**Chapter 11: Online Appendices**

# Online Appendix 11.A: Random-reader random-case (RRRC) formulae

This appendix is restricted to comparing two treatments. This section gives the sample-size formulae for the DBMH and ORH methods regarding both readers and cases as random factors.

## Online Appendix 11.A.1: DBM non-centrality parameter

According to book Equation 9.34, the non-centrality parameter ∆ is (*J* = number of readers and *K* = number of cases in the *pivotal* study and the *Y* subscript is used to denote pseudovalue derived variances and mean-squares):

 .

Here *d* is the effect size:

 .

This result is stated in Eqn. 6 in Hillis and Berbaum1. The Hillis-denominator degrees of freedom for the DBM model is (book Equation 9.22; the subscript Y, indicating pseudovalue derived quantities, has been added since in this section both pseudovalue and FOM based quantities are being used):

 .

The critical value of the F-distribution is  and a significant result is obtained when the observed value of the DBMH F-statistic, defined in book Equation 9.23, exceeds the critical value. The expected statistical power is:

 .

Here is a sample from the non-central F-distribution with non-centrality parameter  and numerator and denominator degrees of freedom equal to 1 and **, respectively.

## Online Appendix 11.A.2: OR non-centrality parameter derivation

The starting point is Equation . It can be rewritten as:

 .

In the ORH method a distinction needs to be made between the numbers of cases in the pilot study, , and that in the pivotal study, . All variances on the right hand side of Equation are estimated from the pilot dataset. It is necessary to transform them to the corresponding OR parameters. The necessary correspondences are reproduced below (Table 5 in Ref. 2):

Table : Correspondences between DBM parameters (first column) and OR parameters (second column); reproduced from Table 5 in Ref. 2.

|  |  |
| --- | --- |
| DBM | OR |
|  |  |
|  |  |
|  |  |

Using these relations, it follows from Equation that:

 .

 .

 .

 .

As per Hillis recommendation, in going from Equation to Equation a non-negativity constraint has been imposed on . Hillis3 further recommends that either a similar constraint be imposed on  or a positive value should be posited.

Since the OR covariances are determined from the pilot study, one must account for the inverse dependence of covariances on the number of cases – if the number of cases is doubled, the covariances will be halved compared to those determined from the pilot study. This explains the factor  in Equation .

[The inverse scaling is not present in the DBM formulae as these are expressed in terms of pseudovalue variance components, which are intrinsic quantities whose *estimates* are independent of the numbers of cases. The *accuracy* or *variability* of the estimates do depend on the numbers of cases: the estimates become more accurate, i.e., less variable, with increasing numbers of cases.]

## Online Appendix 11.A.3: OR degrees of freedom derivation

The starting point is book Equation 10.45, shown below for *I* = 2 (i.e., two treatments).

 .

All values on the right hand side are estimated from the pilot study with  cases and the question is how to transform these values to a study with  cases. To this end one needs to express  in terms of the covariance matrix parameters.

Equation 9 in Ref. 2 can be rewritten as:

 .

Using this result and temporarily ignoring the non-negativity constraints, the numerator of the ratio in Equation is:

 .

To account for the differences in numbers of cases between pilot and pivotal studies, all variances and covariances involving case-sampling need to be multiplied by . Therefore, Equation can be expressed as:

 .

To ensure non-negativity of the degrees of freedom, Hillis imposes a constraint on the  term. Therefore, the Hillis degrees of freedom for the pivotal study is:

 .

The inverse scaling is not applied to the treatment reader variance component, which is independent of case-sampling variability. [The argument is similar to the reason for not applying a similar scaling to the DBM variance components; see discussion following Equation .]

Finally, the expected power is given by Equation .

# Online Appendix 11.B: Fixed-readers random-case (FRRC) formulae

The formulae given above apply to random-readers random-cases. The formulae for fixed reader and fixed case analysis are, to the best of the author's knowledge, unpublished. The author has held on to a pdf document titled "*MRMC Sample Size Program User Guide*" by Hillis and Berbaum, downloaded around Feb 2011 from the U of Iowa website. This document is no longer available on the cited website but it is included in the supplementary material. It describes usage of the DBM method for sample size estimation including the special cases of FRRC and RRFC generalizations. The basic idea is that for FRRC and RRFC analyses, for the non-centrality parameter one sets  and , respectively, and for the denominator degrees of freedom one sets  and , respectively.

In Equation , setting , the non-centrality parameter for the DBM method for FRRC generalization is:

 .

To obtain the corresponding ORH expression, one starts by setting  in Equation . The non-centrality parameter is:

 .

With appropriate substitutions from Table 1, this can be expressed as follows:

 .

Applying the non-negativity constraint,

 .

The corresponding *ddf* is defined by:

 .

The expected power is given by Equation .

# Online Appendix 11.C: Random-reader fixed-cases (RRFC) formulae

This time one sets  and  in Equation . The non-centrality parameter for the DBM method for RRFC generalization is:

 .

The corresponding expression for the ORH method is obtained by setting  in Equation and making appropriate substitutions from Table 1 (setting  implies ) and imposing a positivity constraint:

 .

The corresponding *ddf* is defined by:

 .

The expected power is given by Equation .

# Online Appendix 11.D: Power table

A power table is a listing of different combinations of numbers of readers and cases that are expected to yield a desired power, typically chosen to be 80%. This allows the user to choose the most appropriate combination consistent with ones resources. The following code listing, in file mainSsDbmh.R, shows an application of the DBMH method to generate a power table. Dataset02 corresponds to the Van Dyke data file4. The DBMH method is invoked at line 12. [Dataset03 corresponds to the Franken data file5. Online Appendix 24 details how these "precompiled" datasets were generated. RStudio reads them, when needed, from the file datasets.RData in the ~/RJafroc/data directory; the OptimisticScenario flag, currently set to FALSE, is explained in book section 11.8.6.]

This section shows, via an example, that the DBMH and ORH methods give identical results for all three desired generalizations, RRRC, FRRC and RRFC.

## Online Appendix 11.D.1: Code Listing

rm(list = ls()) #mainSsDbmh.R

library(RJafroc)

OptimisticScenario <- FALSE

fileName <- "VanDyke.lrc" # fileName <- "Franken1.lrc"

cat("File name = ", fileName, "\n")

if (fileName == "VanDyke.lrc") dataset <- dataset02 else dataset <- dataset03

retDbm <- StSignificanceTesting(

dataset02,

FOM = "Wilcoxon",

method = "DBMH")

varYTR <- retDbm$varComp$varComp[3]

varYTC <- retDbm$varComp$varComp[4]

varYEps <- retDbm$varComp$varComp[6]

effectSize <- retDbm$ciDiffTrtRRRC$Estimate

cat("effect size =", effectSize, "\n")

effectSize <- abs(effectSize)

sigma <- (retDbm$ciDiffTrtRRRC$`CI Upper`

-retDbm$ciDiffTrtRRRC$`CI Lower`)/4

if (OptimisticScenario == TRUE) {

if (fileName == "VanDyke.lrc") {

effectSize <- effectSize -2\*sigma

}

if (fileName == "Franken1.lrc") {

effectSize <- effectSize +2\*sigma

}

}

cat("p-value = ", retDbm$pRRRC,

"\nanticipated effectSize = ", effectSize,

"\nCI Lower =", retDbm$ciDiffTrtRRRC$`CI Lower`,

"\nCI Upper =", retDbm$ciDiffTrtRRRC$`CI Upper`, "\n")

powTab <- SsPowerTable(

effectSize = effectSize,

method = "DBMH",

varYTR = varYTR,

varYTC = varYTC,

varYEps = varYEps)

print(powTab)

Line 9 – 12 applies DBMH significance testing to the pilot dataset, the Van Dyke dataset in the shown code. Lines 13 – 15 extract the three needed DBM variance components. Again, the Y indicates that these are pseudovalue-based quantities. The observed effect size is extracted at line 16. If the OptimisticScenario flag is TRUE, the anticipated effect size is chosen as shown in lines 22 – 27. In the current example, the effect size is set equal to the observed effect size. The power table is generated at line 34 – 39.

If one sources the code, one sees (book section 11.8.4; the first column, which is the line number of the output, can be ignored; the information is in the remaining 3 columns; notice that the numbers under power are close to 0.8 – these are the actual values as yielded by Eqn. which are expected to be close to the desired power – this is an argument to SsPowerTable that does not need to be explicitly stated since the default value is 0.8):

## Online Appendix 11.D.2: Code Output

> source(...)

alpha = 0.05

$powerTableRRRC

numReaders numCases power

1 3 >2000 <NA>

2 4 1089 0.8

3 5 344 0.801

4 6 251 0.801

5 7 211 0.801

6 8 188 0.801

7 9 173 0.801

8 10 163 0.802

...

$powerTableFRRC

numReaders numCases power

1 3 209 0.800

2 4 182 0.801

3 5 166 0.802

4 6 155 0.801

5 7 147 0.801

6 8 141 0.801

7 9 137 0.802

8 10 133 0.801

...

$powerTableRRFC

numReaders numCases power

1 3 >2000 <NA>

2 4 >2000 <NA>

3 5 289 0.8

4 6 151 0.801

5 7 102 0.8

6 8 78 0.803

7 9 63 0.804

8 10 53 0.805

...

The last three lines correspond to the results in the first row of the results in book Table 11.2.

The following code listing, in file mainSsOrh.R, shows an application of the ORH method to generate a power table. Notice that unlike the DBMH method, the ORH method needs to know the number of cases in the pilot study. Also, while the ordering of the parameters makes no difference, the parameter names are case sensitive.

## Online Appendix 11.D.3: Code Listing

rm(list = ls()) #mainSsOrh.R

library(RJafroc)

OptimisticScenario <- FALSE

fileName <- "VanDyke.lrc" # fileName <- "Franken1.lrc"

cat("File name = ", fileName, "\n")

if (fileName == "VanDyke.lrc") dataset <- dataset02 else dataset <- dataset03

KStar <- length(dataset$NL[1,1,,1])

retOR <- StSignificanceTesting(

dataset,

FOM = "Wilcoxon",

method = "ORH")

effectSize <- retOR$ciDiffTrtRRRC$Estimate

varCompOR <- retOR$varComp

varTR <- varCompOR$varCov[2]

cov1 <- varCompOR$varCov[3]

cov2 <- varCompOR$varCov[4]

cov3 <- varCompOR$varCov[5]

varEps <- varCompOR$varCov[6]

varTR <- max(varTR,0)

effectSize <-retOR$ciDiffTrtRRRC$Estimate

sigma <- (retOR$ciDiffTrtRRRC$`CI Upper`-retOR$ciDiffTrtRRRC$`CI Lower`)/4

if (OptimisticScenario == TRUE) {

if (fileName == "VanDyke.lrc") {

effectSize <- effectSize -2\*sigma

}

if (fileName == "Franken1.lrc") {

effectSize <- effectSize +2\*sigma

}

}

cat("p-value = ", retOR$pRRRC, ", postulated effectSize = ",

effectSize, ", CI Lower =", retOR$ciDiffTrtRRRC$`CI Lower`,

", CI Upper =", retOR$ciDiffTrtRRRC$`CI Upper`, "\n")

powTab <- SsPowerTable(

effectSize = effectSize,

method = "ORH",

KStar = KStar,

varTR = varTR,

cov1 = cov1,

cov2 = cov2,

cov3 = cov3,

varEps = varEps)

print(powTab)

If one sources the code, one sees an output identical to that in Online Appendix 11.A.4.

# Online Appendix 11.E: Details on effect-size specification

This section is related to book section 11.12. Listed below is the code mainEffectSizeFixedAucEs.R to demonstrate the consequence of specifying a fixed effect-size in AUC units, 0.05 in the code, on the implied increase in the separation parameter necessary to achieve that effect-size.

### Online Appendix 11.E.1: Code Listing

rm(list = ls()) #mainEffectSizeFixedAucEs.R

source("AzDeePrimeTransformations.R")

require(ggplot2)

esAuc = 0.05

nBins <- 100

AzNh <- seq(0.55,1,length.out = nBins)

esDp <- effectSizeDeePrime(esAuc, AzNh)

AzNh <- AzNh[1:length(esDp)]

esDpRatio <- esDp / azToDeePrime(AzNh)

ConstAucPlot <- rep(esAuc, length.out = length(esDpRatio))

df <- data.frame(Az = c(AzNh, AzNh),DpMultipler = c(ConstAucPlot, esDpRatio),

truth = c(rep('ES: dp multiple', length(AzNh)), rep('Const. Az ES', length(AzNh))))

myPlot <- ggplot(df, aes(x = Az, y = DpMultipler, color = truth)) +

geom\_line(size = 1) +

scale\_colour\_manual(values=c("black","darkgrey")) +

theme(legend.title = element\_blank(), legend.position = c(0.52, 0.85)) +

theme(axis.title.x = element\_text(size = 30,face="bold"),

axis.title.y = element\_text(size = 30,face="bold")) +

theme(legend.text = element\_text(size = 15, face = "bold"),legend.key.size = unit(2.5, "lines"))

print(myPlot)

Line 5 specifies the effect-size, 0.05 in AUC units. Line 7 defines AzNh , an equally spaced Az null hypothesis array of length 100 ranging from 0.55 to 1. Line 8 calls function effectSizeDeePrime() to convert the specified AUC effect-sizes to  effect-sizes esDp. The reader should examine the code (using the debug abilities of RStudio) to be convinced that this is indeed what it does. The esDp array is of shorter length than 100, since some of the AUC values would cause the AH AUC to exceed unity; line 9 discards those values. Line 10 divides the effect-sizes in  units by the baseline  values, and the ratio esAucFixed is the quantity plotted as the heavy line in book Figure 11.4 (a). For comparison, the light line shows the constant AUC effect-size that was used in this code.

Listed below is the code mainEffectSizeFixedDpMultiple.R to demonstrate the consequence of specifying a fixed effect-size in  units, expressed as a multiple of the baseline , 0.2 in the code, on the implied increase in AUC to achieve that effect-size.

### Online Appendix 11.E.2: Code Listing

rm(list = ls()) #mainEffectSizeFixedDpMultiple.R

source("AzDeePrimeTransformations.R")

require(ggplot2)

nBins <- 100

dpMultiple = 0.2

azNh <- seq(0.55,0.99,length.out = nBins)

dpAH <- azToDeePrime(azNh) \* (1 + dpMultiple)

azAh <- deePrimeToAz(dpAH)

esAz <- azAh - azNh

ConstDpPlot <- rep(dpMultiple, length.out = nBins)

df <- data.frame(Az = c(azNh, azNh),esAuc = c(esAz, ConstDpPlot),

truth = c(rep('Az ES for fixed dp ES', length(azNh)), rep('ES: dp multiplier', length(azNh))))

myPlot <- ggplot(df, aes(x = Az, y = esAuc, color = truth)) +

geom\_line(size = 1) +

scale\_colour\_manual(values=c("black","darkgrey")) +

theme(legend.title = element\_blank(), legend.position = c(0.52, 0.7)) +

theme(axis.title.x = element\_text(size = 30,face="bold"),

axis.title.y = element\_text(size = 30,face="bold")) +

theme(legend.text = element\_text(size = 15, face = "bold"),legend.key.size = unit(2.5, "lines"))

print(myPlot)

Line 6 specifies a fixed effect-size 0.2 as a multiple of . Three values are used, 0.02, 0.04 and 0.06. Line 8 uses the function azToDeePrime() to convert the specified NH AUC effect-sizes to  effect-sizes and multiplies the results by to get dpAH. Line 9 and 10 converts this array to an array of effect-sizes in AUC units, esAz, which is plotted as the heavy line in book Figure 11.4 (b - d). For comparison, the light line shows the constant  effect-size that was used in this code.

# Online Appendix 11.F: Details of the sample size estimation process

In the examples shown above, the details of the sample size estimation procedure were "cloaked" in an RJafroc function. The file mainSsOrhDetails.R, for "*sample size ORH method details*" shows the steps of the sample size estimation procedure using the ORH method.

### Online Appendix 11.E.2: Code Listing

rm(list = ls());#mainSsOrhDetails.R

library(RJafroc)

fileName <- "VanDyke.lrc"

#fileName <- "Franken1.lrc"

rm(list = ls());#mainSsOrhDetails.R

library(RJafroc)

fileName <- "VanDyke"

#fileName <- "Franken"

cat("fileName = ", fileName, "\n")

if (fileName == "VanDyke") rocData <- dataset02 else rocData <- dataset03

KStar <- length(rocData$NL[1,1,,1])

JStar <- length(rocData$NL[1,,1,1])

retORH <- StSignificanceTesting(rocData,FOM = "Wilcoxon", method = "ORH")

effectSize <- c(retORH$ciDiffTrtRRRC$Estimate)

d <- effectSize

Cov1 <- retORH$varComp$varCov[3]

Cov2 <- retORH$varComp$varCov[4]

Cov3 <- retORH$varComp$varCov[5]

Var <- retORH$varComp$varCov[6]

VarTR <- retORH$varComp$varCov[2]

VarTR <- max(VarTR,0)

cat("VarTR = ", VarTR, "Var = ", Var,

", Cov1 = ", Cov1, ", Cov2 = ", Cov2, ", Cov3 = ", Cov3, ", d = ", d, "\n")

alpha <- 0.05

JArr <- c(JStar, 2\*JStar)

KArr <- c(KStar, 2\*KStar)

optionArr <- c("RRRC", "FRRC", "RRFC")

for (j in 1:(length(JArr))) {

for (k in 1:(length(KArr))) {

J <- JArr[j]

K <- KArr[k]

DeltaNum <- J\*d^2/2

cat("J = ", J, ", K = ", K, "\n")

for (o in 1:(length(optionArr))) {

option <- optionArr[o]

if (option == "RRRC") {

DeltaDenom <- max(VarTR,0)+(KStar/K)\*(Var-Cov1+(J-1)\*max(Cov2-Cov3,0))

Delta <- DeltaNum/DeltaDenom

ddfNumSqBrkt <- (KStar/K)\*(Var-Cov1+(J-1)\*max(Cov2-Cov3,0))

ddfDenomSqBrkt <- (KStar/K)\*(Var-Cov1-max(Cov2-Cov3,0))

ddf <- (J-1)\*((VarTR+ddfNumSqBrkt)/(VarTR+ddfDenomSqBrkt))^2

} else if (option == "FRRC") {

DeltaDenom <- (KStar/K)\*(Var-Cov1+(J-1)\*max(Cov2-Cov3,0))

Delta <- DeltaNum/DeltaDenom

ddf <- (K-1)

} else if (option == "RRFC") {

DeltaDenom <- max(VarTR,0)+(KStar/K)\*(Var-Cov1-max(Cov2-Cov3,0))

Delta <- DeltaNum/DeltaDenom

ddf <- (J-1)

} else stop ("Bad value for option")

power <- SsPowerGivenJK(

J,K,

effectSize = d,

option = option,

method = "ORH",

cov1 = Cov1,

cov2 = Cov2,

cov3 = Cov3,

varTR = VarTR,

varEps = Var,

KStar = KStar)

if (o == 1) power <- power$powerRRRC

if (o == 2) power <- power$powerFRRC

if (o == 3) power <- power$powerRRFC

FCrit <- qf(1 - alpha, 1, ddf)

Power <- pf(FCrit, 1, ddf, ncp = Delta, FALSE)

cat(option, "FCrit = ", FCrit, ", ddf = ", ddf, ", Delta = ", Delta, ", power = ", Power,

", Power RJafroc = ", power, "\n")

next

}

}

}

Since line 5 is currently commented, line 7 selects the Van Dyke pilot dataset, dataset02. Line 8 initializes , the number of cases in the pilot study (114). Line 10 performs ORH significance testing, saving the results to retORH. Line 11 extracts the effect size. Lines 12 – 17 extract the OR variance components and applies a non-negativity constraint. Summary information is printed out. Calculations are done for two values of J and two values of K, specified at lines 22-23. Line 29 calculates the numerator of the OR expression for the non-centrality parameter . The reader should confirm that line 34-38 implements the formulae for RRRC analysis, line 40-42 implements the formulae for FRRC analysis and line 44-46 implements the formulae for RRFC analysis.

Table 2 summarizes the results of sourcing this file for the Van Dyke and the Franken datasets. The first three columns list the dataset name, the parameters of the sample size estimation method and their values. The body of the table lists the numbers of readers and cases, the critical value , , , and Power: in each case three values are listed, separated by slashes, corresponding to the three generalizations.

This table clearly shows that not only is the effect-size much smaller for the Franken dataset, it is also characterized by larger variability components.

Table : Summary of ORH parameters and intermediate quantities involved in random-reader random-case (RRRC), fixed-reader random-case (FRRC) and random-reader fixed-case (RRFC) power analyses for two datasets. VD = Van Dyke and FR = Franken. *J* / *K* denote numbers of readers and numbers of cases considered. In all cases *ndf* = 1 and  = 0.05. Where three values are listed, separated by slashes, they correspond to RRRC/FRRC/RRFC. Note the much larger non-centrality parameter and smaller variances in ORH model for the VD dataset, leading to higher power values.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | Parameter | Value | Sample size parameters for RRRC/FRRC/RRFC | | | | |
|  |  |  |  | Power |
| VD | -0.0438 (-0.088, 0.000) | | | | | | |
|  | 0.0008 | 5 / 114 | 4.53 / 3.93 / 7.71 | 15.3 / 113 / 4 | 4.46 / 5.48 / 8.70 | 0.507 / 0.641 / 0.606 |
|  | 0.00035 | 5 / 228 | 4.79 / 3.88 / 7.71 | 11.5 / 227 / 4 | 7.51 / 11.0 / 12.8 | 0.708 / 0.909 / 0.761 |
|  | 0.00034 | 10 / 114 | 3.97 / 3.93 / 5.12 | 76.0 / 113 / 9 | 5.99 / 6.85 / 17.4 | 0.676 / 0.737 / 0.959 |
|  | 0.00024 | 10 / 228 | 4.03 / 3.88 / 5.12 | 51.8 / 227 / 9 | 10.6 / 13.7 / 25.5 | 0.893 / 0.958 / 0.994 |
|  | 0.00020 |  | | | | |
| FR | 0.011 (-0.005, 0.027) | | | | | | |
|  | 0.0015 | 4 / 100 | 10.1 / 3.94/ 10.1 | 3 / 99 /3 | 0.321 / 0.321 /0.321 | 0.070 / 0.087 /0.087 |
|  | 0.00079 | 4 / 200 | 10.1 / 3.89 / 10.1 | 3 / 199 / 3 | 0.642 / 0.642 /0.642 | 0.089 / 0.125 / 0.089 |
|  | 0.00048 | 8 / 100 | 5.59 / 3.94 / 5.59 | 7 / 99 / 7 | 0.642 /0.642 / 0.642 | 0.107 / 0.125 / 0.107 |
|  | 0.00051 | 8 / 200 | 5.59 / 3.89 / 5.59 | 7 / 199 / 7 | 1.28 / 1.28 / 1.28 | 0.166 / 0.204 / 0.166 |
|  | 0 |  | | | | |

In the Van Dyke dataset the constraint applied by Hillis to maintain the correct ordering of  and  is redundant, as they already have the correct ordering. In the Franken dataset, these covariances are incorrectly ordered. Therefore, according to the defining equation, *ddf* is *J-1*, see *ddf* column of Table 2. *Positing a conjectured positive value for  would further decrease power for this dataset.* Given the small observed effect-size, a pivotal study is not warranted for the Franken dataset, even though the effect-size is in the "right direction", i.e., digital > plain film (modality 1 is digital). Even using the upper limit of the 95% CI as the anticipated effect-size, one needs 10 readers and 203 cases for RRRC generalization. Any larger anticipated effect-size would be so overly optimistic, given the technology at that time, as to be wishful thinking. A pivotal study was *not* in fact conducted.

It is interesting that in the early days of reporting the results of ROC studies, failure to reject the NH was sometimes taken as evidence that the NH is correct, i.e., the modalities are "equivalent" or "comparable", which is a fallacy. This is a common mistake, even nowadays. The conclusion of the Franken paper5 states: "*The study suggest that for pediatric plain film images, video images offer diagnostic information comparable with that of conventional radiographs for neonatal examinations*".

There is a human tendency to put the best face on new technology. This example also shows that it is more informative to quote a CI in addition to a quoting p-value. The late Prof. Metz has an interesting paper on this6 appropriately titled "*Quantification of Failure to Demonstrate Statistical Significance: The Usefulness of Confidence Intervals*". *A study that fails to reject the NH is nevertheless useful in the archival literature if it reports a CI, because others can use it to select a realistic effect-size.*

# References

1. Hillis SL, Berbaum KS. Power Estimation for the Dorfman-Berbaum-Metz Method. *Acad. Radiol.* 2004;11(11):1260-1273.

2. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Acad Radiol.* 2008;15(5):647-661.

3. Hillis SL, Obuchowski NA, Berbaum KS. Power Estimation for Multireader ROC Methods: An Updated and Unified Approach. *Academic Radiology.* 2011;18(2):129-142.

4. Van Dyke CW, White RD, Obuchowski NA, Geisinger MA, Lorig RJ, Meziane MA. Cine MRI in the diagnosis of thoracic aortic dissection. *79th RSNA Meetings.* 1993.

5. Franken EA, Jr., Berbaum KS, Marley SM, Smith WL, Sato Y, Kao SCS, Milam SG. Evaluation of a Digital Workstation for Interpreting Neonatal Examinations: A Receiver Operating Characteristic Study. *Investigative Radiology.* 1992;27(9):732-737.

6. Metz CE. Quantification of Failure to Demonstrate Statistical Significance: The Usefulness of Confidence Intervals. *Investigative Radiology.* 1993;28:59-63.