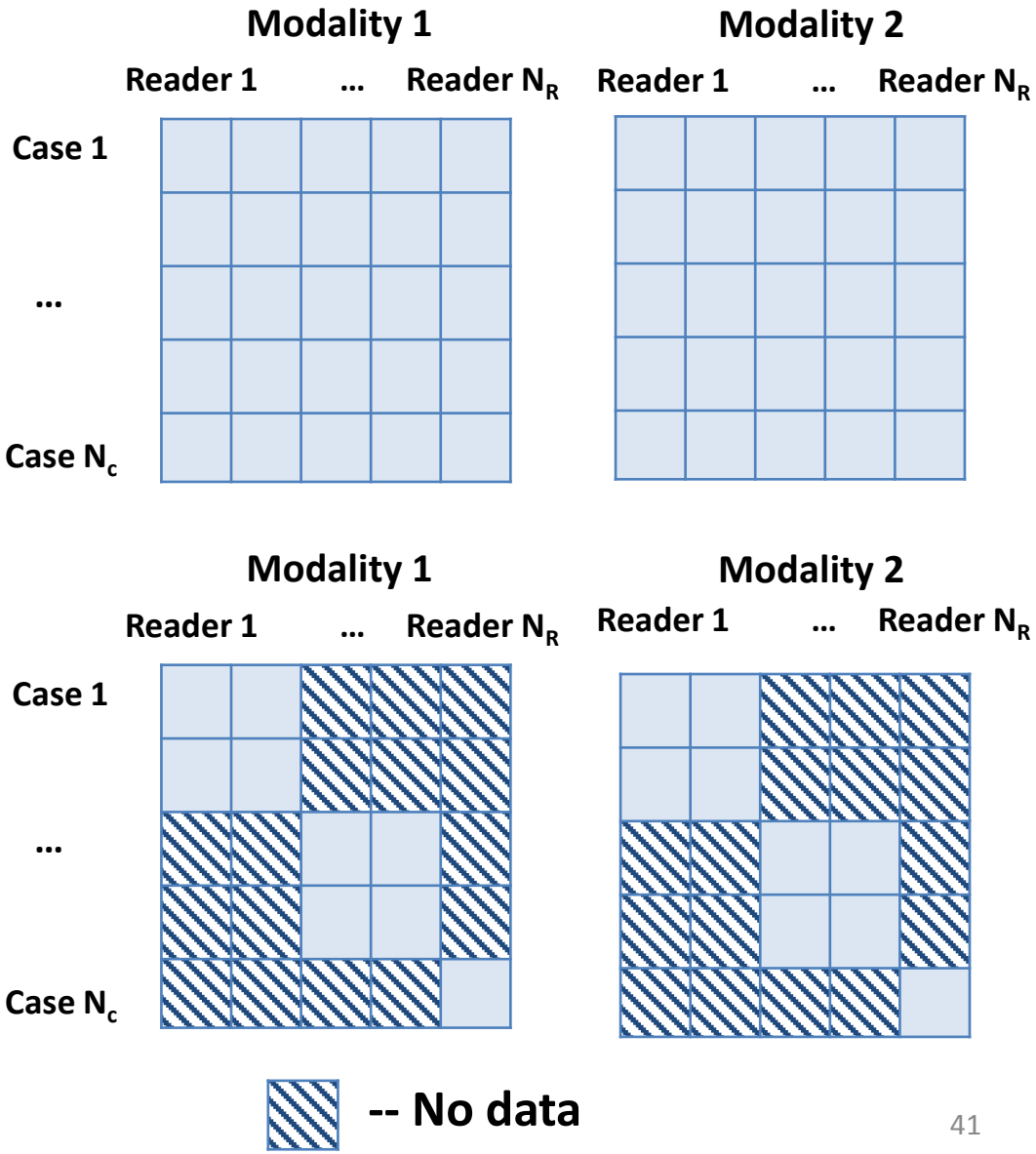
Outline

- Introduction to multi-reader multi-case reader (MRMC) studies
- Simulation model: binary data
- Introduction to analysis methods
- Use of simulation model for validation of analysis methods
- Use of simulation model for sizing a new study
- A real example: laboratory MRMC studies
- Discussion and summary

Discussions

- **Study design**: assumed to be **fully crossed**
  - Every reader reads every case for both modalities
- Alternative: **Split-plot design**
  - Readers read their own group of cases

    Obuchowski, Gallas, Hillis, Acad Radiol 2012



Summary

- Simulation model for simulating binary data in MRMC reader studies
  - Validation of an analysis method
  - Aid sizing a study
- Analysis methods
  - Account for both case and reader variability
  - Literature
  - ORH and Gallas methods: statistical properties investigated
- Software (1.0Beta) freely available: http://code.google.com/p/imrmc/wiki/iMRMC_Binary (a new version will be released by the end of summer)

References

- W. Chen, A. Wunderlich, N. Petrick, B. Gallas. “A general framework for MRMC reader studies with binary assessments: simulation, analysis, validation, and sizing.” SPIE Journal of Medical Imaging, under review.
- D. D. Dorfman, K. S. Berbaum, and C. E. Metz, “Receiver operating characteristic rating analysis - generalization to the population of readers and patients with the jackknife method,” Investigative Radiology 27(9), 723–731 (1992).
- N. Obuchowski and H. E. Rockette, “Hypothesis testing of diagnostic accuracy for multiple readers and multiple tests: An ANOVA approach with dependent observations,” Communications in Statistics-Simulation and Computation 24(2), 285–308 (1995).
- B. D. Gallas, “One-shot estimate of MRMC variance: AUC,” Acad Radiol 13(3), 353–362 (2006).
- S. L. Hillis, “A comparison of denominator degrees of freedom methods for multiple observer ROC analysis,” Stat Med 26(3), 596–619 (2007).
- S. L. Hillis, “A marginal-mean ANOVA approach for analyzing multireader multicase radiological imaging data,” Stat Med 33, 330–360 (2014).
- W. Chen, N. Petrick, and B. Sahiner, “Hypothesis testing in noninferiority and equivalence MRMC ROC studies.,” Acad Radiol 19, 1158–1165 (2012).
- N. Obuchowski, B. D. Gallas, and S. L. Hillis, “Multi-reader ROC studies with split-plot designs: A comparison of statistical methods,” Acad Radiol 19, 1508–1517 (2012).

Thank you

- Questions?
- Contact:

Weijie Chen, Ph.D.
Division of Imaging and Applied Mathematics
Office of Science and Engineering Laboratories
FDA/CDRH
10903 New Hampshire Avenue, WO62-4104
Silver Spring, MD 20993-0002
301-7962663 (phone) 301-796-9925 (fax)
weijie.chen@fda.hhs.gov