



How do we decide how many readers
and cases we need in a reader study?

S Zabala Travers, MD; B Gallas, PhD; WC Cheng, PhD;
T Kimpe; A Badano, PhD.

RSNA Annual Meeting, 2015.

Importance

Results should be generalizable to new readers and new cases.

Pilot study

The study analysis needs to account for:

- ✓ reader variability
- ✓ case variability

Precision

iMRMC

1- Load data

Results should be conclusive and reproducible.

2 - Analyze

3 - Size

Sample sizes of readers and cases controls precision.

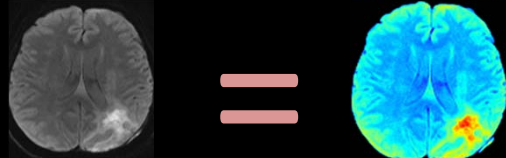
Summary

Is there an effect of color?



HYPOTHESIS TEST

Null Hypothesis:
There is no effect of color.



Performance
with gray scale

Performance with
rainbow scale

Alternative Hypothesis:
There is an effect of color.



Performance
with gray scale

Performance with
rainbow scale

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Zabala-Travers *et al*, *Med Physics*, Jun 2015

Is there an effect of color?



PILOT READER STUDY

Importance

Pilot study

Precision

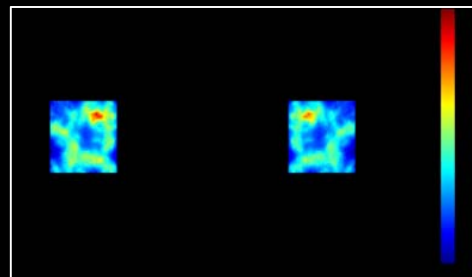
iMRMC

1- Load data

2 - Analyze

3 - Size

Summary



Reading session, user interface

600 CASES

17 READERS

- ✓ 2 alternative forced choice
- ✓ *Which image has the higher intensity?*
- ✓ We scored number of correct answers
- ✓ 2 sessions using same set of cases with a different color scale
- ✓ We compared percentage of correct answers between both modalities.

Zabala-Travers et al, Med Physics, Jun 2015

Results

Importance

Pilot study

Precision

iMRMC

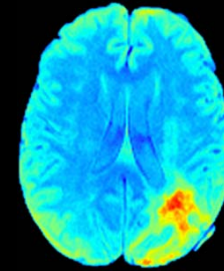
1- Load data

2 - Analyze

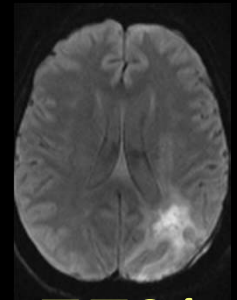
3 - Size

Summary

Average percent correct (PC) of answers for all readers and cases:



95%



77%

EFFECT SIZE:

Difference in performance (Δ PC)

18%

Zabala-Travers *et al*, *Med Physics*, Jun 2015

Results

Importance

Pilot study

Precision

iMRMC

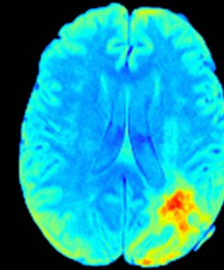
1- Load data

2 - Analyze

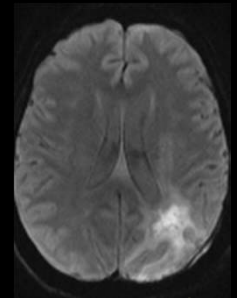
3 - Size

Summary

Average percent correct (PC) of answers for all readers and cases:



95%



77%

EFFECT SIZE:

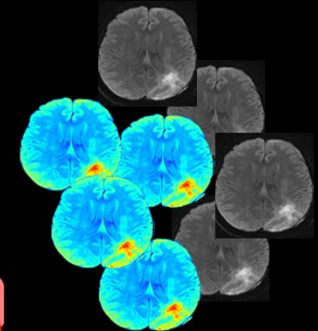
Difference in performance (Δ PC)

18%

Results are an average!

There is variability coming from:

- ✓ Readers with different skills
- ✓ Cases of different difficulty
- ✓ The interaction of readers and cases



Zabala-Travers *et al*, *Med Physics*, Jun 2015

Results

Importance

Pilot study

Precision

iMRMC

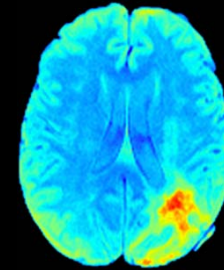
1- Load data

2 - Analyze

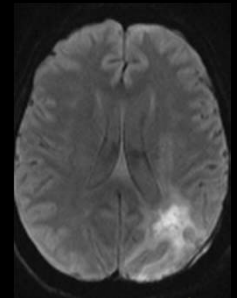
3 - Size

Summary

Average percent correct (PC) of answers for all readers and cases:



95%



77%

EFFECT SIZE:

Difference in performance (ΔPC)

18%

PRECISION or VARIABILITY:

Standard error (SE)

6%

95% confidence interval (CI = $\Delta PC \pm 2SE$)

CI (6,30)

HYPOTHESIS TEST:

p - value

p < 0.05

Zabala-Travers *et al*, *Med Physics*, Jun 2015

Importance

Pilot study

Precision

iMRMC

1- Load data

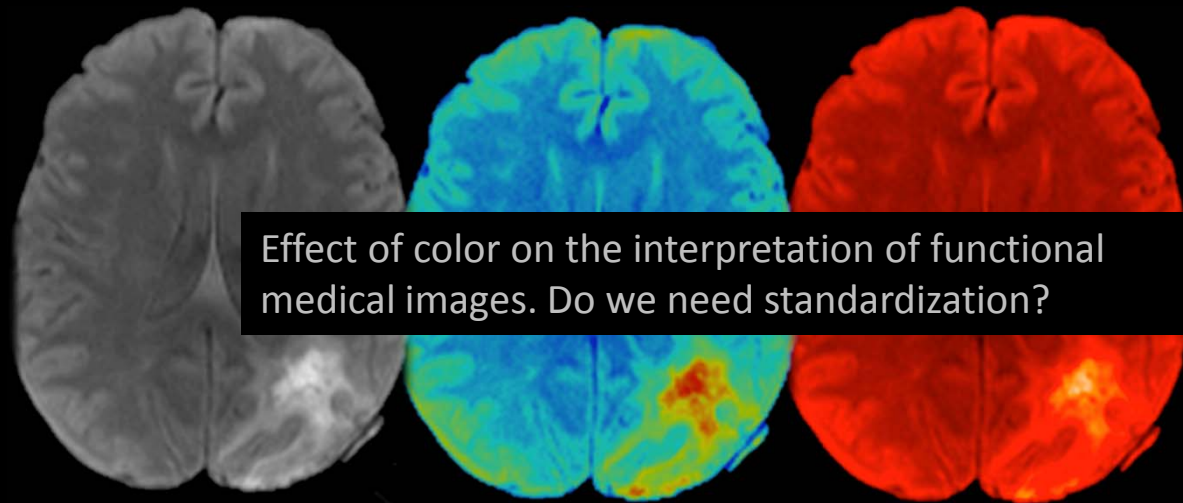
2 - Analyze

3 - Size

Summary

Conclusion of the pilot study:

- Color effect on reader performance is worth investigating in the clinical setting.



Ongoing multi – institutional reader study.

- How many readers and cases are necessary to study this effect in a clinical setting?

The sample size is a balance between resources and precision.

Importance

Pilot study

Precision

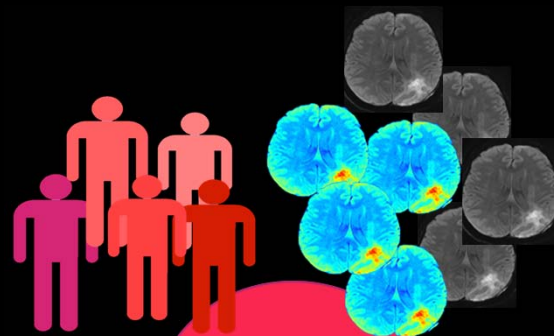
iMRMC

1- Load data

2 - Analyze

3 - Size

Summary



Precision

"Costs"

As a general rule:
If we want:
✓ Higher precision
✓ Less variability
We need
✓ A larger sample size

Larger sample size implies:
✓ Longer reading times
✓ Larger number of readers and cases

Variability: readers' component.

Importance

Pilot study

Precision

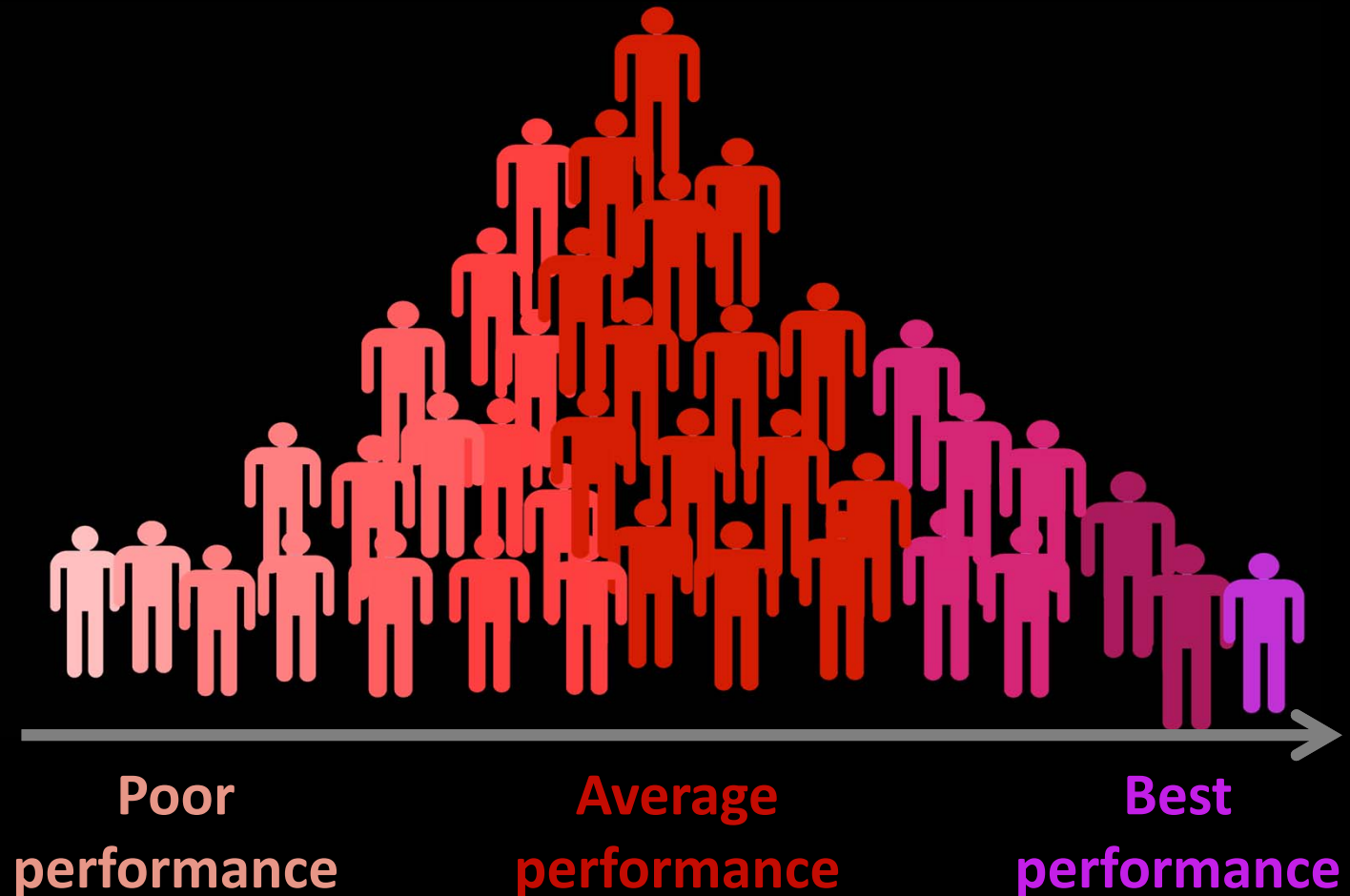
iMRMC

1- Load data

2 - Analyze

3 - Size

Summary



Variability: readers' component.

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Study 1:
Sample size: 3 readers

Study 2:
Sample size: 7 readers

Result 1:

Excellent performance



Result 2:

Average performance



Result 3:

Average to bad performance



Result 1:

Average performance



Result 2:

Average to good performance



Result 3:

average performance

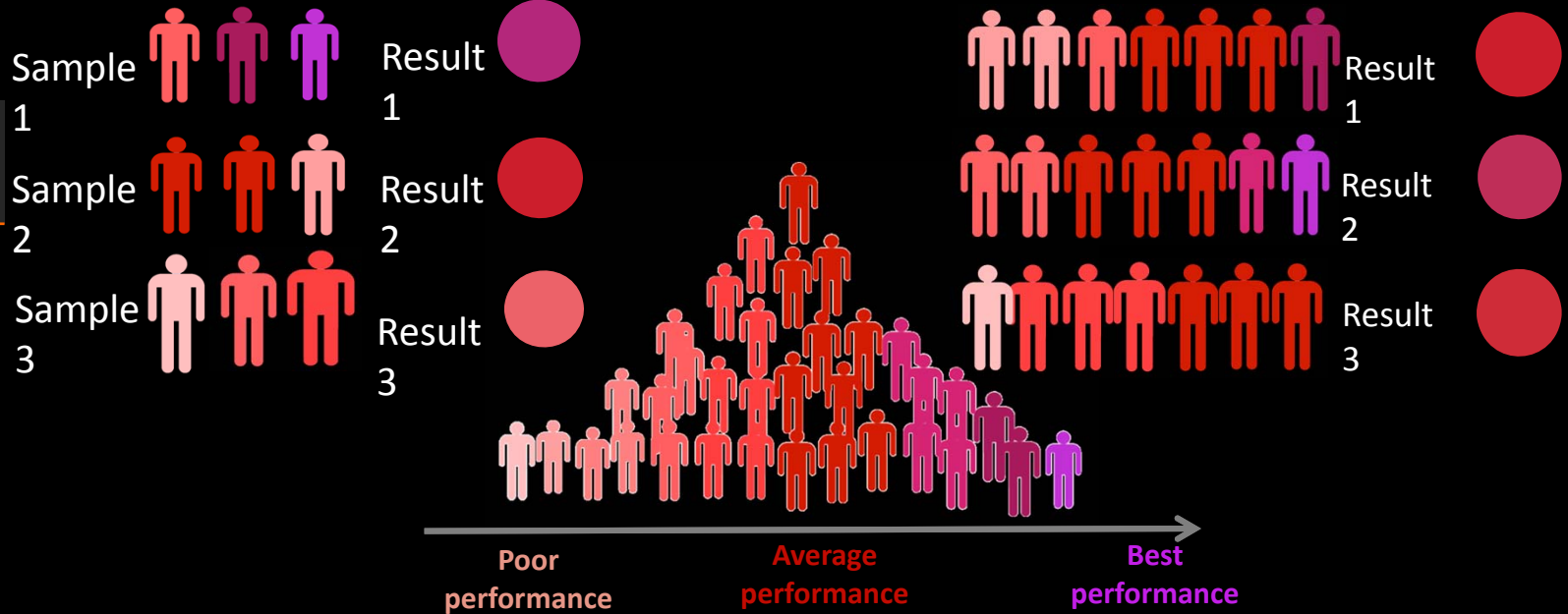


Variability: readers' component.

Importance

A larger number of readers and cases decreases variability.

Pilot study



Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

iMRMC software tool: 3 steps

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

The screenshot shows the iMRMC software interface with three main sections highlighted by colored boxes:

- Step 1: LOAD DATA FILE** (Green box): Shows the 'Help and Info' window with a text input field containing '.imrmc file' and a file path 'alysed/MRMC_LL_Eizojetvsgray.imrmc'. There are 'Browse...' and 'Input Statistics' buttons.
- Step 2: STATISTICAL ANALYSIS** (Red box): Shows the 'MRMC Variance Analysis (Difference)' window. It includes a hypothesis test: $H_0: AUC_A - AUC_B = 0.00$, two-sided alternative, 95% significance, 8 Readers. Results show $AUC_A = 0.954$, $AUC_B = 0.774$, $AUC_A - AUC_B = 0.179$, $\sqrt{\text{total var}} = 0.047$. The p-value is 0.0001. A table of component means and variances is also visible.
- Step 3: SAMPLE SIZING** (Yellow box): Shows the 'Sizing Analysis' window. It includes input fields for 'Significance level' (0.05), 'Effect Size' (0.05), '#Reader' (8), and '#Normal' (8). It displays the formula for the Normal Approximation: $df = \frac{8 \times 0.047^2}{0.05^2}$.

	M1	M2	M3	M4	M5	M6	M7	M8	
comp M0	9.53672E...	9.53672E...	9.10070E...	9.10070E...	9.13446E...	9.13446E...	9.09332E...	9.09332E...	
coeff M0	1.91516E...	5.63057E...	5.63057E...	1.65539E...	9.57579E...	2.81528E...	2.81528E...	-1.72307...	
comp M1	7.74334E...	7.74334E...	6.01588E...	6.01588E...	6.07748E...	6.07748E...	5.99134E...	5.99134E...	
coeff M1	1.64156E...	4.82620E...	4.82620E...	1.41890E...	9.84939E...	2.89572E...	2.89572E...	-1.48658...	
comp prod...	7.34915E...	7.34915E...	7.32418E...	7.32418E...	7.43930E...	7.43930E...	7.39261E...	7.39261E...	
-coeff prod...	2.73594E...	8.04367E...	8.04367E...	2.36484E...	2.02460E...	5.95231E...	5.95231E...	-2.50020...	
total	1.08687E...	3.19539E...	2.13627E...	6.28062E...	-3.28666...	-9.66279...	-1.05360...	-6.09203...	total var=2.218E-3

It's free and open source! Download software and guidelines here:
<http://github.com/DISDR/iMRMC/releases/tag/iMRMC-v2p7>

iMRMC software tool: 3 steps

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Step 1:
LOAD DATA FILE

Input file for the iMRMC:
✓ .imrmc or .csv files

N1	600	#Number of lesion present cases	
N0	600	#Number of lesion absent cases	
NR	17	#Number of readers	
NM	2	#Number of modalities	
#READER ID	#CASE ID	#MODALITY	#SCORE
BEGIN DATA:			
Reader1	Case1	grayscale	1
Reader1	Case2	grayscale	1
Reader1	Case3	grayscale	0
Reader1	Case4	grayscale	1

It's free and open source! Download software and guidelines here:
<http://github.com/DISDR/iMRMC/releases/tag/iMRMC-v2p7>

iMRMC software tool: 3 steps

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

MLE (avoid negatives) 1 rainbow 2 gray MRMC Variance Analysis (Difference) Show Reader AUCs

H0: AUC_A - AUC_B = 0.00, two-sided alternative, 95% significance, 8 Readers, 295 Normal cases, 295 Disease cases.

AUC_A = 0.954, AUC_B = 0.774, AUC_A - AUC_B = 0.179, sqrt(total var) = 4.709E-2, T Statistic = 3.808E0

T-stat df(Normal Approx) = ∞ p-Value = 0.0001 Conf. Int. = (0.0870, 0.2716) Reject Null? = 1.0000

df(BDG) = 5.50 p-Value = 0.0125 Conf. Int. = (0.0583, 0.3004) Reject Null? = 1.0000

BDG BCK DBM OR MS

	M3	M4	M5	M6	M7	M8
coeff M1	1.64156E...	4.82620E...	4.82620E...	1.41890E...	9.84939E...	2.89572E...
coeff M2	1.64156E...	4.82620E...	4.82620E...	1.41890E...	9.84939E...	2.89572E...
comp prod...	7.34915E...	7.34915E...	7.32418E...	7.32418E...	7.43930E...	7.43930E...
-coeff prod...	2.73594E...	8.04367E...	8.04367E...	2.36484E...	2.02460E...	5.95215E...
total	1.08687E...	3.19539E...	2.13627E...	6.28062E...	-3.28666E...	-9.66111E...

total var=2.218E-3

Step 2: STATISTICAL ANALYSIS

It's free and open source! Download software and guidelines here:
<http://github.com/DISDR/iMRMC/releases/tag/iMRMC-v2p7>

iMRMC software tool: 3 steps

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

The POWER of a reader study is the probability of rejecting the null hypothesis when it is false

It is the likelihood that we will detect a difference in performance when it exists

Power improves when precision improves

**Step 3:
SAMPLE SIZING**

Significance level 0.05 Effect Size 0.05 #Reader 8 #Normal 295 #Diseased 295 Size a Trial

Sizing Analysis: SE = , Test Stat=

Normal Approx: df= ∞ , Power=

BDG: df= , Lambda= Power=

Hillis 2011: df= , Lambda= , Power=

It's free and open source! Download software and guidelines here:
<http://github.com/DISDR/iMRMC/releases/tag/iMRMC-v2p7>

iMRMC software tool

STEP 3: Sample sizing



Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Significance level Effect Size #Reader #Normal #Diseased

Sizing Analysis: SE = 5.194E-2, Stat= 3.466E0
Normal Approx: df= ∞, Power= 0.93
BDG: df= 10.41, Lambda= 12.01, Power= 0.88
Hillis 2011: df= 7.00, Lambda= 0.17, Power= 0.07

Significance level Effect Size #Reader #Normal #Diseased

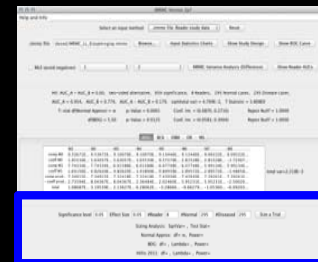
Sizing Analysis: SE = 4.660E-2, Stat= 3.863E0
Normal Approx: df= ∞, Power= 0.97
BDG: df= 14.12, Lambda= 14.92, Power= 0.95
Hillis 2011: df= 9.01, Lambda= 0.21, Power= 0.07

Significance level Effect Size #Reader #Normal #Diseased

Sizing Analysis: SE = 4.393E-2, Stat= 4.098E0
Normal Approx: df= ∞, Power= 0.98
BDG: df= 11.44, Lambda= 16.79, Power= 0.96
Hillis 2011: df= 9.00, Lambda= 0.21, Power= 0.07

iMRMC software tool

STEP 3: Sample sizing



Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Significance level Effect Size # Reader #Normal #Diseased

Sizing Analysis: SqrtVar= , Test Stat=

Normal Approx: df= ∞ , Power=

BDG: df= , Lambda= , Power=

Hillis 2011: df= , Lambda= , Power=

Where do we get this number from?

- ✓ Pilot study results
- ✓ Reported results from studies using the same modality
- ✓ A relevant difference in performance

iMRMC software tool

STEP 3: Sample sizing

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

3 - Size

Summary

Significance level Effect Size # leader #Normal #Diseased

Sizing Analysis: SE = 4.393E-2 , Stat= 4.098E0

Normal Approx: df= ∞ , Power= 0.98

BDG: df= 11.44 , Lambda= 16.79 , Power= 0.96

Hillis 2011: df= 9.00 , Lambda= 0.21 , Power= 0.07

Significance level Effect Size # leader #Normal #Diseased

Sizing Analysis: SqrtVar=4.393E-2 , Stat= 2.276E0

Normal Approx: df= ∞ , Power= 0.62

BDG: df= 11.44 , Lambda= 5.13 , Power= 0.55

Hillis 2011: df= 9.00 , Lambda= 0.06 , Power= 0.06

Summary

Importance

Pilot study

Precision

iMRMC

1- Load data

2 - Analyze

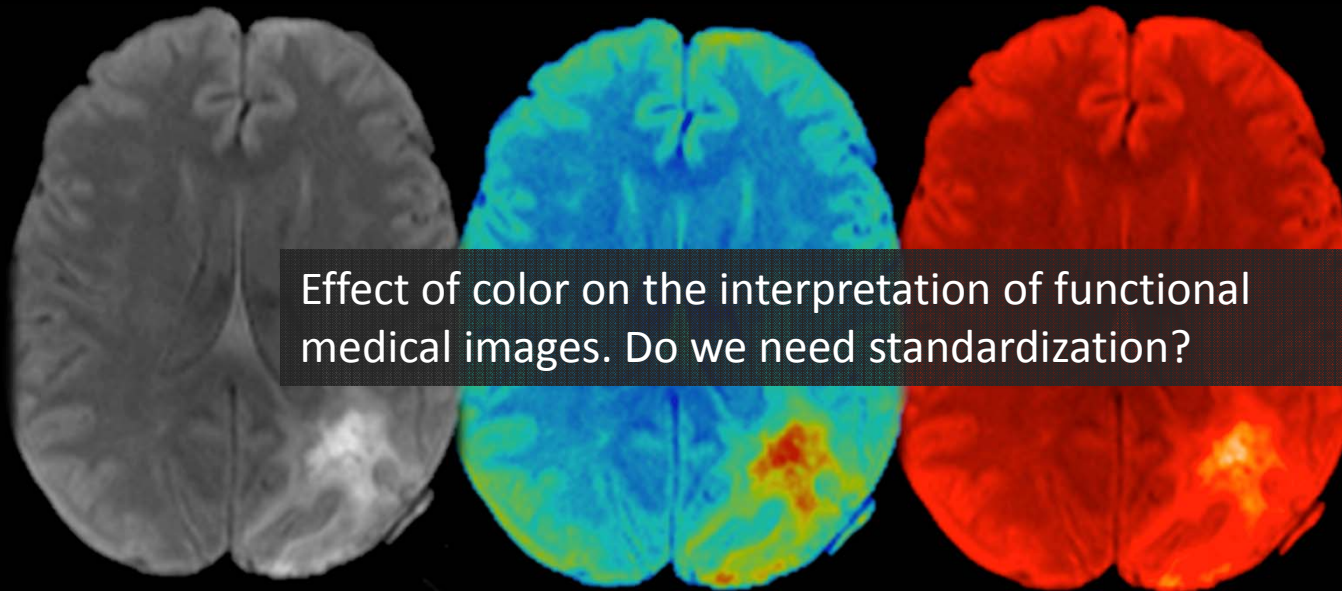
3 - Size

Summary

✓ Sizing a reader study is essential to obtain precise results.

✓ The iMRMC software is an open-source statistical tool for the analysis and sample sizing of reader studies.

✓ It is free and easy to use!



Effect of color on the interpretation of functional medical images. Do we need standardization?

Participate as a reader in an on-going color study!!!!
South – Hall A - BARCO Booth 2559 😊

Thank you!
s.zabala.travers@gmail.com