iMRMC

User Manual 2016

Table of Contents

1. Introduction

The primary objective of this Java application (iMRMC) is to assist investigators with analyzing and sizing multi-reader multi-case (MRMC) reader studies that compare the difference in the area under Receiver Operating Characteristic curves (AUCs) from two modalities. The core elements of this java application include the ability to do MRMC variance analysis and the ability to size an MRMC trial. A database containing components of variance from past MRMC studies is planned.

When publishing results using our software, please reference with one of the following:
- Gallas, B. D.; Bandos, A.; Samuelson, F. & Wagner, R. F. (2009), 'A Framework for Random-Effects ROC Analysis: Biases with the Bootstrap and Other Variance Estimators', Commun Stat A-Theory 38 (15), 2586-2603.
- Gallas, B. D. (2006), 'One-Shot Estimate of MRMC Variance: AUC.' Acad Radiol, 13 (3), 353-362.

The software treats arbitrary study designs that are not "fully-crossed". Please refer to the following for more information.
- Gallas, B. D. & Brown, D. G. (2008), 'Reader Studies for Validation of CAD Systems.' Neural Networks Special Conference Issue, 21, (2-3), 387-397.
- Obuchowski, N.; Gallas, B. D. & Hillis, S. L. (2012), 'Multi-Reader ROC Studies with Split-Plot Designs: A Comparison of Statistical Methods.' Acad Radiol, 19, 1508-1517.
- Chen, W; Gong, Q; Gallas, B. D, 'Paired split-plot designs of multireader multicase studies,' J. Med. Imag. 5(3), 031410 (2018).

Credits

- Brandon Gallas 2010 - Present
    - Project Leader

- Xin He 2010 - 2011
    - Original Creator

- Rowhan Pathare 2013
    - Begin version control (SVN)
    - SNV code shared on google code
    - Arbitrary study designs, ROC Plots

- Qi Gong 2014 - Present
    - Create new features, debug and improve performance for iMRMC
    - Public code and release version to Github

- Yousun Ko 2015 - 2016
    - Transitioned everything to GitHub
    - Consolidated on the documentation (User Manual)

- Jessica Handoko 2016
    - Update the user manual
    - Upload all documents about iMRMC to GitHub

2. Getting Started

The primary objective of this Java application (iMRMC) is to assist investigators with analyzing and sizing multi-reader multi-case (MRMC) reader studies that compare the difference in the area under Receiver Operating Characteristic curves (AUCs) from two modalities. The core elements of this java application include the ability to do MRMC variance analysis and the ability to size an MRMC trial. A database containing components of variance from past MRMC studies is planned.

Download Java

The software is written in Java. It requires Java Runtime Environment(JRE) 1.8 and higher version.

User can check the JAVA version from http://java.com/en/download/testjava.jsp

User can download JRE from ORACLE website.

Download iMRMC

Go to https://github.com/DIDSR/iMRMC/releases to download and save the latest version of iMRMC file.

3. Run iMRMC

iMRMC software can be run under 2 models, GUI interface and command line. Both of them are available on Windows, Mac and Linux operation systems.

    In **GUI interface**, users use buttons to load input file and manage software.

    In **command line**, users can load input file by command line, and iMRMC will automatically do all modalities combinations MRMC analysis.

iMRMC accepts 2 kinds of input files.

    1. **Study Raw Data** : It includes study size, case truth and MRMC studies ROC scores.

    2. **Summary Data** : It include study information and part of MRMC AUC analysis results (BDG moments and AUCs).

iMRMC can do MRMC statistical analysis and Sizing Future Study.

    1. **MRMC Statistical analysis:** iMRMC estimates the AUC components of variance for input file.

    2. **Sizing Future Study:** iMRMC predicts a new study (study size and study design) base on statistical analysis reusults.
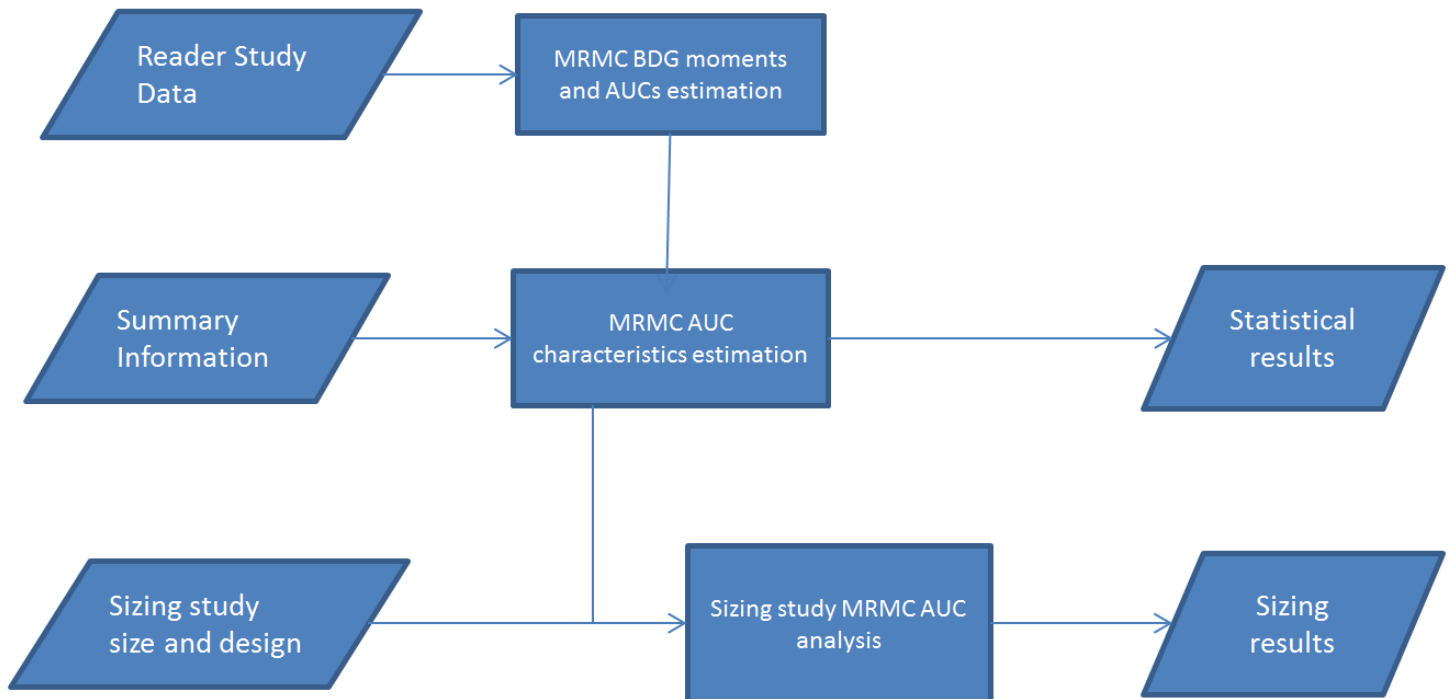
3.1 GUI Interface

This is the iMRMC main GUI. It include 4 parts: 1. Manage panel, 2. Statistical Analysis Result, 3. Sizing Future Study, 4. Export Results.
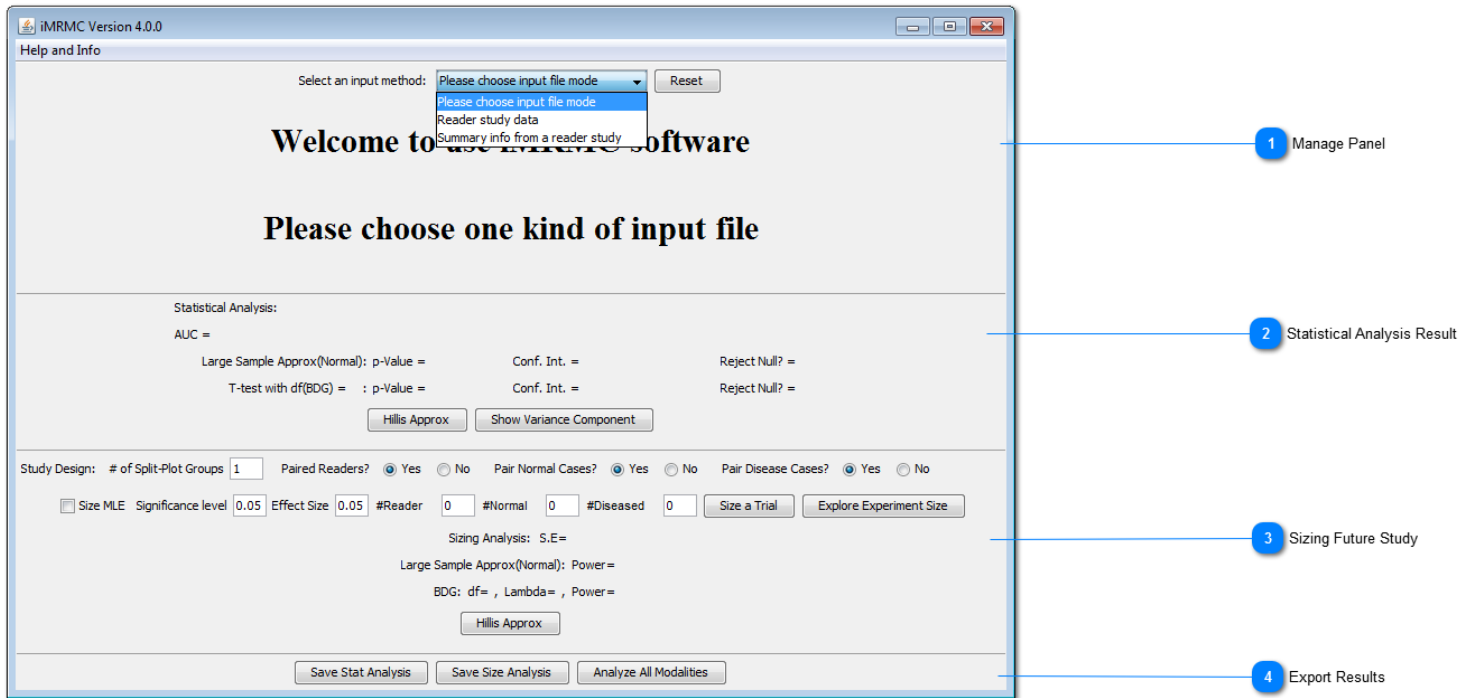
Manage Panel
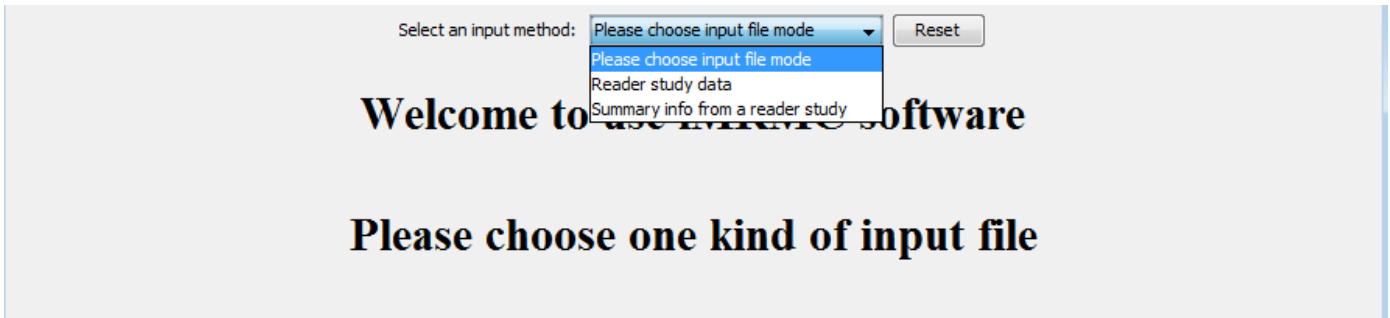
Default Panel is a welcome page.

User could use "Select an input method" pull down manual to choose input method.

Reader Study Data: Analysis Study Raw Data.

Summary Info From Reader Study Data: Analysis Summary Data.

User could use "Reset" button to clean all analysis settings and results.

Manage panel will change depending on different types of input methods.

Statistical Analysis Result
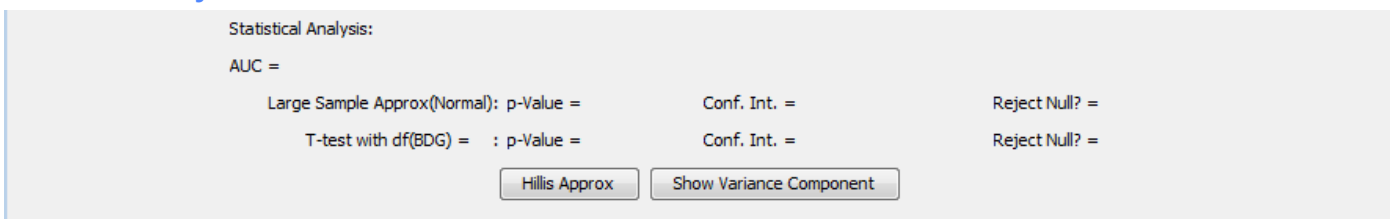
Displays statistical analysis reuslt

Sizing Future Study

Study Design: # of Split-Plot Groups 1   Paired Readers? ⊙ Yes ○ No   Pair Normal Cases? ⊙ Yes ○ No   Pair Disease Cases? ⊙ Yes ○ No

☐ Size MLE Significance level 0.05 Effect Size 0.05 #Reader 0   #Normal 0   #Diseased 0   [Size a Trial]   [Explore Experiment Size]

Sizing Analysis: S.E=

Large Sample Approx(Normal): Power=

BDG: df= , Lambda= , Power=

[Hillis Approx]

Sets sizing study parameter and displays results.

4 **Export Results**

[Save Stat Analysis]   [Save Size Analysis]   [Analyze All Modalities]

Export analyses results.

Study Design: # of Split-Plot Groups 1   Paired Readers? ⊙ Yes ○ No   Pair Normal Cases? ⊙ Yes ○ No   Pair Disease Cases? ⊙ Yes ○ No

☐ Size MLE Significance level 0.05 Effect Size 0.05 #Reader 0   #Normal 0   #Diseased 0   [Size a Trial]   [Explore Experiment Size]

Sizing Analysis: S.E=

Large Sample Approx(Normal): Power=

BDG: df= , Lambda= , Power=

[Hillis Approx]

3.1.1 Reader Study Data

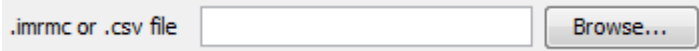In the Reader Study Data input method, welcome panel changes to following panel. User should load Study Raw Data as input file.
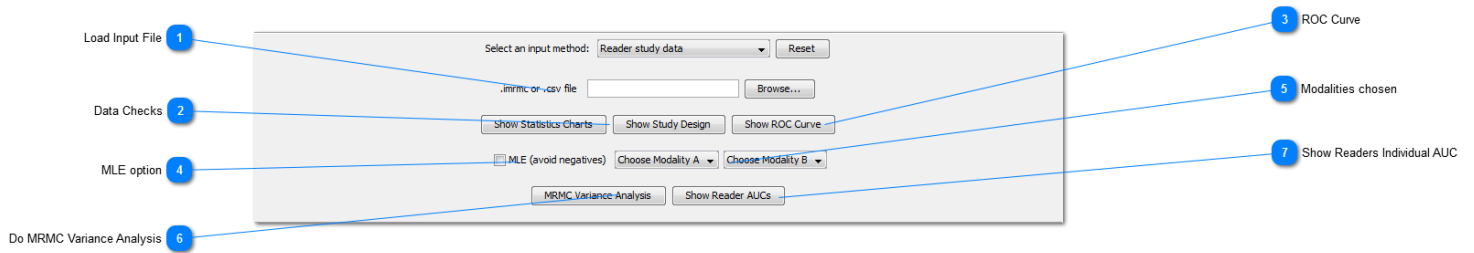
1 Load Input File

User can use the "Browse..." button to load one Study Raw Data input file. The input file should be either in ".imrmr" or ".csv" format.

2 Data Checks

User can use these two buttons to check the input file data.

3 ROC Curve
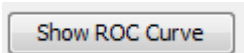
Display ROC curve for all readers and modalities.

4 MLE option
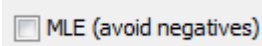
Check box to define whether use Maximum Likelihood Estimate (MLE) analysis.

5 Modalities chosen

Two pull down manual to choose analysis modalities.
    Choose one modality: Estimates MRMC AUC characteristic for one modality.
    Choose two modalities: Estimates MRMC AUC characteristic for two modalities, and difference between modalities.
Note: Two chosen modalities cannot be the same.

6 Do MRMC Variance Analysis
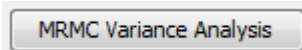
Do MRMC Variance Analysis for chosen modalities
Note: User should choose Modalities first.

7 Show Readers Individual AUC
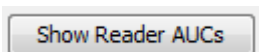
Show All readers AUC for each modality.
Note: User should do analysis first.

3.1.1.1 Data Checks

User can use **Show Statistics Chart** and **Show Study Design** buttons to check input file. They help user to easily determine if the study has been input correctly.

Statistics Charts

The left chart displays the number of cases that each reader scored, and the right one displays the number of readers that scored each case. The bars allow for easy visual recognition of where the study is incomplete. Hovering the mouse over a particular bar displays the relevant information for the data.
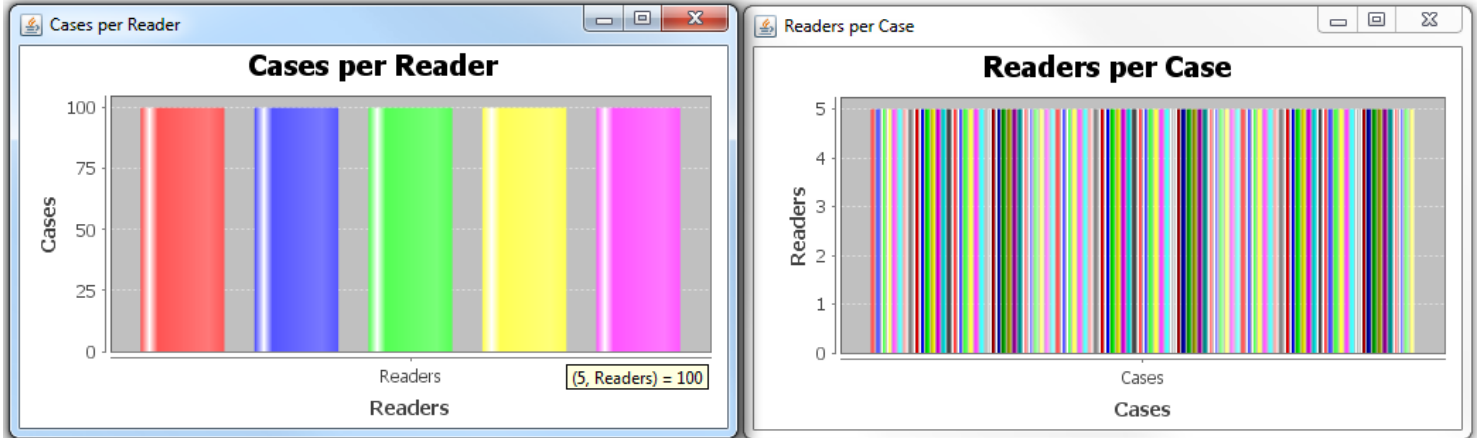
Study Design

For a more in-depth view of the completeness of the study, the **Show Study Design** button allows the user to view the presence or absence of score data per modality. At first user should select a modality to view.
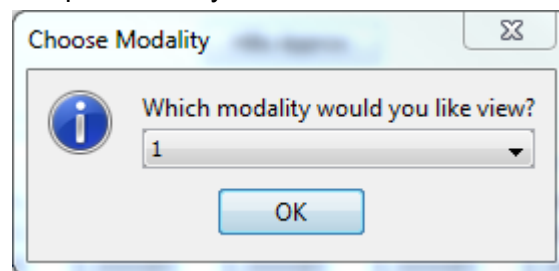
Study Design chart displays each case and each reader appears for chosen modality. A black square represents an existing score for said modality, case, and reader. A white square represents missing data at this point. Hovering the mouse over a particular square displays this information as "Case Index / Reader Index". Note: In the chart, the Y axis label is reader ID, while X axis label is case index.
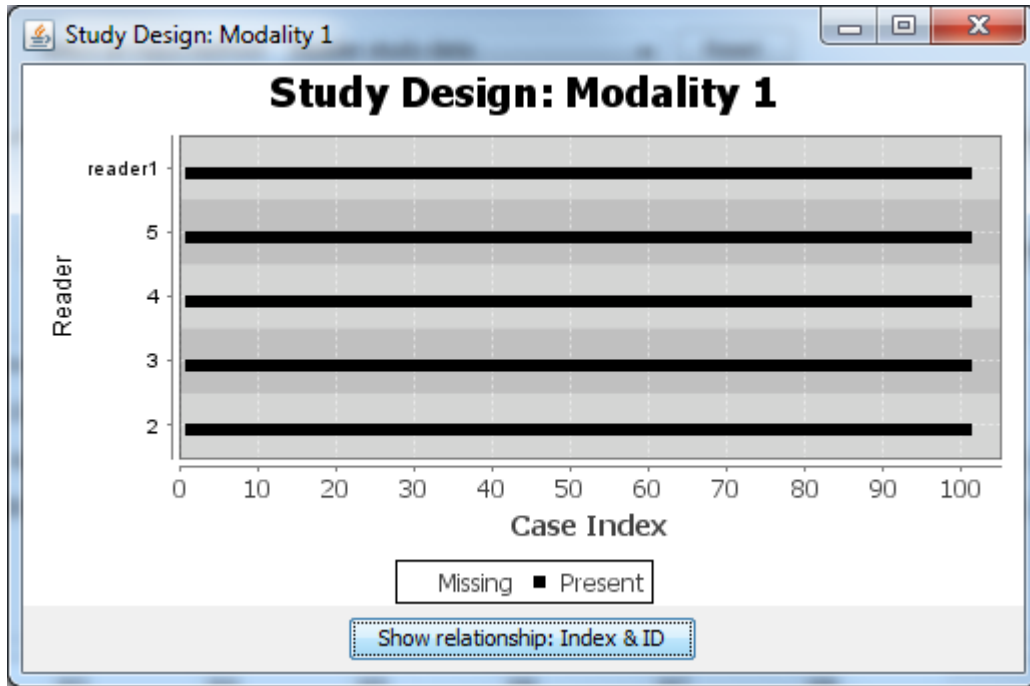
Show relationship: Index& ID button can display the relationship between Index and ID for reader and case in two windows. The "Export" button at the botton of windows could export the relationship to input file directory.

3.1.1.2 ROC Curve

The **Show ROC Curve** button can display readers individual and average ROC curves for multiple modalities, and then displays the ROC curves for the given modality. At first, user should select a modality to view (multiple choices is available) .
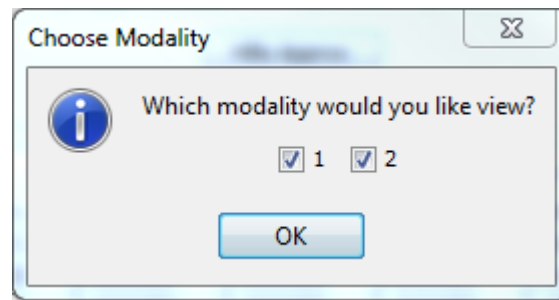
Selecting the various checkboxes on the bottom border of the chart window enables/disables ROC curves for each individual reader, as well as average ROC curves, of which there are four types.

   "Reader ID" controls the individual curve for this reader.

   Horizontal Average averages in specificity at every possible specificity.

   Vertical Average averages in sensitivity at every possible specificity.

   Diagonal Average averages in the direction of sensitivity+specificity at every sensitivity-specificity (along x=y diagonal line).

   Pooled Average gathers all reader scores into one large set and calculates the ROC curve across that set.

   Show Readers controls all readers individual curves.

   Show Averages controls all average curves.
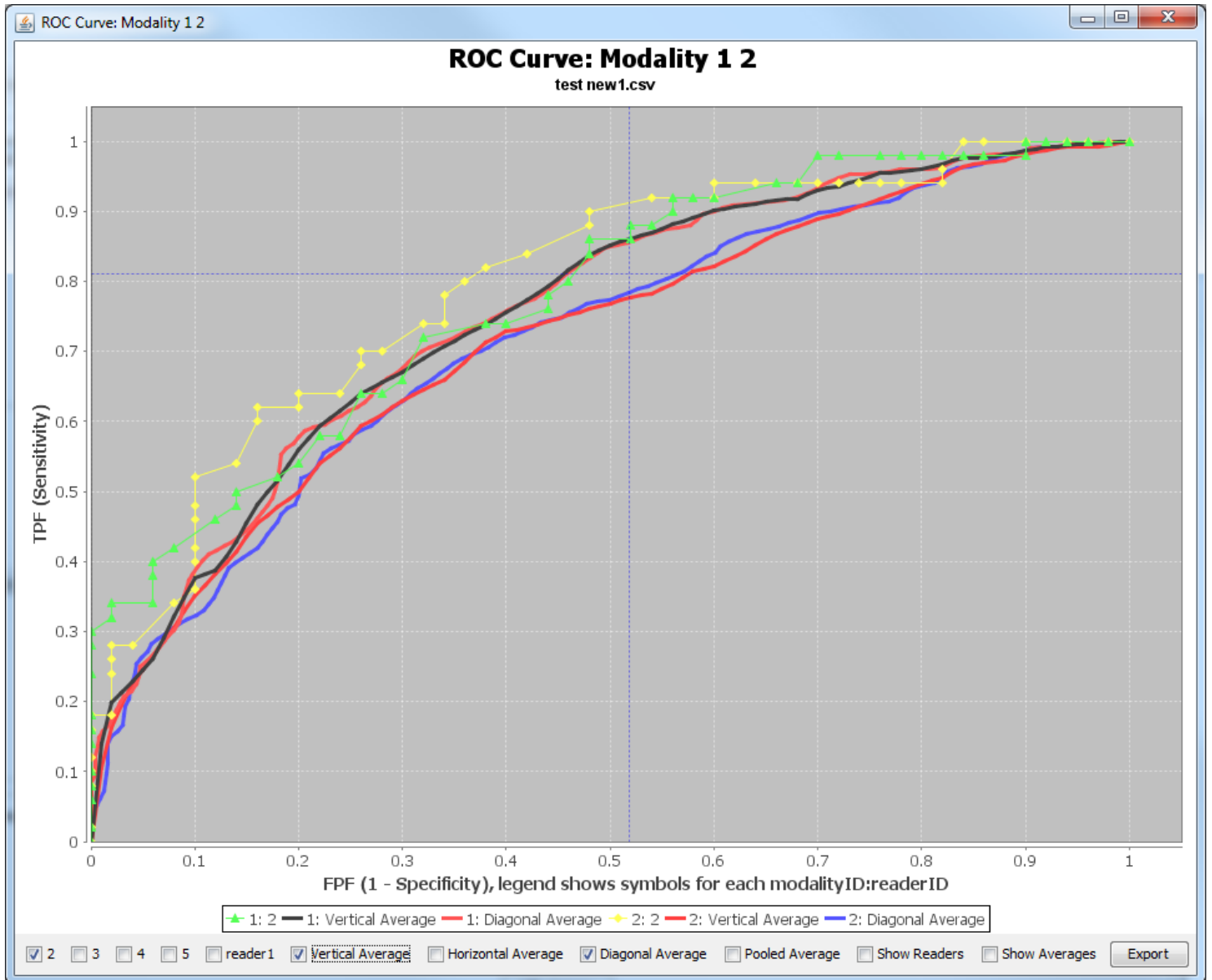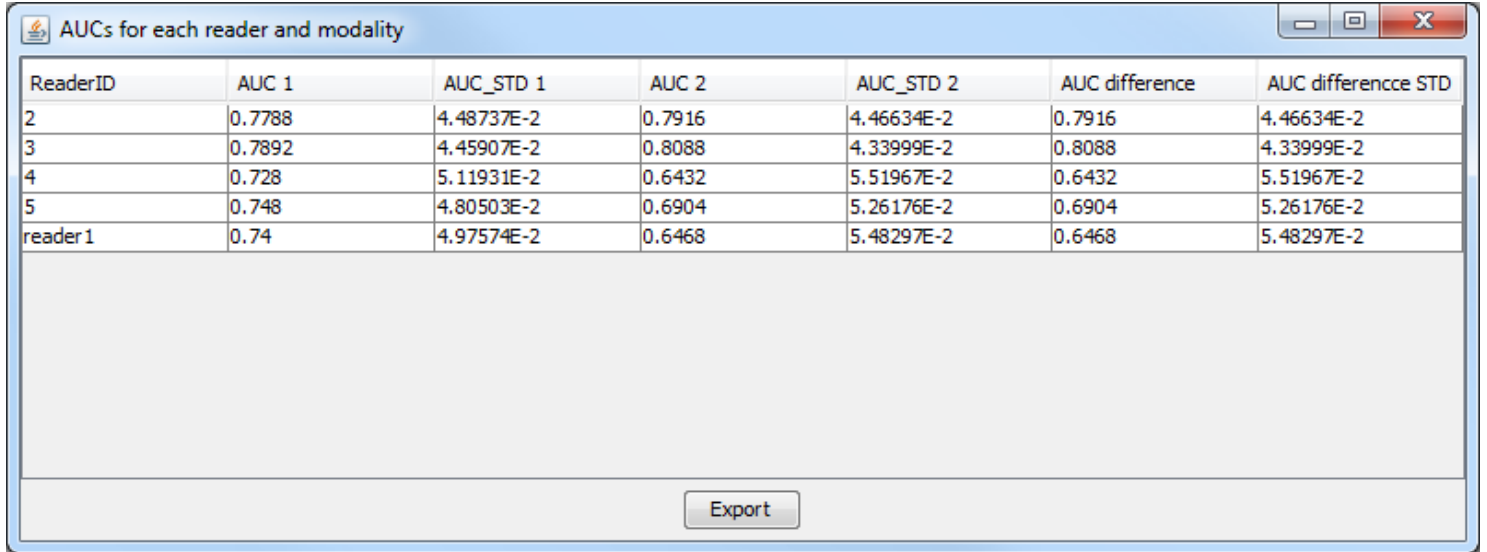
The chart is resizable, and its contents will scale to the new size automatically. Additionally, clicking and dragging from an upper left to lower right direction zooms the view in on that area. Clicking and dragging from a lower right direction to an upper left direction returns the view to its default zoom.

Export button can gather all curves data to one file and export to input file directory.

3.1.1.3 Reader AUCs

After statistical analysis, user can use the **Show Reader AUCs** button to display readers individual AUCs for analyzed modalites.

ReaderID	AUC 1	AUC_STD 1	AUC 2	AUC_STD 2	AUC difference	AUC differencce STD
2	0.7788	4.48737E-2	0.7916	4.46634E-2	0.7916	4.46634E-2
3	0.7892	4.45907E-2	0.8088	4.33999E-2	0.8088	4.33999E-2
4	0.728	5.11931E-2	0.6432	5.51967E-2	0.6432	5.51967E-2
5	0.748	4.80503E-2	0.6904	5.26176E-2	0.6904	5.26176E-2
reader1	0.74	4.97574E-2	0.6468	5.48297E-2	0.6468	5.48297E-2

Export

Export button can export AUCs to one file. User has option to choose the saving directory.

3.1.2 Summary Information From Reader Study Data

In this input method, welcome panel changes to following panel. User should load Summary Data as input file. It is one of the output files of **Save Stat Analysis**button.

As the input file does not contain raw reading scores, there are no data checks, show ROC curves and show Reader AUCs options. The **MRMC Variance Analysis**button is not available . iMRMC will do analysis when user chooses one modality options from **Modalities Chosen** pull down manual.

The main porpose of this analysis version is for sizing furture study. The statistical panel only does fullycrossed analysis.
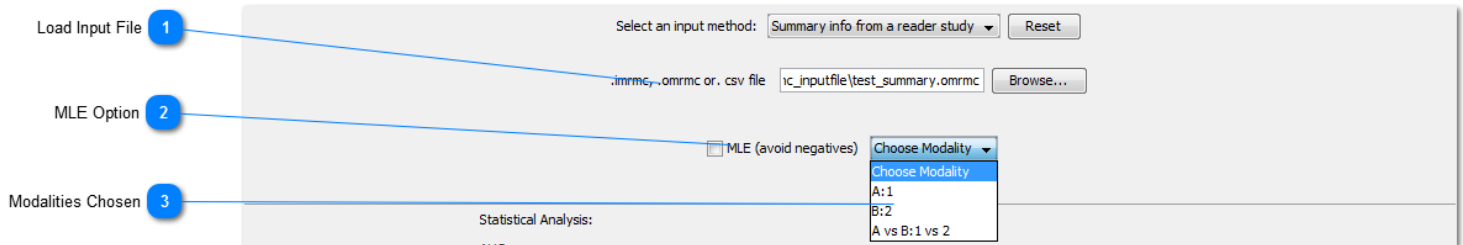
Load Input File

User can use the "Browse..." button to load one Summary Data input file. The input file should be either in ".imrmr", ".omrmc" or ".csv" format.

MLE Option

Check box to define whether use Maximum Likelihood Estimate (MLE) analysis.

Modalities Chosen

Choose analysis modality. After user chooses one option, iMRMC will do analysis automatically.

> **A:1:** Analyze modality "1" as the first modality
> **B:2:** Analyze modality "2" as the second modality
> **A vs B: 1 vs 2:** Analyze difference between two modalities, modality "1" as the first and modality "2" as the second modalities.

3.1.3 Statistical Analysis Results

This panel displays Statiscal Analysis Results
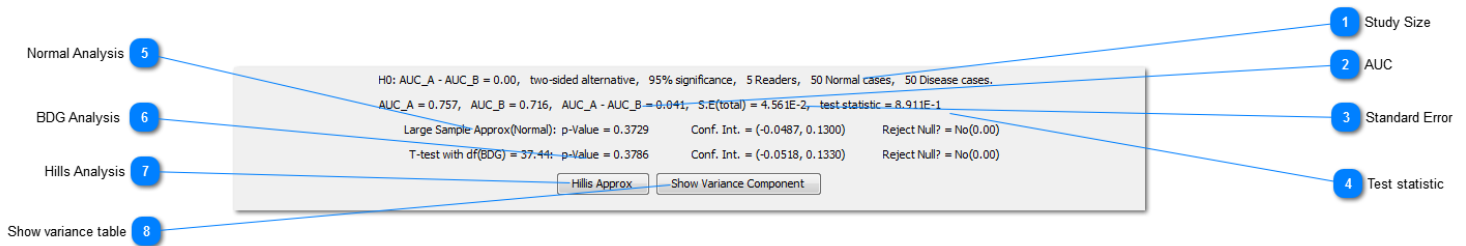
Study Size

5 Readers, 50 Normal cases, 50 Disease cases.

Reader and case sizes of analyzed data.

AUC

AUC_A = 0.757, AUC_B = 0.716, AUC_A - AUC_B = 0.041,

AUC for each modality and difference between them. If user only analyzes one modality, there is one AUC term.

Standard Error

S.E(total) = 4.561E-2,

BDG analysis Standard Error.

Test statistic

test statistic = 8.911E-1

Test statistic for Normal, BDG, Hillis methods

Normal Analysis

Large Sample Approx(Normal): p-Value = 0.3729   Conf. Int. = (-0.0487, 0.1300)   Reject Null? = No(0.00)

This method assumes test statistic comes from a normal distration (z-test). Results include P-Value, Confidence Interval, and Hypothesis test result (Yes = 1, No = 0).

BDG Analysis

T-test with df(BDG) = 37.44: p-Value = 0.3786   Conf. Int. = (-0.0518, 0.1330)   Reject Null? = No(0.00)

This method assumes test statistic comes from a t-distribution with BDG degree of freedom (t-test). Results include P-Value, Confidence Interval, and Hypothesis test result (Yes = 1, No = 0).

Hills Analysis

Hillis Approx

Hills analysis P-Value, Confidence Interval and Hypothesis test result (Yes = 1, No = 0). Click button to see the result.

Show variance table

Show Variance Component

The botton open Variance Component window. It shows the components of variance, which can be given in fiv representations. For the BDG and BCK representation we have 7 total rows. The "**comp MA**" and "**comp MB**" components for each modality, and the "**coeff MA**, **MB**" rows give the corresponding coefficients/weights for ea

variance. The "**comp product**" shows the covariance components for both modalities, and the "**- coeff produc**
weight. The "**total**" row displays the contribution towards the total variance by each component. The contributi
produce the total variance, and the square root of that is the standard error, which is displayed to the right of th
and MS tabs are only available for fully crossed analysis data. They are similar as the frist 2 tabs, but only con
the components, coefficients, and total without separating by modality.

If any of the resulting components are negative, the user will be asked if they would like to use MLE estimates
statistics estimates. The MLE estimates of the components are never negative, and this can help with calculati
sometimes. If MLE estimates are being used, a double asterisk (**) will appear next to the decomposition nam

Variance Component

BDG | BCK | DBM | OR | MS

	M1	M2	M3	M4	M5	M6	M7	M8
comp MA	7.56800E-1	6.27442E-1	6.27187E-1	5.71009E-1	6.34960E-1	6.01083E-1	5.99914E-1	5.71446E-1
coeff MA	8.00000E-5	3.92000E-3	3.92000E-3	1.92080E-1	3.20000E-4	1.56800E-2	1.56800E-2	-2.31680E-1
comp MB	7.16160E-1	5.84519E-1	5.71693E-1	5.15359E-1	5.76680E-1	5.43780E-1	5.38985E-1	5.10390E-1
coeff MB	8.00000E-5	3.92000E-3	3.92000E-3	1.92080E-1	3.20000E-4	1.56800E-2	1.56800E-2	-2.31680E-1
comp product	5.94000E-1	5.69580E-1	5.63642E-1	5.42661E-1	5.63920E-1	5.52715E-1	5.52043E-1	5.41132E-1
- coeff product	1.60000E-4	7.84000E-3	7.84000E-3	3.84160E-1	6.40000E-4	3.13600E-2	3.13600E-2	-4.63360E-1
total	2.27968E-5	2.85376E-4	2.80653E-4	2.00896E-4	2.68160E-5	6.18317E-4	5.45869E-4	9.92413E-5

total var=

3.1.4 Sizing Future Study

The third panel is for sizing a trial using the components of variance from statistical analysis.

    1. Set the number of split-plot groups, pairing of readers, pairing of cases and study size.

    2. Set the significance level and the effect size

    3. Click **"Size a Trial"** or **"Explore Experiment Size"** buttons.

Note: Sizing MLE analysis has not be validated
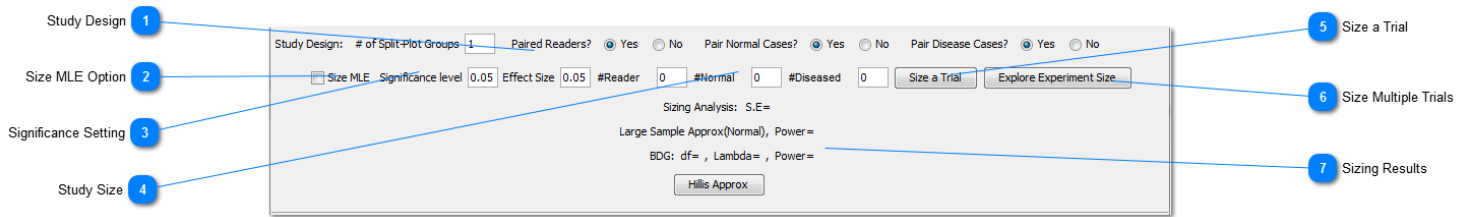
1   Study Design

Sets the number of split-plot groups, pairing of readers, pairing of cases.

2   Size MLE Option

Check box to define whether use Maximum Likelihood Estimate (MLE) in sizing analysis.

3   Significance Setting

Sets the significance level and the effect size.

4   Study Size

Sets number of readers, nondiseasedcases, and diseased cases.

5   Size a Trial

Uses components of variance from statistical analysis and sizing parameter to estimate a new MRMC study analysis.

6   Size Multiple Trials

Uses components of variance from statistical analysis to estimate different sizes MRMC studies analyses, and displays results in **Explore Experiment Size** window.

7   Sizing Results

Normal, BDG and Hills analyses results.

3.1.4.1 Explore Experiment Size

This window shows analyses results for different sizes MRMC studies base on components of variance from statistical analysis. This feature could analyze different size of studies (reader size = [5, 10, 15, 20]; case size = [32, 64, 128, 256, 512, 1024, 2048, 4096, 8192]). It can help user to have a base idea about what study size they need to reach their ideal variance.

Note: All analyses in this feature assume the studys are fully-crossed and paired between two modalities.
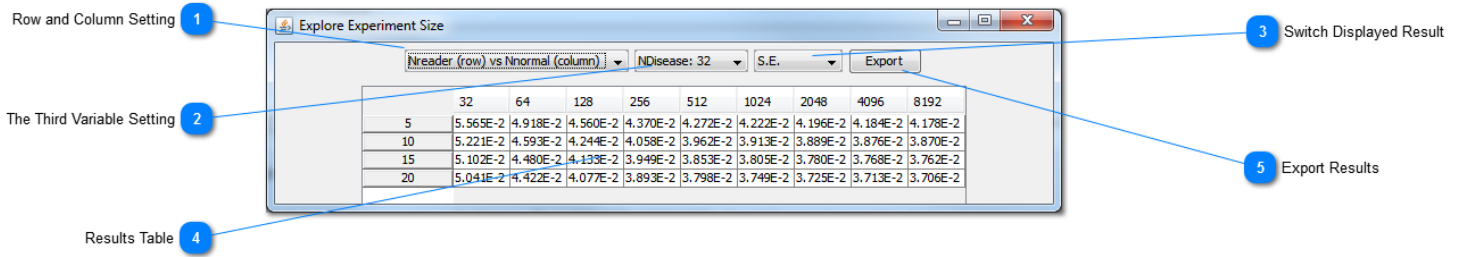
Row and Column Setting
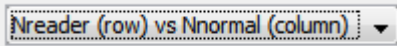
Chooses Results Table row and column combination.

Row	Column
Nreader	Nnormal
Nreader	Ndisease
Nnormal	Ndisease

The Third Variable Setting
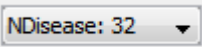
Sets the size for the third variable, which is not in the Results Table.

Switch Displayed Result

Chooses one AUC characteristic to be estimated and displayed (Standard error or BDG power).

Results Table

	32	64	128	256	512	1024	2048	4096	8192
5	5.565E-2	4.918E-2	4.560E-2	4.370E-2	4.272E-2	4.222E-2	4.196E-2	4.184E-2	4.178E-2
10	5.221E-2	4.593E-2	4.244E-2	4.058E-2	3.962E-2	3.913E-2	3.889E-2	3.876E-2	3.870E-2
15	5.102E-2	4.480E-2	4.133E-2	3.949E-2	3.853E-2	3.805E-2	3.780E-2	3.768E-2	3.762E-2
20	5.041E-2	4.422E-2	4.077E-2	3.893E-2	3.798E-2	3.749E-2	3.725E-2	3.713E-2	3.706E-2

Analysis resutls for different study sizes.
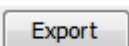
Export Results

Export current Results Table to file.

3.1.5 Export results

User can use buttons in this panel to export analyses results.

| Save Stat Analysis | Save Size Analysis | Analysis All Modalities |

Save Stat Analysis:

> User should do statistical analysis first. This button can export current statistical analysis results (Results in **Statistical Analysis Results**).
>
> User will be asked to create a folder. There will be two files saved in that folder.
>> **omrmc format:** Readable by humans and can be loaded as a **Summary Data** on the iMRMC software.
>>
>> **csv format:** Easier to be loaded by other software.

Save Size Analysis:

> User should do sizing analysis first. This button export sizing analysis results (Reuslts in **Sizing Future Study**).
>
> Output file is in omrmc format, which is readable by humans. But it can't be loaded to iMRMC as summary data.

Analysis All Modalities:

> After loading input file, user can use this button to do statistical analysis for all modalities combinations in input file and export results. User doesn't have to do statistical analysis or sizing analysis first.
>
> Software will create a folder in input file directory, and save all outputs in it. All the outputs are in csv format. Two analysis models have different outputs.
>> **Reader Study Data:**
>>> **1. ROC Curves Data:** All ROC curves data for all modalities.
>>>
>>> **2. MRMC Variance Components Tables:** BDG, BCK, DBM, MS, OR variance components for all modalities combinations. Each method is in individual file, which includes both U-statistics and MLE analyses tables.
>>>
>>> **3. Readers individual AUC:** Readers individual AUC for all modalities combinations.
>>>
>>> **4. Readers covariance:** Readers covariance for all modalities combinations.
>>>
>>> **5. U-statistics analyses results:** MRMC U-statistics Analysis results (AUCs, Standard Error, Confidence Interval, and Hypothesis test results) for all modalities combinations.
>>>
>>> **6. MLE analyses results:** MRMC MLE Analysis results (AUCs, Standard Error, Confidence Interval, and Hypothesis test results) for all modalities combinations.
>>
>> **Summary Informations From Reader Study Data:**
>>> **1. MRMC Variance components tables:** BDG, BCK, DBM, MS, OR variance components for all modalities combinations. Each method is in individual file, which includes both U-statistics and MLE analyses tables.
>>>
>>> **2. U-statistics analyses results:** MRMC U-statistics Analysis results (AUCs, Standard Error, Confidence Interval, and Hypothesis test results) for all modalities combinations.
>>>
>>> **3. MLE analyses results:** MRMC MLE Analysis results (AUCs, Standard Error, Confidence Interval, and Hypothesis test results) for all modalities combinations.

3.2 Command Line
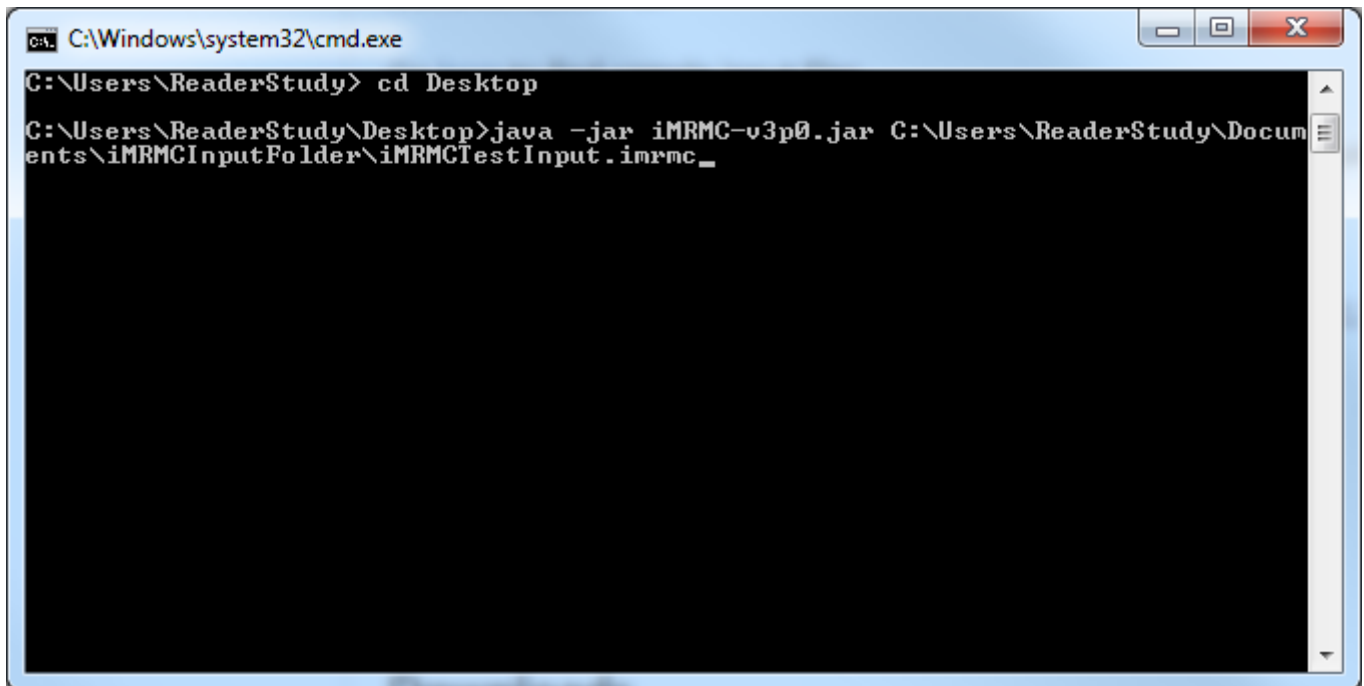
iMRMC can be launched and run by command line. In this model user can input both **Reader Study Data** and **Summary Data.**

The software does the same processes as **Analysis All Modalities** button. With this feature, users can also call iMRMC from other softwares.

User can launch and run iMRMC in command line by the following steps:
 1. Open up the command window (cmd, terminal) or other compatible software to visit the iMRMC software saving directory.
 2. Type the command "java -jar iMRMC-v3p0.jar", along with the full input file name and path.
 3. User can add second arguement after input file to define output file saving directory.
 Note: In the command line version, the input file name and path shouldn't have any space (bad input filename: iMRMC Test.imrmc)

The software can also be called by other program language
 1. R: The R project is in https://github.com/DIDSR/iMRMC/tree/master/Rpackage. The project can build Rpakcage for more convenience future usage.
  a. simMRMC.R can generate variance configuration and simulate Study Raw Data.
  b. iMRMC.R loads Study Raw Data and run the iMRMC software by command line. Then it loads in the analysis results and saves the results as variables in the R environment.

 2. Matlab: The function is in https://github.com/DIDSR/iMRMC/tree/master/MatlabFunction.
  The function can convert the LABMRMC format study data to our software readable .imrmc Study Raw Data. Then it load Study Raw Data and run iMRMC software by command line. Finally, it loads in the analysis results and saves the results as variables in the Matlab environment.

4. Input and output

Users can create an input file with all the settings with any text editing software, like notepad and excel.

The software can remember the user's last input and output file directories, and use them as default directories for next load and export processing.

4.1 Input file

There are two kinds of input file **Study Raw Data** and **Summary Data**.

The .omrmc version output file of **Save Stat Analysis** button can also be loaded to iMRMC as summary data.

4.1.1 Study Raw Data

Study Raw Data is the raw MRMC reading scores. It include 2 parts: 1. Study Description, 2. List of ROC ratings.

Study Description can include any information as free text, or no information at all. iMRMC looks for four lines corresponding to the size of the experiment.

N0 = number of cases without disease
N1 = number of cases with disease
NR = number of readers
NM = number of modalities

The **list of ROC ratings** begins with "BEGIN DATA:'. It has a row for each score—each reader, case, modality combination observed—in any order. It also has rows specifying the truth state of each case (given by the reference gold-standard reader = truth). iMRMC can handle data that is not fully-crossed for any number of modalities, though only two of those modalities may be used at a time for variance analysis.

For **rows recording reader scores**, there are four fields in the following order:

reader I.D. (string)
> The reader I.D. can be any string (or number) to identify a study reader. Make sure that this reader I.D. appears the same way each time or the program will think that there are multiple readers. For example, iMRMC believes that "novice0247", "novice 0247", "NOVICE0247" are three different readers. Do not change the I.D. of a reader when recording his or her scores from a different modality.
> Note: Do not use "-1" as that indicates a row recording the truth state.

case I.D. (string)
> The case I.D. can be any string (or number) to identify the case imaged. Make sure that this case appears the same way each time or the program will think that there are multiple cases. For example, iMRMC believes that "TeacherMale15", "Male Teacher15", "TeacherM15" are three different cases. Do not change the I.D. of a case when recording its scores from a different modality.

modality I.D. (string)
> The modality I.D. can be any string (or number) to identify the imaging modality. Some examples could be: "film", "digital", "digital+CAD", "MRI". Again be careful how you type the modality I.D.

score (integer or float)
> The score is the level of confidence, likelihood, rank, or severity of disease that the reader gives to the case for the modality indicated. The score can have any scale and can be negative : percent, fraction, integer (1 to 7, 0 to 100, -20 to +20), or floating point. Low scores correspond to truth state 0 (normal or no disease) and high scores correspond to truth state 1 (abnormal or disease). If low scores correspond to truth state 1 and high scores correspond to truth state 0, AUC calculations and ROC curve displays will be incorrect.

For **rows recording truth states**, there are also four fields in the following order:

Truth flag (-1 or truth)
> Rows recording case truth states are denoted with a -1 or truth in the first field.

Case ID (Integer)
> Every case I.D. scored by a reader (appears in a row recording a reader score described above) must have one and only one row recording the truth state. Each case must be defined as 0 (normal or no disease) or 1 (abnormal or disease).

Arbitrary String (String)
> This string can be anything and it needs to be something. It is currently unused by iMRMC. You can use "0" as a reminder of what the row is recording.

Truth state (0 or 1)
> The truth state 0 indicates normal or no disease. The truth state 1 indicates abnormal or disease.

Higher reader scores should indicate higher likelihood or confidence of disease and low ratings should indicate lesser likelihood or confidence of disease.

FIELDS MUST BE SEPARATED BY COMMAS!*

For example, one input file could look like the following:

```
 1 Simulated iMRMC input from 15-01-22-17-49-14
 2
 3 NR: 2
 4 N0: 1
 5 N1: 1
 6 NM: 2
 7
 8 BEGIN DATA:
 9 -1,case1,0,1
10 -1,case2,0,0
11 reader1,case1,film,1.414840794
12 reader1,case1,digital,1.986055857
13 reader2,case1,film,-0.414840794
14 reader2,case1,digital,0.689055857
15 reader1,case2,film,-1.114840794
16 reader1,case2,digital,-0.874555857
17 reader2,case2,film,0.239740794
18 reader2,case2,digital,-0.186555857
19
```

① Study Description

```
 1 Simulated iMRMC input from 15-01-22-17-49-14
 2
 3 NR: 2
 4 N0: 1
 5 N1: 1
 6 NM: 2
```

2 readers, 1 nondiseased case, 1 diseased case and 2 modalities

② List of ROC Ratings

```
 8 BEGIN DATA:
 9 -1,case1,0,1
10 -1,case2,0,0
11 reader1,case1,film,1.414840794
12 reader1,case1,digital,1.986055857
13 reader2,case1,film,-0.414840794
14 reader2,case1,digital,0.689055857
15 reader1,case2,film,-1.114840794
16 reader1,case2,digital,-0.874555857
17 reader2,case2,film,0.239740794
18 reader2,case2,digital,-0.186555857
```

Line 9 specifies truth state for "case1" is abnormal
    Field 1: -1 indicates row records truth state
    Field 2: case I.D. = "case1"
    Field 3: arbitrary string = "0"
    Field 4: truth state = 1
Line 10 specifies truth state for case2 is normal
    Field 1: -1 indicates row records truth state
    Field 2: case I.D. = "case2"
    Field 3: arbitrary string = "0"
    Field 4: truth state = 0

Lines 11,12 show "reader1" scores for "case1" on modalities "film" and "digital".
Lines 13,14 show "reader2" scores for "case1" on modalities "film" and "digital".
Lines 15,16 show "reader1" scores for "case2" on modalities "film" and "digital".
Lines 17,18 show "reader2" scores for "case2" on modalities "film" and "digital".

4.1.2 Summary Data

Summary Data is previous MRMC analysis results. Summary file should begin with " BEGIN SUMMARY" and end with "END SUMMARY".
It includes 3 parts: 1. Study Information, 2. AUCs, 3. BDG Variance Components. Summary Data file could be generated by **Save Stat Analysis** button.
If users want to create Summary Data manually, they should follow the format exactly.

```
BEGIN SUMMARY
NReader=  5
Nnormal=  50
NDisease= 50

Modality A = 1
Modality B = 2

Reader-Averaged AUCs
AUC_A = 0.757
AUC_B = 0.716
AUC_A - AUC_B = 0.041

Reader Specific AUCs
Reader ID,      AUC_A,      AUCs_B,    AUC_A - AUCs_B
        2,  7.78800E-1,  7.91600E-1,     -1.28000E-2
        3,  7.89200E-1,  8.08800E-1,     -1.96000E-2
        4,  7.28000E-1,  6.43200E-1,      8.48000E-2
        5,  7.48000E-1,  6.90400E-1,      5.76000E-2
   reader1,  7.40000E-1,  6.46800E-1,      9.32000E-2
********************BDG Moments************************
          Moments,       M1,       M2,       M3,       M4,       M5,       M6,       M7,       M8
Modality1(AUC_A), 7.56800E-1, 6.27442E-1, 6.27187E-1, 5.71009E-1, 6.34960E-1, 6.01083E-1, 5.99914E-1, 5.71446E-1,
Modality2(AUC_B), 7.16160E-1, 5.84519E-1, 5.71693E-1, 5.15359E-1, 5.76680E-1, 5.43780E-1, 5.38985E-1, 5.10390E-1,
    comp product, 5.94000E-1, 5.69580E-1, 5.63642E-1, 5.42661E-1, 5.63920E-1, 5.52715E-1, 5.52043E-1, 5.41132E-1,
END SUMMARY
```

Study Information — 1
AUCs — 2
BDG Variance Components — 3

1 Study Information

```
NReader=  5
Nnormal=  50
NDisease= 50

Modality A = 1
Modality B = 2
```

1. Reader and Case Size.
2. Modality ID.

2 AUCs

```
Reader-Averaged AUCs
AUC_A = 0.757
AUC_B = 0.716
AUC_A - AUC_B = 0.041

Reader Specific AUCs
Reader ID,       AUC_A,        AUCs_B,     AUC_A - AUCs_B
        2,   7.78800E-1,   7.91600E-1,       -1.28000E-2
        3,   7.89200E-1,   8.08800E-1,       -1.96000E-2
        4,   7.28000E-1,   6.43200E-1,        8.48000E-2
        5,   7.48000E-1,   6.90400E-1,        5.76000E-2
   reader1,   7.40000E-1,   6.46800E-1,        9.32000E-2
```

1. Reader-Averaged AUCs for individual modalities and their difference.
2. Reader Specific AUCs in for each modality.

3 BDG Variance Components

```
********************BDG Moments************************
          Moments,       M1,       M2,       M3,       M4,       M5,       M6,       M7,       M8
Modality1(AUC_A), 7.56800E-1, 6.27442E-1, 6.27187E-1, 5.71009E-1, 6.34960E-1, 6.01083E-1, 5.99914E-1, 5.71446E-1,
Modality2(AUC_B), 7.16160E-1, 5.84519E-1, 5.71693E-1, 5.15359E-1, 5.76680E-1, 5.43780E-1, 5.38985E-1, 5.10390E-1,
    comp product, 5.94000E-1, 5.69580E-1, 5.63642E-1, 5.42661E-1, 5.63920E-1, 5.52715E-1, 5.52043E-1, 5.41132E-1,
```

BDG Variance Components for individual modality and the production between them.

4.1.3 Possible Formatting Errors

Currently, iMRMC does not provide detailed feedback on formatting errors. The feedback is to provide a report of the experiment size after successfully reading an input file.

Typical errors in formatting are as follows:

- \# The truth state of every case must be defined once and only once.
    - o ERROR: Truth states not defined for all the cases.
    - o ERROR: Truth states multiply defined.
- \# The row containing "BEGIN DATA:" must be typed exactly as given.
    - o ERROR: "BEGIN DATA:" not all upper-case.
    - o ERROR: "BEGIN DATA:" not terminated with a colon (:).
- \# Commas with no spaces must be used to separate the four fields of data describing each ROC rating.
    - o ERROR: Spaces, tabs, or semicolons(;) separate the data in the list of ROC ratings. Only commas should be used.

4.2 Output File

iMRMC could export several outputs by buttons

Outputs	Windows or Buttons
Index and ID relationship for Study Design label	**Study Design window**
ROC Curve Data	**ROC Curve window** and **Analysis All Modalities button**
Reader Specific AUCs	**Reader AUCs window** and **Analysis All Modalities button**
Statistical Analysis result (omrmc and csv)	**Save Stat Analysis button**
Sizing Analysis result	**Save size Analysis button**
Sizing Analysis results for multiple studies sizes	**Explore Experiment Size window**
All modalities combination statistical analysis tables	**Analysis All Modalities button**
All modalities combination statistical analysis results (U-statistics and MLE)	**Analysis All Modalities button**

4.2.1 Save Stat Analysis Button

The button exports the analysis result for chosen modalities. It creates a folder with three files:

A. PDF version: A human read friendly verion, includes study size, AUC, AUC standard error, and BDG analysis result (p-value, confidence interval, and hypothesis test result) .

B. CSV version: two-line table, includes study size, AUC, AUC variance, analysis results by Normal, BDG and Hillis method (p-value, confidence interval, and hypothesis test result).

C. OMRMC version: It can be used as **Summary Data** input file with following details.

File Information ①

```
MRMC summary statistics from iMRMC Version 4.0.0
Summary statistics based on input file named:
C:\Users\ReaderStudy\Documents\qigong\iMRMC\imrmc_inputfile\test_new1.imrmc

BEGIN SUMMARY
NReader=  5
Nnormal=  50
NDisease= 50

Modality A = 1
Modality B = 2

Reader-Averaged AUCs
AUC_A = 0.757
AUC_B = 0.716
AUC_A - AUC_B = 0.041
```

Summary Data ②

```
Reader Specific AUCs
MOD A
Reader ID,  Normal_Case,  Disease_Case,        AUC,       STDAUC
        2,           50,            50,   7.78800E-1,   4.48737E-2
        3,           50,            50,   7.89200E-1,   4.45907E-2
        4,           50,            50,   7.28000E-1,   5.11931E-2
        5,           50,            50,   7.48000E-1,   4.80503E-2
  reader1,           50,            50,   7.40000E-1,   4.97574E-2
MOD B
Reader ID,  Normal_Case,  Disease_Case,        AUC,       STDAUC
        2,           50,            50,   7.91600E-1,   4.46634E-2
        3,           50,            50,   8.08800E-1,   4.33999E-2
        4,           50,            50,   6.43200E-1,   5.51967E-2
        5,           50,            50,   6.90400E-1,   5.26176E-2
  reader1,           50,            50,   6.46800E-1,   5.48297E-2
Difference between MODs A and B
Reader ID,  Normal_Case,  Disease_Case,        AUC,       STDAUC
        2,           50,            50,  -1.28000E-2,   4.93243E-2
        3,           50,            50,  -1.96000E-2,   5.33648E-2
        4,           50,            50,   8.48000E-2,   5.88014E-2
        5,           50,            50,   5.76000E-2,   5.30532E-2
  reader1,           50,            50,   9.32000E-2,   5.44117E-2

********************BDG Moments************************
      Moments,        M1,        M2,        M3,        M4,        M5,        M6,        M7,        M8
Modality1(AUC_A), 7.56800E-1, 6.27442E-1, 6.27187E-1, 5.71009E-1, 6.34960E-1, 6.01083E-1, 5.99914E-1, 5.71446E-1,
Modality2(AUC_B), 7.16160E-1, 5.84519E-1, 5.71693E-1, 5.15359E-1, 5.76680E-1, 5.43780E-1, 5.38985E-1, 5.10390E-1,
   comp product, 5.94000E-1, 5.69580E-1, 5.63642E-1, 5.42661E-1, 5.63920E-1, 5.52715E-1, 5.52043E-1, 5.41132E-1,
END SUMMARY

********************StatPanel outputs (U-statistics, Not MLE)************************
Modality A = 1
Modality B = 2

AUC_A = 0.757,   AUC_B = 0.716,   AUC_A - AUC_B = 0.041
```

Statistical Analysis Results ③

```
Statistical Tests:
H0: AUC_A - AUC_B = 0.00,   two-sided alternative,   95% significance,   5 Readers,   50 Normal cases,   50 Disease cases.
    AUC_A = 0.757,   AUC_B = 0.716,   AUC_A - AUC_B = 0.041,   S.E(total) = 4.561E-2,   test statistic = 8.911E-1
            Large Sample Approx(Normal):    p-Value = 0.3729    Conf. Int. = (-0.0487, 0.1300)
                 T-test with df(BDG) = 37.44:   p-Value = 0.3786    Conf. Int. = (-0.0518, 0.1330)
    T-test with df(Hillis 2008) = 52.48:    p-Value = 0.3729    Conf. Int. = (-0.0487, 0.1300),
********************BDG output Results (U-statistics, Not MLE)************************
Moments,M1,M2,M3,M4,M5,M6,M7,M8
comp MA,7.56800E-1,6.27442E-1,6.27187E-1,5.71009E-1,6.34960E-1,6.01083E-1,5.99914E-1,5.71446E-1,
coeff MA,8.00000E-5,3.92000E-3,3.92000E-3,1.92080E-1,3.20000E-4,1.56800E-2,1.56800E-2,-2.31680E-1,
comp MB,7.16160E-1,5.84519E-1,5.71693E-1,5.15359E-1,5.76680E-1,5.43780E-1,5.38985E-1,5.10390E-1,
coeff MB,8.00000E-5,3.92000E-3,3.92000E-3,1.92080E-1,3.20000E-4,1.56800E-2,1.56800E-2,-2.31680E-1,
comp product,5.94000E-1,5.69580E-1,5.63642E-1,5.42661E-1,5.63920E-1,5.52715E-1,5.52043E-1,5.41132E-1,
-coeff product,1.60000E-4,7.84000E-3,7.84000E-3,3.84160E-1,6.40000E-4,3.13600E-2,3.13600E-2,-4.63360E-1,
total,2.27968E-5,2.85376E-4,2.80653E-4,2.00896E-4,2.68160E-5,6.18317E-4,5.45869E-4,9.92413E-5,

********************BCK output Results (U-statistics, Not MLE)************************

Moments,N,D,N~D,R,N~R,D~R,R~N~D
comp MA,2.84685E-2,2.96375E-2,5.40826E-3,-4.36468E-4,2.77091E-2,2.67948E-2,6.77725E-2,
coeff MA,2.00000E-2,2.00000E-2,4.00000E-4,2.00000E-1,4.00000E-3,4.00000E-3,8.00000E-5,
comp MB,2.85951E-2,3.33902E-2,4.30446E-3,4.96916E-3,2.77386E-2,3.57696E-2,7.10026E-2,
coeff MB,2.00000E-2,2.00000E-2,4.00000E-4,2.00000E-1,4.00000E-3,4.00000E-3,8.00000E-5,
comp product,1.09111E-2,1.15829E-2,2.93803E-4,1.52922E-3,1.00700E-2,1.53361E-2,3.14473E-3,
-coeff product,4.00000E-2,4.00000E-2,8.00000E-4,4.00000E-1,8.00000E-3,8.00000E-3,1.60000E-4,
total,7.04828E-4,7.97236E-4,3.65005E-6,2.94850E-4,1.41231E-4,1.27569E-4,1.05989E-5,

********************DBM output Results (U-statistics, Not MLE)************************

Components,R,C,R~C,T~R,T~C,T~R~C
components,1.52922E-3,4.49998E-2,5.09379E-2,7.37126E-4,7.52857E-2,6.98498E-2,
coeff,0.00000E0,0.00000E0,0.00000E0,4.00000E-1,2.00000E-2,4.00000E-3,
total,0.00000E0,0.00000E0,0.00000E0,2.94850E-4,1.50571E-3,2.79399E-4,

********************OR output Results (U-statistics, Not MLE)************************
```

Variance Components Tables ④

① **File Information**

```
MRMC summary statistics from iMRMC Version 4.0.0
Summary statistics based on input file named:
C:\Users\ReaderStudy\Documents\qigong\iMRMC\imrmc_inputfile\test_new1.imrmc
```

Software version and output file name.

② Summary Data

```
BEGIN SUMMARY
NReader= 5
Nnormal= 50
NDisease= 50

Modality A = 1
Modality B = 2

Reader-Averaged AUCs
AUC_A = 0.757
AUC_B = 0.716
AUC_A - AUC_B = 0.041

Reader Specific AUCs
MOD A
Reader ID,  Normal_Case,  Disease_Case,        AUC,      STDAUC
       2,           50,            50,   7.78800E-1,  4.48737E-2
       3,           50,            50,   7.89200E-1,  4.45907E-2
       4,           50,            50,   7.28000E-1,  5.11931E-2
       5,           50,            50,   7.48000E-1,  4.80503E-2
  reader1,          50,            50,   7.40000E-1,  4.97574E-2
MOD B
Reader ID,  Normal_Case,  Disease_Case,        AUC,      STDAUC
       2,           50,            50,   7.91600E-1,  4.46634E-2
       3,           50,            50,   8.08800E-1,  4.33999E-2
       4,           50,            50,   6.43200E-1,  5.51967E-2
       5,           50,            50,   6.90400E-1,  5.26176E-2
  reader1,          50,            50,   6.46800E-1,  5.48297E-2
Difference between MODs A and B
Reader ID,  Normal_Case,  Disease_Case,        AUC,      STDAUC
       2,           50,            50,  -1.28000E-2,  4.93243E-2
       3,           50,            50,  -1.96000E-2,  5.33648E-2
       4,           50,            50,   8.48000E-2,  5.88014E-2
       5,           50,            50,   5.76000E-2,  5.30532E-2
  reader1,          50,            50,   9.32000E-2,  5.44117E-2

********************BDG Moments************************
       Moments,        M1,        M2,        M3,        M4,        M5,        M6,        M7,        M8
Modality1(AUC_A), 7.56800E-1, 6.27442E-1, 6.27187E-1, 5.71009E-1, 6.34960E-1, 6.01083E-1, 5.99914E-1, 5.71446E-1,
Modality2(AUC_B), 7.16160E-1, 5.84519E-1, 5.71693E-1, 5.15359E-1, 5.76680E-1, 5.43780E-1, 5.38985E-1, 5.10390E-1,
    comp product, 5.94000E-1, 5.69580E-1, 5.63642E-1, 5.42661E-1, 5.63920E-1, 5.52715E-1, 5.52043E-1, 5.41132E-1,
END SUMMARY
```

Summary Data for current analysis.

③ Statistical Analysis Results

```
********************StatPanel outputs (U-statistics, Not MLE)************************
Modality A = 1
Modality B = 2

AUC_A = 0.757,   AUC_B = 0.716,   AUC_A - AUC_B = 0.041
Statistical Tests:
H0: AUC_A - AUC_B = 0.00,  two-sided alternative,  95% significance,  5 Readers,  50 Normal cases,  50 Disease cases.
    AUC_A = 0.757,   AUC_B = 0.716,   AUC_A - AUC_B = 0.041,   S.E(total) = 4.561E-2,  test statistic = 8.911E-1
            Large Sample Approx(Normal):   p-Value = 0.3729   Conf. Int. = (-0.0487, 0.1300)
               T-test with df(BDG) = 37.44:   p-Value = 0.3786   Conf. Int. = (-0.0518, 0.1330)
    T-test with df(Hillis 2008) = 52.48:   p-Value = 0.3729   Conf. Int. = (-0.0487, 0.1300),
```

Study information and MRMC Statistical Analysis results (AUCs, Standard Error, Confidence Interval, and Hypothesis test results).

④ Variance Components Tables
MRMC 5 methods statistical analysis variance components tables (component, coefficient).

4.2.2 Save Size Analysis Button

The button exports statistical results for the sizing study. It creates a folder with two files:

A. CSV version: two-line table, includes study size, statistical results by Normal, BDG and Hillis method (degree freedom, standard error, and power).

B. OMRMC version: Has following details.

```
                          MRMC size statistics from iMRMC Version 3p0
 File Information    1     Size statistics written to file named:
                          test_new1MRMCsize201608191651.omrmc


                          ********************Sizing parameters*************************
 Sizing Parameters   2     Effective Size = 0.05,Significance Level = 0.05
                          NReaderSize=  5,NnormalSize=  50,NDiseaseSize= 50
                          ***********************************************************
                          Sizing Results:
                          S.E=4.561E-2
 Sizing Results      3     Large Sample Approx(Normal) ,  Power= 0.19
                          BDG:  df= 37.44,  Power= 0.19
                          Hills:df= 52.48, Power= 0.19
                          ***********************************************************


                          ********************BDG output Results (U-statistics, Not MLE)*************************
                          Moments,M1,M2,M3,M4,M5,M6,M7,M8
                          comp MA,7.56800E-1,6.27442E-1,6.27187E-1,5.71009E-1,6.34960E-1,6.01083E-1,5.99914E-1,5.71446E-1,
                          coeff MA,8.00000E-5,3.92000E-3,3.92000E-3,1.92080E-1,3.20000E-4,1.56800E-2,1.56800E-2,-2.31680E-1,
                          comp MB,7.16160E-1,5.84519E-1,5.71693E-1,5.15359E-1,5.76680E-1,5.43780E-1,5.38985E-1,5.10390E-1,
                          coeff MB,8.00000E-5,3.92000E-3,3.92000E-3,1.92080E-1,3.20000E-4,1.56800E-2,1.56800E-2,-2.31680E-1,
                          comp product,5.94000E-1,5.69580E-1,5.63642E-1,5.42661E-1,5.63920E-1,5.52715E-1,5.52043E-1,5.41132E-1,
 Variance Components 4     -coeff product,1.60000E-4,7.84000E-3,7.84000E-3,3.84160E-1,6.40000E-4,3.13600E-2,3.13600E-2,-4.63360E-1,
 Tables                    total,2.27968E-5,2.85376E-4,2.80653E-4,2.00896E-4,2.68160E-5,6.18317E-4,5.45869E-4,9.92413E-5,


                          ********************BCK output Results (U-statistics, Not MLE)*************************

                          Moments,N,D,N~D,R,N~R,D~R,R~N~D
                          comp MA,2.84685E-2,2.96375E-2,5.40826E-3,-4.36468E-4,2.77091E-2,2.67948E-2,6.77725E-2,
                          coeff MA,2.00000E-2,2.00000E-2,4.00000E-4,2.00000E-1,4.00000E-3,4.00000E-3,8.00000E-5,
                          comp MB,2.85951E-2,3.33902E-2,4.30446E-3,4.96916E-3,2.77386E-2,3.57696E-2,7.10026E-2,
                          coeff MB,2.00000E-2,2.00000E-2,4.00000E-4,2.00000E-1,4.00000E-3,4.00000E-3,8.00000E-5,
                          comp product,1.09111E-2,1.15829E-2,2.93803E-4,1.52922E-3,1.00700E-2,1.53361E-2,3.14473E-3,
                          -coeff product,4.00000E-2,4.00000E-2,8.00000E-4,4.00000E-1,8.00000E-3,8.00000E-3,1.60000E-4,
                          total,7.04828E-4,7.97236E-4,3.65005E-6,2.94850E-4,1.41231E-4,1.27569E-4,1.05989E-5,
```

1 File Information

```
MRMC size statistics from iMRMC Version 3p0
Size statistics written to file named:
test_new1MRMCsize201608191651.omrmc
```

Software version and output file name.

2 Sizing Parameters

```
********************Sizing parameters*************************
Effective Size = 0.05,Significance Level = 0.05
NReaderSize=  5,NnormalSize=  50,NDiseaseSize= 50
```

Parameter for sizing analysis

3 Sizing Results

```
***********************************************************
Sizing Results:
S.E=4.561E-2
Large Sample Approx(Normal) ,  Power= 0.19
BDG:  df= 37.44,  Power= 0.19
Hills:df= 52.48, Power= 0.19
***********************************************************
```

Sizing Analysis results (Standard Error, power).

4 Variance Components Tables

BDG and BCK sizing analysis variance components tables (moment and coefficient for compared two modalities, and their product).

4.2.3 Analyze All Modalities

The button exports all individual modalities and combinations results in 10 csv files

File name	Content
statAnalysis	U-statistics analysis results: AUC, varAUC, analysis results by Normal, BDG and Hillis method (p-value, confidence interval, and hypothesis test result)
statAnalysisMLE	MLE analysis results: AUC, varAUC, analysis results by Normal, BDG and Hillis method (p-value, confidence interval, and hypothesis test result)
ROCcurves	FPF, TPF value for ROC cruves: reader individual and averages (horizontal average, vertical average, diagonal average, pooled average)
AUCperReader	Individual readers AUC and varAUC
readerCovariance	Covariance between readers
BDGtable	BDG moment and coefficient for compared two modalities, and their product
BCKtable	BDK marginal and coefficient for compared two modalities, and their product
DBMtable	DBM component and coefficient
ORtable	OR component and coefficient
MStable	MS component and coefficient

BDG table explanation:

compOrCoeff column	Description
comp_A	BDG 8 moments for modality A AUC variance
coeff_A	Coefficients corresponding to BDG 8 moments for modality A
comp_B	BDG 8 moments for modality B AUC variance
coeff_B	Coefficients corresponding to BDG 8 moments for modality B
comp_product	BDG 8 moments for modality A and B AUC covariance
coeff_product	Coefficients corresponding to BDG 8 moments covariance
Total	comp_A * coeff_A + comp_A * coeff_B - comp_product * coeff_product

Other tables are in similar format.

5. Methods

In this software, we utilize two methods to estimate the MRMC variance components of four widely used variance decompositions of the reader-averaged AUC. Then we use these variance decompositions to size a trial. The software can analyze data from a reader-study of two modalities that is fully crossed with readers and cases paired across modalities (every reader reads every case in two modalities).

5.1 Introduction to ROC Reader Studies

ROC reader studies are designed to evaluate and compare imaging devices and acquisition protocols, or generally evaluate image quality according to an objective task. The ROC task is a classification task, e.g., classifying a patient as non-diseased or diseased. Image quality is then defined as the ability of a reader (e.g., a radiologist) to perform such a task. ROC studies have been used to evaluate imaging and computer-aided detection and diagnosis devices at the FDA[13].

In a typical ROC reader study the reader is presented with one of two mutually exclusive alternatives (e.g. a tumor-present image or a tumor-absent image). The observer is then asked to rate his or her confidence level of which alternative is presented (e.g., the confidence level of tumor presence on an image). Any number of responses may be used to rate the confidence level. For example, in a traditional clinical reader study, a set of five confidence level responses is used with 1 representing "absolutely sure there is no tumor" and 5 representing "absolutely sure there is tumor present". Alternatively, reader studies may ask the observer to use a "continuous" rating scale. Such scales are not really continuous but allow the reader to rate each case with a whole number ranging from 1 to 100. The rating values are collected for both non-diseased and diseased cases. [Example electronic case report form], [Example Instructions for ROC scores]

Given ratings for non-diseased and diseased cases, an ROC curve can be traced out by calculating the sensitivity/specificity (TPF/TNF) pair for each confidence level, or threshold, possible [1]. An ROC curve illustrates the tradeoff between sensitivity and specificity of the reader across all thresholds. This tradeoff is realized by a change in the reader's threshold. In the case of breast cancer screening via mammography, when the threshold is made more aggressive the reader recalls more patients for additional imaging, increasing his or her sensitivity at the price of lower specificity. If the reader's threshold is moved in the opposite direction, the reader will recall fewer patients; the reader is less aggressive, decreasing his or her sensitivity with the concomitant result of increased specificity. The area under this ROC curve (AUC) is a summary figure-of-merit for describing how well a reader is able to separate the population of diseased patients from non-diseased patients.

One interpretation of AUC is that it is the reader's average sensitivity over all possible specificities. As such, it is a global summary of task performance that avoids thresholds entirely. AUC is also mathematically equivalent to the probability that a random reader will correctly choose the signal-present image over the signal-absent image when a pair is presented side-by-side or sequentially, as is done in a 2-alternative forced choice (2AFC) task [12].

To account for the variability in readers, an ROC study is often conducted in a multi-reader paradigm. The endpoint of such an ROC reader study is the reader-averaged AUC value. The uncertainty in the reader-averaged AUC suffers from two sources of variability: the readers and the cases. To account for both sources of variability, reader studies often involve several trained readers in addition to a dataset of diseased and non-diseased cases. One popular study design for estimating AUC is the fully crossed study design in which every reader reads every case.

Statistical methods have been proposed in the literature to analyze MRMC data [12][13]. The origin of each method differs, and consequently, the estimation process of each method differs. Additionally, each method has at its foundation a different decomposition, or representation, of the total variance. These are discussed in the next section "Variance Estimation".

Another study design, the split-plot, was introduced the field of medical imaging, borrowed from studies in agriculture [12]. In a split-plot study the readers and cases are split into groups. Then all the readers in a group read all the case in their group. This kind of study can be much less burdensome than a fully-crossed without sacrificing much statistical precision. Current study shows that paired split-plot (PSP) design can gain statistical precision as compared with the FC design with the same number of readers and readings [14].

5.2 Variance Estimation

The first variance estimation method uses U-statistics to provide unbiased estimates of the variance components [4]. This method lacks a positivity constraint and can lead to negative estimates of variance components and total variance. The second variance estimation method uses the non-parametric maximum likelihood estimate (MLE) of the distribution of readers and cases, which is the empirical distribution of readers and cases [2]. Efron and Tibshirani also refer to this estimation method as the "ideal" bootstrap. The MLE estimate of variance components and total variance cannot go negative. The tradeoff for positive variance estimates is a positive bias.

5.3 Components of Variance

Regardless of their original developments, the four variance decompositions that are included in this software can be derived from first principles for the reader average of *empirical* AUCs. As such they are related to one another with simple mappings [3]. The decompositions are:

BDG components [3-5]: The author Brandon D. Gallas (BDG) decomposed the total variance into eight moments from first principles in a fashion equivalent to U-statistics, which decomposes the total variance into seven conditional covariances. The "extra moment" is the mean squared, which is a part of each conditional covariance. There is a term for non-diseased cases, diseased cases, readers, and all combinations. The decomposition treats non-diseased cases separately from diseased cases such that the total variance can be easily generalized to new readers, new non-diseased cases, and new diseased cases.

BCK components [6]: The authors Barrett, Clarkson, and Kupinski (BCK) decomposed the total variance into seven marginal variances from first principles. They are marginal in the sense that they average over the non-random effects rather than conditioning on them as above. The BCK components are thought of as pure variance terms. There is a term for non-diseased cases, diseased cases, readers, and all combinations. The BCK decomposition treats non-diseased cases separately from diseased cases such that the total variance can be easily generalized to new readers, new non-diseased cases, and new diseased cases.

DBM components [7]: The authors Dorfman, Berbaum, and Metz (DBM) decomposed the total variance into six components based on a mixed-model ANOVA. The components are: reader effect, case effect, reader-case effect, modality-reader effect, modality-case effect, and modality-reader-case effect. The DBM decomposition does not treat non-diseased cases separately from diseased cases. So, while the total variance can be easily generalized to new readers and to total cases (with a fixed disease prevalence), additional modeling is needed to separately generalize to new non-diseased cases and new diseased cases (i.e., population with arbitrary disease prevalence).

MS components [7]: The MS decomposition is based on the same mixed-model ANOVA as the DBM components. MS stands for mean squares, which are estimated from the data first and then mapped to the DBM components. There are six MS components: reader effect, case effect, reader-case effect, modality-reader effect, modality-case effect, and modality-reader-case effect. The MS decomposition does not treat non-diseased cases separately from diseased cases. So, while the total variance can be easily generalized to new readers and to total cases (with a fixed disease prevalence), additional modeling is needed to separately generalize to new non-diseased cases and new diseased cases (i.e., population with arbitrary disease prevalence).

OR components [8]: The authors Obuchowski and Rockette (OR) decomposed the total variance into six components based on a two-factor ANOVA by modeling the accuracy of the jth reader using the ith diagnostic test. The components are: reader effect, modality-reader effect, same-reader-different-modality covariance, different-reader-same-modality covariance, different-reader-different-modality covariance, and residual error. The OR decomposition does not treat non-diseased cases separately from diseased cases. So, while the total variance can be easily generalized to new readers and to total cases (with a fixed disease prevalence), additional modeling is needed to separately generalize to new non-diseased cases and new diseased cases (i.e., population with arbitrary disease prevalence).

We emphasize that the software presents the four different variance decompositions of the reader average of *empirical* AUCs estimated by U-statistics and MLE [3]. The software does not estimate the DBM and OR components as originally proposed [7, 8]. They are obtained through linear combinations of the BDG components [3]. It is worth pointing out that the U-statistics and MLE estimation methods in this software are specific to the reader average of *empirical* AUCs, whereas the original DBM and OR *estimation* methods can be used for other performance measures. Estimation of components when using fully-crossed data is applicable to the BDG, BCK, DBM, OR and MS decompositions. If using data with a non-fully-crossed study design, only the BDG and BCK components are currently able to be estimated. Calculations for the other decompositions are in development.

5.4 Hypothesis Testing and Confidence Intervals:

• The null hypothesis for comparing two modalities is AUC_A - AUC_B = 0.00 (or AUC = 0.5 for signal modality).

• The test statistic is the (AUC_A - AUC_B) / se(AUC_A - AUC_B) for all methods.

 i. For the Large Sample Approx(Normal) results, we assume the test statistic comes from a normal distribution. This is referred to as a Wald or z-test.

 ii. For the T-test with df(BDG) results: we assume the test statistic comes from a t-distribution with df reported (df = degrees of freedom are presented in Obuchowski et al. 2012 [11]). This is a t-test.

 iii. For the T-test with df(Hillis) results: we assume the test statistic comes from a t-distribution with df reported (df = degrees of freedom are presented in Hillis et al. 2008 [10]). This is a t-test.

• Each method results include: p-value of the null hypothesis and the confidence interval on the difference (or the confidence around the single modality AUC).

5.5 Sizing a Future Study

The software calculates the expected statistical power of the new study. The software implements two kinds of hypothesis testing. One is for a single modality; the null hypothesis is that the AUC is equal to a specified value. The second compares modalities, the null hypothesis is that the two modalities have the same AUC. The statistical power is computed in two ways. One uses the **Z-test**, and the other uses the **F-test** from Hillis [9].

Z-test:

Assumes the variance is known even though it is actually estimated. The Z-test uses the ratio of the affect size over the square root of its estimated variance as the test statistic, and assumes the test statistic follows a standard normal distribution.

F-test:

Does not assume the variance is known. It estimates the non-centrality parameter and the denominator degrees of freedom values using the components of variance (input or estimated) for the specified case and reader sample size and effect size. The degrees of freedom is also determined with an alternate calculation from Obuchowski et al. 2012 [11]. The statistical power is then computed using F distributions, unless the degrees of freedom is greater than 50, at which point the power is computed using a standard normal distribution.

Status

The sizing analysis has been debugged and enabled. We believe all elements are now functioning properly. Confirmatory validation of statistical and sizing analysis by Roe and Metz simulation is in processing. Please contact us with any issues via the issues page.

6. Database

At the moment we do not have an available database, we are in the process of creating an updated and fully functional database.
Stay up to date by checking the GitHub release page every few weeks.

The software includes a database that consists of results from simulated data and results from previous MRMC reader studies. We have accrued many reader study datasets and are working on the letters of permission to include them. The datasets are from a variety of sources including FDA sponsor data for imaging device approvals, as well as academic research studies. Each dataset contains

> \# **Basic information of the study:** this includes the source of the study, a summary of the study, related publications, and other information.

> \# **Key information of the study:** number of readers, number non-diseased cases, number of diseased cases, modalities used, and task.

> \# **Variance representations** as outlined above**.**

See Appendix A for details regarding the summary information.

At the bottom of the applet means we are sharing the whole database (basic information of each study, size of the study, and variance decompositions). The "database" is populated by simulated datasets right now. The buttons at the bottom are related to downloading a spread sheet of that data (work in progress).

7. References and Other Information

References, Abbrevations, and Appendixes used in the iMRMC Manual.

7.1 References

[1] Metz, C. E., "Basic principles of ROC analysis.", Semin Nucl Med. 1978 Oct;8(4):283-98.

[2] Efron, B. & Tibshirani, R. J, "An Introduction to the Bootstrap: Monographs on Statistics and Applied Probability", Chapman & Hall, New York, N.Y, 1993.

[3] Gallas, B. D.; Bandos, A.; Samuelson, F. & Wagner, R. F. (2009), 'A Framework for Random-Effects ROC Analysis: Biases with the Bootstrap and Other Variance Estimators', Commun Stat A-Theory 38(15), 2586-2603

[4] Gallas, B. D. (2006), 'One-Shot Estimate of MRMC Variance: AUC', *Acad Radiol* **13**(3), 353-362.

[5] Gallas, B. D. & Brown, D. G. (2008), 'Reader Studies for Validation of CAD Systems', *Neural Networks* **21**(2-3), 387-397.

[6] Clarkson, E.; Kupinski, M. A. & Barrett, H. H. (2006), 'A Probabilistic Model for the MRMC Method. Part 1.Theoretical Development', *Acad Radiol* **13**(11), 1410-1421.

[7] Dorfman, D. D.; Berbaum, K. S. & Metz, C. E. (1992), 'Receiver Operating Characteristic Rating Analysis: Generalization to the Population of Readers and Patients with the Jackknife Method', *Invest Radiol* **27**(9), 723-731.

[8] Obuchowski, N. A. (1995), 'Multireader, Multimodality Receiver Operating Characteristic Curve Studies: Hypothesis Testing and Sample Size Estimation Using an Analysis of Variance Approach with Dependent Observations', *Acad Radiol* **2**(Suppl 1), S22-S29.

[9] Hillis SL, Obuchowski NA, Berbaum KS. (2011), 'Power estimation for multireader ROC methods: An updated and unified approach'. *Acad Radiol*, **18**(2), 129-142.

[10] Hillis, S. L.; Berbaum, K. S. & Metz, C. E. (2008), 'Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis.', *Acad Radiol* **15**(5), 647--661.

[11] Obuchowski, N.A.; Gallas, B. D. & Hillis, S. L. (2012), 'Multi-reader ROC Studies with Split-plot Designs: A Comparison of Statistical Methods', *Acad Radiol* **19**(12), 1508-1517.

[12] Wagner, R. F.; Metz, C. E. & Campbell, G. (2007), 'Assessment of Medical Imaging Systems and Computer Aids: A Tutorial Review.' *Acad Radiol*, **14**(6), 723-748.

[13] Gallas, B. D.; Chan, H.-P.; D'Orsi, C. J.; Dodd, L. E.; Giger, M. L.; Gur, D.; Krupinski, E. A.; Metz, C. E.; Myers, K. J.; Obuchowski, N. A.; Sahiner, B.; Toledano, A. Y. & Zuley, M. L. (2012), 'Evaluating Imaging and Computer-aided Detection and Diagnosis Devices at the FDA.' *Acad Radiol*, **19**(4), 463-477.

[14] Chen, W; Gong, Q; Gallas, B. D, 'Paired split-plot designs of multireader multicase studies,' J. Med. Imag. **5**(3), 031410 (2018)

7.2 Related Papers and Presentations

[1] Gallas, B. D.; Pisano, E.; Cole, E. & Myers, K. (2017), "Impact of different study populations on reader behavior and performance metrics: initial results." In Medical Imaging 2017: Image Perception, Observer Performance, and Technology, Matthew A. Kupinski & Robert M. Nishikawa, ed., 10136, pp. 0A. (Link to presentation)

[2] Gallas, B. D. (2015),"Resources for Reader Studies." SPIE Medical Imaging Conference Joint Workshop: Physics & Image Perception, Observer Performance, and Technology, SPIE - the international society for optics and photonics. (Link to presentation)

[3] Gallas, B. D.; Gong, Q. & Myers, K. J. (2015),"Training Readers to Use Multi-Level ROC Scores." MIPS Conference XVI, Medical Imaging Perception Society. (Link to presentation)

[4] Gallas, B. D. & Hillis, S. L. (2014), 'Generalized Roe and Metz ROC Model: Analytic Link Between Simulated Decision Scores and Empirical AUC Variances and Covariances.' J Med Img, 1, (3), 031006. (Link to paper)

[5] Gallas, B. D. (2013), "ROC Analysis: A tribute to Charlie Metz and an assessment of the State of the Art." In SPIE Medical Imaging Conference Workshop: Image Perception, Observer Performance, and Technology, Proc SPIE. (Link to presentation)

[6] Gallas, B. D. (2012),"Multi-Reader Diagnostic Imaging Case Studies: Components of Variance." Joint Statistical Meetings. (Link to presentation)

7.3 Abbrevations

Abbrevation	Meaning
AFC	Alternative Forced Choice
ROC	Receiver operating characteristic
MRMC	Multi Reader – Multi Case Studies
MLE	Maximum Likelihood Estimate
AUC	Area under the curve
JRE	Java Runtime Enivronment
ANOVA	Analysis of Variance
BDG, BCK, DBM, OR, MS	Name Abbrevations

7.4 Appendix A

Basic Information

The following fields summarize the data source, setting, experimental conditions, and reference material related to the reader ROC ratings.

1. **Source: (Owner of data, free text)** Data owner or representative, affiliation, mailing address, phone, and email.
2. **Summary: (free text)** A succinct description of the reader study (readers, cases, reference method, rating method) with more details on the task and modalities (treatments) under investigation.
3. **Publication: (free text)** Full references to articles that describe or utilize the reader study or data. This field should also include other sources of information related to the study or data (links to webpages, protocol identifiers).
4. **Notes: (free text)** This field can alert the user to any peculiar issues related to the data. For example, is there any missing data or corrections to the data?
5. **Acknowledgements: (free text)**
6. **NR: (Number of readers, integer)**
7. **N0: (Number of normal cases, integer)**
8. **N1: (Number of abnormal cases, integer)**
9. **NM: (Number of modalities, integer)**
10. **AUC1: (AUC Modality 1, floating point)**
11. **AUC2: (AUC Modality 2, floating point)**
12. **…**
13. **AUCn: (AUC Modality n, floating point)**
14. **DAUC: (Difference in AUCs, floating point)**

Key Information

The following items are critical for determining the application area of the components of (co)variance. They may be renamed and regrouped over time as deemed useful.

15. **Mod1: (Modality 1, limited choices)**

16. **Mod2: (Modality 2, limited choices)**
    a. simulation
    b. film radiography
    c. digital radiography
    d. film with CAD
    e. film mammography
    f. digital mammography
    g. SE MRI
    h. CINE MRI
    i. Multi-detector row CT
    j. Digital breast tomosythesis
    k. Mammography + ultrasound
    l. Mammography + ultrasound + CAD
    m. Ultrasound
    n. Ultraound + CAD
17. **Task: (limited choices)**
    a. simulation
    b. chest and abdominal abnormality detection
    c. breast abnormality detection
    d. breast cancer detection
    e. detection of thoracic aortic dissection
18. **TaskAdd: (limited choices)** This field indicates a modification or qualifier to the task. Multiple qualifiers may be indicated. More liberty will be given to data suppliers in defining these choices. This list is expected to grow.
    a. pediatric
    b. malignant mass
    c. malignant calcification

    d.   nodule
    e.   pneumothorax
    f.   interstitial disease