Published in final edited form as:

Acad Radiol. 2012 December ; 19(12): 1518–1528. doi:10.1016/j.acra.2012.09.011.

Simulation of Unequal-Variance Binormal Multireader ROC Decision Data: An Extension of the Roe and Metz Simulation Model

Stephen L. Hillis, PhD

Departments of Radiology and Biostatistics, University of Iowa; Comprehensive Access and Delivery Research and Evaluation (CADRE) Center, Iowa City VA Health Care System

1. INTRODUCTION

Roe and Metz (RM) [1] proposed a model for the purpose of simulating multireader (MRMC) data that emulate confidence-of-disease data collected from the typical factorial study design where each case (i.e., patient) undergoes each of several diagnostic tests (or modalities) and the resulting images are interpreted once by each reader. Studies that have used this model include References [2–11]. To account for the MRMC study design, the RM model generates data according to an equal-variance binormal model for which the separation of the diseased and nondiseased decision-variable distributions varies across modality-reader combinations. However, it is generally recognized that for real data the decision variable distributions for diseased and nondiseased cases will often have different variances, with the diseased distribution typically having the larger variance [12, p. 83]. Thus this equal-variance assumption is an important limitation of the RM model. The purpose of this paper is to extend the RM model by allowing the diseased and nondiseased decision-variable distributions to have unequal variances for each reader, while keeping intact other important characteristics of the original RM simulation structures.

2. MATERIALS AND METHODS

2.1. Original RM model formulation

Let X_{ijkt} denote the value of the RM-model decision variable for test i , reader j , case k , and truth t ($t = \text{“-”}$ for a nondiseased case image, “+” for a diseased case image); i.e., X_{ijkt} represents the reader’s degree of confidence that the image is abnormal. The RM decision model, using the notation of Roe and Metz [1], is given by

$$X_{ijkt} = \mu_t + \tau_{it} + R_{jt} + C_{kt} + (\tau R)_{ijt} + (\tau C)_{ikt} + (RC)_{jkt} + (\tau RC)_{ijkt} + E_{ijkt} \quad (1)$$

Without loss of generality they set

$$\mu_- = 0 \quad (2)$$

© 2012 The Association of University Radiologists. Published by Elsevier Inc. All rights reserved.

Address: Department of Radiology, 3170 Medical Laboratories, 200 Hawkins Drive, Iowa City, IA 52242-1077 Telephone: 319-338-0581 x7680 steve-hillis@uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and

$$\tau_{i-}=0 \text{ for all } i \quad (3)$$

All effects are random except for μ_t and τ_{it} . The random effects are mutually independent and normally distributed with zero means. Roe and Metz denote the corresponding variance components by $\sigma_R^2, \sigma_C^2, \sigma_{\tau R}^2, \sigma_{\tau C}^2, \sigma_{RC}^2, \sigma_{\tau RC}^2$, and σ_E^2 . They note that the three-way interaction and error variance components, $\sigma_{\tau RC}^2$ and σ_E^2 , are inseparable (i.e., only the sum can be estimated) because there are no replications, and hence define

$$\sigma_\varepsilon^2 = \sigma_{\tau RC}^2 + \sigma_E^2$$

In their simulations Roe and Metz set $\sigma_{\tau RC}^2 = 0$, which is equivalent to omitting $(\tau RC)_{ijkl}$ from Eq. 1, resulting in σ_ε^2 representing only the pure error in the model.

Because a case is either normal or abnormal, but not both (i.e., case is nested within truth), then any effect involving case must depend on the truth level t , which is the reason that C_{kt} , $(\tau C)_{ikt}$, $(RC)_{jkt}$, and $(\tau RC)_{ijkl}$ each contain a t subscript. For the other three effects, μ_b , τ_{it} , and R_{jt} , the presence of the t subscript in Eq. 1 indicates an interaction with truth. (A more standard statistical notation would include a truth factor symbol in the notation of the effects representing interactions with truth and include the t subscript within parentheses for effects nested within truth.) Note that all of the effects in the RM model have a t subscript, i.e., effects are either interactions with truth or nested within truth. This is justified by the invariance of the ROC curve to location shifts of the decision variable, which implies that the ROC curve for a given reader does not depend on an effect unless the effect is an interaction with truth or is nested within truth.

2.2. Creating the unequal-variance RM model

Statistical properties for a fixed reader (as opposed to a random reader) can be derived from Eq. 1 by treating the reader and test-by-reader effects that do not involve case, R_{jt} and $(\tau R)_{ijt}$, as fixed. In particular, for test i and fixed reader k , Eq. 1 implies an equal variance model with the nondiseased and diseased decision-variable distributions having means $\mu_- + \tau_{i-}$ and $\mu_+ + \tau_{i+}$, respectively, and the same variance

$$\sigma^2 = \sigma_C^2 + \sigma_{\tau C}^2 + \sigma_{RC}^2 + \sigma_\varepsilon^2 \quad (4)$$

To create the unequal-variance RM model, I allow the variance components $\sigma_C^2, \sigma_{\tau C}^2, \sigma_{RC}^2$, and σ_ε^2 to depend on truth. Specifically, using an obvious notation to show the dependence of the variance components on t , I modify the RM model (1) by defining

$$\sigma_{C(+)}^2 = \frac{1}{b^2} \sigma_{C(-)}^2, \sigma_{\tau C(+)}^2 = \frac{1}{b^2} \sigma_{\tau C(-)}^2, \sigma_{RC(+)}^2 = \frac{1}{b^2} \sigma_{RC(-)}^2, \sigma_{\varepsilon(+)}^2 = \frac{1}{b^2} \sigma_{\varepsilon(-)}^2 \quad (5)$$

for some $b > 0$. In Eq. 5 the value of truth state t is in parentheses.

For test i and fixed reader k , Eqs. 1 and 5 imply an unequal-variance binormal model with the nondiseased and diseased decision-variable distributions having respective means $\mu_- + \tau_{i-} = 0$ and $\mu_+ + \tau_{i-}$ and variances

$$\sigma_{-}^2 = \sigma_{C(-)}^2 + \sigma_{\tau C(-)}^2 + \sigma_{RC(-)}^2 + \sigma_{\varepsilon(-)}^2$$

and

$$\sigma_{+}^2 = \sigma_{C(+)}^2 + \sigma_{\tau C(+)}^2 + \sigma_{RC(+)}^2 + \sigma_{\varepsilon(+)}^2 \quad (6)$$

It follows from Eq. 5 that b is the ratio of the standard deviations for the nondiseased and diseased decision-variable distributions:

$$b = \frac{\sigma_{-}}{\sigma_{+}}$$

Thus b is the conventional binormal “slope” parameter. I will refer to the modified Roe and Metz model, defined by Eqs. 1 and 5, as the *unequal-variance RM model*.

2.3. Selecting inputs

In Appendix A I describe in detail steps for selecting input values for the unequal-variance RM model. These steps ensure that for the typical reader the relationship between the binormal means and standard deviations has a mean-to-sigma ratio [13] equal to 4.5. The mean-to-sigma ratio for a particular reader, denoted by r , is defined as the difference of the binormal means divided by the difference of the binormal standard deviations for a particular reader (see Section Appendix A.2 for a discussion of r .) What is important is that $r = 4.5$ describes a typical relationship between fitted binormal model means and standard deviations for real data sets. For example, Green and Swets [14, p 95] note that $r \approx 4$ describes the relationship between the latent means and standard deviations for many studies. Thus I required that the median value of r across readers be equal to 4.5. At the same time, other characteristics of the original RM model are maintained, including (1) AUC for a typical reader; (2) fixed-reader outcome correlations ρ_{BR} and ρ_{WR} , as defined by Roe and Metz, due to reading the same cases; and (3) between-reader variability, as measured by 95% distribution intervals for the latent AUCs (for each modality and reader the latent AUC is the expected AUC, which can be thought of as the AUC estimate if the numbers of normal and abnormal cases were infinitely large.)

One other requirement that I imposed in selecting input values was that no more than 2.5% of the latent (or true) ROC curves be noticeably *improper* [15]; a binormal ROC curve is improper if it crosses the chance line and hence has a “hook” in it [16]. An improper binormal model implies that larger values of the decision variable are *less* indicative of disease over a false-positive-fraction interval, and thus does not correspond to a meaningful decision variable. Hillis and Berbaum [16] classify the improperness of a binormal ROC curve as visibly *noticeable* if $|A| > 2$. Because noticeably improper ROC curves do not correspond to meaningful decision variables, I imposed the requirement that no more than 2.5% of the latent ROC curves have a mean-to-sigma ratio less than 2. See Section Appendix A.2 for details and further discussion of this approach.

The null-simulation input parameter values used by Roe and Metz for the RM model are given in Table 1. For these null simulations they set $\mu_{-} = \tau_{i-} = \tau_{i+} = 0$, $i = 1, 2$, and thus μ_{+} is the median separation between the diseased and nondiseased distributions across readers. In comparison, the proposed null simulation parameter values for the unequal-variance RM model, created using the steps given in Section Appendix A.3.2 are presented in Table 2.

Table 1 has twelve lines instead of eight lines in the original table given by Roe and Metz; the first six lines correspond to the first two lines in the original table. This modification was made to better show the relationship between Table 1 and Table 2.

For Table 2 each of the simulation parameter combinations results in a median mean-to-sigma ratio of $r = 4.5$. A comparison of Tables 1 and 2 shows that parameters that remain the same are AUC, ρ_{WR} , ρ_{BR} and variance components for the normal cases. Also note that in each table that $\sigma_R^2 = \sigma_{rR}^2$. Parameters that differ are μ_+ , b (for the original RM model $b = 1$), variance components for the diseased cases, and reader and test-by-reader variance components. The $r_{.025}$ column contains the 2.5th percentile of the mean-to-sigma ratio distribution across readers; note that, as required, the lowest value ($r_{.025} = 2.03$) is larger than 2. (For $r = 4.25$ the lowest value was less than 2, which is why I chose $r = 4.5$ rather than $r = 4$.) As mentioned above, reader and test-by-reader variance component values were computed in Table 2 to result in 95% distribution intervals for the latent AUCs identical to those for Table 1. In summary, the parameters values in Table 2 preserve the main properties of the original model while resulting in a ratio of unequal fixed-reader nondiseased and diseased distribution variances that has a typical relationship to the mean (i.e., $r = 4.5$) and only a small proportion of noticeably improper latent ROC curves.

2.4. Example illustrating unequal-variance binormal parameters

To illustrate mean-to-sigma ratios I consider an example from Carolyn Van Dyke, MD, that my colleagues and I have analyzed in previous papers. The study [17] compared the relative performance of single spin-echo magnetic resonance imaging (spin echo) to cinematic presentation of MRI (cine) for the detection of thoracic aortic dissection. There were 45 patients with an aortic dissection and 69 patients without a dissection imaged with both spin echo and cine. Five radiologists independently interpreted all of the images using a five-point ordinal scale: 1 = definitely no aortic dissection, 2 = probably no aortic dissection, 3 = unsure about aortic dissection, 4 = probably aortic dissection, and 5 = definitely aortic dissection.

2.5. Simulation study

MRMC data were simulated using the unequal-variance RM model. For each of the twelve parameter combinations in Table 2, 2000 samples were simulated for each of nine reader-case sample size combinations resulting from pairing each of three reader-sample sizes (readers = 3, 5, and 10) with each of three case-sample sizes (25+/25-, 50+/50-, 100+/100-, where "+" indicates a diseased case and "-" indicates a nondiseased case), resulting in a total of $12 \cdot 9 = 108$ different simulation settings. AUC outcomes were analyzed using the multireader method proposed by Dorfman, Berbaum and Metz (DBM) [22] with the modifications described by Hillis et al [9] incorporated. I evaluated performance of the DBM method with respect to the empirical type I error rate for testing the null hypothesis of no modality effect at $\alpha = .05$. AUC was estimated using two methods: likelihood (semi-parametric) estimation assuming a latent binormal model [18–20] and the trapezoidal-rule (nonparametric) [21] method. For nonparametric estimation continuous decision-variable values were used; for semi-parametric estimation decision values were transformed to ordinal discrete ratings, taking integer values from one to five, using the same cutpoints as Dorfman et al. [2]. In addition, trapezoid-rule AUC simulations were repeated with the number of normal cases tripled: 25 abnormals and 75 normals, 50 abnormals and 150 normals, and 100 normals and 300 abnormals.

It has been shown [8] that the method proposed by Obuchowski and Rockette [23], updated by the degrees of freedom estimate proposed by Hillis [8], yields results identical to those of DBM when the jackknife [24–26] is used to estimate the error covariances. Thus our

analysis results can be considered to have been produced by either method. Data simulation and analysis were performed using SAS [27]. The semiparametric AUC was computed in SAS using a dynamic link library (DLL), written in Fortran 90 by Don Dorfman and Kevin Schartz, which was accessed from within the IML procedure in SAS; this DLL can be downloaded from <http://perception.radiology.uiowa.edu>.

3. RESULTS

Table 3 presents the estimated binormal parameters and the corresponding estimated mean-to-sigma ratio for the example data set. The ROC curves are presented in Figure 1. The median mean-to-sigma ratio is 3.42, in line with the observation by Green and Swets [14, p 95] that $r \approx 4$ is typical. From Figure 1 we see that only the ROC curve for reader 5 using cine, with $r = 1.98$, has an obvious hook. This is the only ROC curve with $|r| \geq 2.0$, and thus is the only ROC curve with improperness classified as noticeable by Hillis and Berbaum [16].

For the simulation study, the empirical type I error rates for $\alpha = .05$ are shown in Figure 2 for trapezoid AUC estimation and in Figure 3 for semiparametric AUC estimation. Conclusions are similar for each method. With $n = 2000$ simulated samples, the probability is approximately .95 that the empirical rate will be between .04 and .06 if the true type I error rate is equal to .05. In both figures we see that performance for five or ten readers was excellent for all three sample sizes for the low (AUC = 0.702) and intermediate (AUC = 0.856) ROC curves, with all of the empirical rates within or close to the interval (0.04, 0.06). For the high (AUC = 0.961) ROC curve one rate slightly exceeded 0.06 in Figure 3, with no rates higher than 0.06 in Figure 2, but there were several rates in the interval (0.03, 0.04) for both figures; thus performance was somewhat conservative, but still quite acceptable.

For three readers, at least half of the rates exceeded 0.06 for the low and intermediate ROC curves in both figures. Thus performance is somewhat liberal. In contrast, for the high ROC curve rates are predominately within the (.04, .06) interval or close to it, making performance acceptable. These results suggest that studies be implemented with more than three readers; if only three readers are used, then I suggest that results be considered preliminary or exploratory, unless the AUC outcomes are high. Results for the trapezoid AUC using a 3:1 normal-to-abnormal case ratio were similar

4. DISCUSSION

Although the Roe and Metz model for simulating MRMC data has been a standard tool for evaluating MRMC analysis methods, a major weakness is that it generates data according to an equal-variance binormal model for each reader. In this paper I have shown how to modify the RM model so that it generates data according to an unequal-variance binormal model for each reader. In doing so, I have preserved other important characteristics of the original simulation structures; specifically, the modified model uses the same AUC values, the same fixed-reader correlations, and the same 95% bounds on the latent reader AUCs.

What is different for the unequal-variance RM model is that the mean-to-sigma ratio, which describes the relationship between the means and variances of the diseased and nondiseased decision-variable distributions, is constrained to have a median value (across readers) of $r = 4.5$, which is representative of many data sets. This value also results in less than 2.5% of latent AUCs being noticeably improper.

A simulation study using the proposed unequal-variance RM model showed that for at least five readers and at least 25 normal and 25 abnormal cases the DBM and Obuchowski-Rockette procedures had excellent performance for low and intermediate ROC curves and

acceptable performance for high ROC curves. On the other hand, when the number of readers was three the performance was somewhat liberal for the low and intermediate ROC curves, suggesting that more than three readers should typically be used. Increasing the ratio of normal to abnormal cases did not change these conclusions.

Although I have only presented input values for null simulations, the results can be easily modified for nonnull simulations which are needed for sample-size and power investigations. An easy way to do this is described in Appendix A.3.4.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Carolyn Van Dyke, MD for sharing her data set for the example. We thank the reviewer for helpful suggestions that clarified the presentation. This research was supported by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), grants R01EB000863 and R01EB013667.

Grant support: National Institute of Biomedical Imaging and Bioengineering (NIBIB), grant R01EB000863 and R01EB013667

Appendix A: Derivation of steps for selecting input values for the unequal-variance RM model

Appendix A.1. Notation and results for the unequal-variance RM model

Recall that the unequal-variance RM model is defined by Eqs. 1 and 5. Furthermore, note that the original RM model is defined by the unequal-variance RM model with $b = 1$.

For the unequal-variance RM model, define

$$X_{ij}^{(kt)} = X_{ijk t}$$

For test i and reader j the model can be written in the form

$$X_{ij}^{(kt)} = \mu_t^{(ij)} + \varepsilon_{kt}^{(ij)} \quad (\text{A.1})$$

where

$$\mu_t^{(ij)} = \mu_t + \tau_{it} + R_{jt} + \tau R_{ijt} \quad (\text{A.2})$$

and

$$\varepsilon_{kt}^{(ij)} = C_{kt} + (\tau C)_{ikt} + (RC)_{ikt} + \tau RC_{ijk t} + E_{ijk t} \quad (\text{A.3})$$

Note that the error term $\varepsilon_{kt}^{(ij)}$ is the sum of all effects involving case, i.e., all effects subscripted by case. I will refer to $\mu_t^{(ij)}$ as the *latent mean* for test i and reader j .

Treating reader j as fixed (i.e., conditioning on R_{jt} and $(\tau R)_{ijt}$, the interactions involving reader that do not involve case), it follows that the $\mu_t^{(ij)}$ are fixed effects and the errors are independently distributed, with

$$\varepsilon_{kt}^{(ij)} \sim N(0, \sigma_t^2), t=-, +$$

where

$$\sigma_-^2 = \sigma_{C(-)}^2 + \sigma_{\tau C(-)}^2 + \sigma_{RC(-)}^2 + \sigma_{\varepsilon(-)}^2 \quad (\text{A.4})$$

$$\sigma_+^2 = \frac{\sigma_-^2}{b^2} \quad (\text{A.5})$$

Note also that, treating reader j as fixed, it follows from Eqs. A.1-A.3 that the latent mean $\mu_t^{(ij)}$ is the expected decision value for test i and reader j across cases.

Furthermore, for test i and reader j the separation between the normal and abnormal distribution means is given by the difference of the latent means:

$$\mu_+^{(ij)} - \mu_-^{(ij)} = \mu_+ + \tau_{i+} + R_{j+} - R_{j-} + \tau R_{ij+} - \tau R_{ij-}$$

Note that the above expression incorporates Eqs. 2 and 3 (Treating readers as random (i.e., R_{jt} and τR_{ijt} are random effects)), it follows for test i that the reader-specific separations²⁷³ are normally distributed:

$$\mu_+^{(ij)} - \mu_-^{(ij)} \sim N(\mu_+ + \tau_{i+}, 2\sigma_R^2 + 2\sigma_{\tau R}^2) \quad (\text{A.6})$$

Hence the expected separation across readers is also the median separation:

$$E(\mu_+^{(ij)} - \mu_-^{(ij)}) = \text{med}(\mu_+^{(ij)} - \mu_-^{(ij)}) = \mu_+ + \tau_{i+} \quad (\text{A.7})$$

Recalling that σ_+^2 and σ_-^2 are the variances of the diseased and nondiseased distributions for a fixed reader, respectively, it follows from Eq. A.7 that the AUC for a typical reader (i.e., a reader having the median separation) using test i is given by

$$\text{AUC} = \Phi\left(\frac{\mu_+ + \tau_{i+}}{\sqrt{\sigma_-^2 + \sigma_+^2}}\right) \quad (\text{A.8})$$

Appendix A.2. Avoiding visibly improper ROC curves

In most practical situations a meaningful decision variable should be an increasing function of the likelihood ratio (likelihood of being diseased divided by likelihood of not being diseased) [12]. A decision variable model having this property and its corresponding ROC curve are said to be *proper* [15]; otherwise it is *improper*. An improper model implies that larger values of the decision variable are *less* indicative of disease for some false-positive-fraction interval, and thus does not correspond to a typical decision variable.

A binormal model is improper if the nondiseased and diseased distribution variances differ, resulting in a single false-positive-fraction value where the ROC curve crosses the chance

line [16]. In practice the “improperness” is typically negligible in the sense that it cannot be seen from looking at the ROC curve. However, there are situations when the improperness is apparent, with the ROC curve visibly crossing below the chance line and having an obvious “hook.”

Because noticeably improper ROC curves do not correspond to meaningful decision variables, for simulating data I recommend avoiding the use of unequal-variance model parameter values that produce more than a small proportion of individual-reader latent ROC curves with visibly noticeable hooks. (The *latent* ROC curve is the underlying or true ROC curve rather than the estimated curve.) This can be accomplished by constraining the mean-to-sigma ratio. For a binormal model with diseased and nondiseased decision-variable means and standard deviations given by (μ_2, σ_2) and (μ_1, σ_1) , respectively, the mean-to-sigma ratio r is defined by

$$r = \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1} \quad (\text{A.9})$$

Equivalently,

$$r = \frac{a}{1 - b}$$

where a and b denote the usual parameters of the binormal model: $a = (\mu_2 - \mu_1)/\sigma_2$ and $b = \sigma_1/\sigma_2$. If $\sigma_1 = \sigma_2$ I define $r = \infty$ if $\mu_2 > \mu_1$, $r = -\infty$ if $\mu_2 < \mu_1$, and r to be undefined if $\mu_2 = \mu_1$.

Defining $\Delta m = \mu_2 - \mu_1$ and $\Delta \sigma = \sigma_2 - \sigma_1$, we can more generally express r as the change in means divided by the change in standard deviations for the decision variable, without having to use specific notation for the means and standard deviations:

$$r = \frac{\Delta m}{\Delta \sigma}$$

The mean-to-sigma ratio was first introduced by Swets et al [13], who noticed that it seemed to be approximately constant for a variety of experiments. Some support for this conclusion was provided by later analyses ([28], [29], [14]). For example, Green and Swets [14, p 95] note that $r \approx 4$ describes the relationship between the latent means and 306 standard deviations for many studies. Hillis and Berbaum [16] classify the improperness of a binormal ROC curve as visibly *indiscernible* if $|r| \geq 3$, *noticeable* if $|r| \geq 2$, and *slight* if $2 < |r| < 3$.

It is easy to show that the binormal ROC curve is uniquely defined in terms of r and AUC.

Furthermore, defining $b = \frac{\sigma_1}{\sigma_2}$ (σ_1 and σ_2 are defined generally as above), there is a one-to-one correspondence between (r, AUC) and $(\Delta m, b)$. Assuming $\mu_2 > \mu_1$ and $b \leq 1$, I show in Appendix B that Δm and b , in terms of r and AUC, are given by

$$\Delta m = \frac{-\tilde{b} + \sqrt{\tilde{b}^2 - 4\tilde{a}\tilde{c}}}{2\tilde{a}} \quad (\text{A.10})$$

and

$$b = \frac{r}{\frac{\Delta m}{\sigma_1} + r} \quad (\text{A.11})$$

where

$$c = \Phi^{-1}(\text{AUC}), \tilde{a} = r^2 - c^2, \tilde{b} = -2c^2 r \sigma_1^2, \tilde{c} = -2c^2 r^2 \sigma_1^2 \quad (\text{A.12})$$

In the next section I derive the unequal-variance RM model parameters by first starting with the median AUC value and variance components for the nondiseased decision-variable distribution and then, setting $r = 4.5$, compute the corresponding values of Δm and b that will be used to generate the simulated data. To ensure that the latent ROC curves are rarely noticeably improper, I then verify that the 2.5th percentile of the mean-to-sigma ratio distribution exceeds 2.0, which implies that less than 2.5% of the latent mean-to-sigma ratios will be visually noticeable, using the Hillis and Berbaum [16] classification.

Appendix A.3. Input values for the unequal-variance simulation model

Appendix A.3.1. Preliminaries

The input values used by Roe and Metz null simulations for comparing two tests are given in Table 1. Without loss of generality I assume

$$\tau_{1+} = 0 \quad (\text{A.13})$$

Under the constraints given by Eqs. 2, 3, and A.13 the null hypothesis of equal test AUCs is given by

$$H_0: \tau_{2+} = 0 \quad (\text{A.14})$$

Under the null hypothesis it follows from Eq. A.7 that μ_+ is the separation between the normal and abnormal distributions for a typical reader (i.e., μ_+ is the median separation across the reader population), with the corresponding median AUC given by Eq. A.8 with $\tau_{t+} = 0$ and $\sigma_-^2 = \sigma_+^2$.

The correlation parameters ρ_{WR} and ρ_{BR} in Table 1 describe correlations, treating readers as fixed, between decision values for the same case and are defined by Roe and Metz as follows:

$$\rho_{WR} = \frac{\sigma_c^2 + \sigma_{\tau c}^2 + \sigma_{RC}^2}{\sigma_c^2 + \sigma_{\tau c}^2 + \sigma_{RC}^2 + \sigma_\varepsilon^2} \quad (\text{A.15})$$

and

$$\rho_{BR} = \frac{\sigma_c^2 + \sigma_{\tau c}^2}{\sigma_c^2 + \sigma_{\tau c}^2 + \sigma_{RC}^2 + \sigma_\varepsilon^2} \quad (\text{A.16})$$

The reader is referred to Roe and Metz [1] for details about the correlations. Since these correlations are for the same case, allowing the variance components in Eqs. A.15-A.16 to depend on truth, as specified by Eqs. 5, will not change the values of these two correlations. The “structure” label in Table 1 describes combinations of values of the two correlations with reader and test-by-reader variance component values.

The reader and test-by-reader variance components used by Roe and Metz for “high reader variance” were chosen so that approximately 95% of the latent AUCs would fall within certain limits, where the latent AUC for a reader corresponds to the difference of the latent means. Thus for test i and reader j the latent AUC is given by $\Phi \left[(\mu_+^{(ij)} - \mu_-^{(ij)}) / \sqrt{2\sigma^2} \right]$ for the RM model and by $\Phi \left[(\mu_+^{(ij)} - \mu_-^{(ij)}) / \sqrt{\sigma_-^2 + \sigma_+^2} \right]$ for the unequal-variance RM model. Under the hypothesis of equal tests (thus $\tau_{1+} = \tau_{2+} = 0$) it follows from Eq. A.6 that 95% of the latent reader AUCs, across the reader population, will be within the interval

$$\left\{ \Phi \left[\frac{\mu_+ - 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{2\sigma^2}} \right], \Phi \left[\frac{\mu_+ + 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{2\sigma^2}} \right] \right\} \quad (\text{A.17})$$

where σ^2 is given by Eq. A.4. From Table 1 we see that Roe and Metz set $\sigma^2 = 1$ and use $\sigma_R^2 + \sigma_{\tau R}^2 = 0.022, 0.060$, and 0.112 for $\mu_+ = 0.75, 1.50$, and 2.50 , respectively. It follows from Eq. A.17 that 95% of the latent AUCs will fall between 0.87 and 0.99 for the high ($\mu_+ = 2.50$, median AUC = 0.961) ROC curves, between 0.72 and 0.94 for the intermediate ($\mu_+ = 1.50$, median AUC = 0.856) ROC curves, and between 0.59 and 0.79 for the low ($\mu_+ = 0.75$, median AUC = 0.702) ROC curves. (I note that Roe and Metz report similar but slightly different ranges, but because they do not give the formula they used it is not possible to know if the slight differences are due to miscalculation or a slightly different formula.)

For the unequal-variance RM model the corresponding interval for the latent reader AUCs is

$$\left\{ \Phi \left[\frac{\mu_+ - 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{\sigma_-^2 + \sigma_+^2}} \right], \Phi \left[\frac{\mu_+ + 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{\sigma_-^2 + \sigma_+^2}} \right] \right\}$$

which is equivalent to

$$\left\{ \Phi \left[\frac{\mu_+ - 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{\sigma_-^2 (1+b^{-2})}} \right], \Phi \left[\frac{\mu_+ + 1.96 \sqrt{2(\sigma_R^2 + \sigma_{\tau R}^2)}}{\sqrt{\sigma_-^2 (1+b^{-2})}} \right] \right\} \quad (\text{A.18})$$

Letting “lower bound” denote the lower endpoint of the interval given in Eq A.18, solving for $\sigma_R^2 + \sigma_{\tau R}^2$ yields

$$\sigma_R^2 + \sigma_{\tau R}^2 = .5 \left[\frac{\mu_+ - \Phi^{-1}(\text{lower bound}) \sqrt{\sigma_-^2 (1+b^{-2})}}{(1.96)} \right]^2 \quad (\text{A.19})$$

To keep the unequal-variance RM model consistent with the RM model, $\sigma_R^2 + \sigma_{\tau R}^2$ will be constrained to yield 95% distribution limits that are the same as for corresponding original RM model.

Appendix A.3.2. Algorithm for computing input values

To create corresponding inputs for the unequal-variance RM model I take the following approach:

1. AUC values from Table 1 are retained. Note that these are actually the median AUC values across readers.
2. Values of μ_+ and b are computed that correspond to the AUC values from Table 1 and $r = 4.5$ for a typical reader, as given by Eq. A.9. These computations are performed using Eqs. A.10-A.12, recognizing from Eq. A.9 that $\Delta m = \mu_+$ because these are null simulations.
3. Values of σ_C^2 , $\sigma_{\tau C}^2$, σ_{RC}^2 , and σ_ε^2 are retained, but only for normal cases; i.e., these values now become the values for $\sigma_{C(-)}^2$, $\sigma_{\tau C(-)}^2$, $\sigma_{RC(-)}^2$, and $\sigma_{\varepsilon(-)}^2$. Thus for fixed readers the nondiseased decision-variable distribution variance and correlations (ρ_{WR} , ρ_{BR}) remain unchanged.
4. Values of $\sigma_{C(+)}^2$, $\sigma_{\tau C(+)}^2$, $\sigma_{RC(+)}^2$, and $\sigma_{\varepsilon(+)}^2$ are obtained by multiplying the corresponding values of $\sigma_{C(-)}^2$, $\sigma_{\tau C(-)}^2$, $\sigma_{RC(-)}^2$, and $\sigma_{\varepsilon(-)}^2$ by $1/b^2$. As a result of this step, for fixed readers the diseased decision-variable correlations (ρ_{WR} and ρ_{BR}) remain unchanged but the variance is increased by a factor of $1/b^2$ (i.e., $\sigma_+^2 = \frac{1}{b^2} \sigma_-^2$).
5. Values of σ_R^2 and $\sigma_{\tau R}^2$ for Table 2 are computed as follows. The sum $\sigma_R^2 + \sigma_{\tau R}^2$ is determined such that approximately 95% of the latent AUCs will fall within the original RM model limits corresponding to the same median AUC. This sum is computed using Eq. A.19,

$$\sigma_R^2 + \sigma_{\tau R}^2 = .5 \left[\frac{\mu_+ - \Phi^{-1}(\text{lower limit}) \sqrt{\sigma_-^2 (1+b^{-2})}}{(1.96)} \right]^2$$

where “lower limit” is set equal to the lower interval limit defined by Eq. A.17 for the RM model. Specifically, for “high reader variance” structures (second letter is “H”) lower limit is given by 0.59469, 0.71293, and 0.86689 for median AUC = .702, .856, and .961, respectively, and for “low reader variance” structures (second letter is “L”) lower limit given by 0.62732, 0.80375, and 0.94088 for median AUC = .702, .856, and .961, respectively. Values of μ_+ and b were determined in step 2 and σ_-^2 is given by Eq. A.4; thus $\sigma_-^2 = 1$.

The resulting null simulation parameter values for the unequal-variance RM model are presented in Table 2. SAS statements used to compute Table 2 are available in Appendix C (available online at www.academicradiology.org).

Appendix A.3.3. Distribution intervals for mean-to-sigma ratio

For the unequal-variance RM model, the latent mean-to-sigma ratio for test i and reader j is given by

$$r_{ij} = \frac{\mu_+^{(ij)} - \mu_-^{(ij)}}{\sigma_+ - \sigma_-}$$

where $\mu_+^{(ij)}$ and $\mu_-^{(ij)}$ are defined by Eq. A.2, $\mu_+^{(ij)} - \mu_-^{(ij)}$ is the latent separation between the diseased and nondiseased distributions, and σ_+ and σ_- are defined by Eqs. A.4 and A.5. It follows from Eqs. A.6, A.13, and A.14 that for null simulations

$$r_{ij} \sim N\left(\frac{\mu_+}{\sigma_- (1/b - 1)}, \frac{2\sigma_R^2 + 2\sigma_{\tau R}^2}{\sigma_-^2 (1/b - 1)^2}\right)$$

and hence the 2.5th percentile for the mean-to-sigma ratio distribution is given by 2.5th pct

(mean-to-sigma ratio) = $\frac{\mu_+ - 1.96 \sqrt{2\sigma_R^2 + 2\sigma_{\tau R}^2}}{\sigma_- (1/b - 1)}$ The 2.5th percentiles are given in the last column of Table 2. Note that all of the lower bounds exceed 2.0; thus less than 2.5% of the latent AUCs will be visibly noticeable, as discussed in Section Appendix A.2. Although I originally considered using $r = 4.0$ for the typical reader, I found that this value, as well as $r = 4.25$, resulted in greater than 2.5% of latent ROC curves that would be noticeably improper. If the magnitude of the reader and modality-by-reader variance components had been smaller, I could have used a smaller median r .

Appendix A.3.4. Nonnull simulation values

Roe and Metz only provide null simulation parameter values. Consider the situation where nonnull simulations are required for computing power for an experiment with two tests. Furthermore, suppose that the investigator desires to investigate power for an effect size specified in terms of the AUC difference, denoted by d (e.g., $d = .05$). For this situation an easy and reasonable approach is the following: (1) use the parameter values in Table 2 for test 1; (2) for test 2, add d to the AUC values and use $\mu_+ + \tau_{2+}$ in place of μ_+ , where $\mu_+ + \tau_{2+}$ is the median separation across readers for test 2 that corresponds to $\text{AUC} + d$. Thus $\mu_+ + \tau_{2+}$ is the solution to

$$\text{AUC} + d = \Phi\left(\frac{\mu_+ + \tau_{2+}}{\sqrt{\sigma_-^2 + \sigma_+^2}}\right)$$

where $\sigma_- = \sqrt{\sigma_{C(-)}^2 + \sigma_{\tau C(-)}^2 + \sigma_{RC(-)}^2 + \sigma_{\varepsilon(-)}^2}$ and $\sigma_+ = \sigma_-/b$. For simplicity I suggest not changing any of the variance component parameter values. Although the median mean-to-sigma ratio will be higher than 4.5 for readers under test 2, the increase usually will not be great because effect sizes of interest are typically relatively small.

Appendix B:

Derivation of Eqs. A.10-A.12

In this section I derive Eqs. A.10-A.12. Let (μ_2, σ_2) and (μ_1, σ_1) denote the mean and standard deviation for the diseased and nondiseased decision-variable distributions for a binormal model with AUC denoting its area under the ROC curve. Let r denote the mean-to-sigma ratio, defined by

$$r = \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1} = \frac{\Delta m}{\Delta \sigma} \quad (\text{B.1})$$

and define

$$b = \frac{\sigma_1}{\sigma_2} \quad (\text{B.2})$$

From the relationship

$$\text{AUC} = \Phi \left(\frac{\Delta m}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right) \quad (\text{B.3})$$

it follows that

$$\Delta m = \Phi^{-1}(\text{AUC}) \sqrt{\sigma_1^2 + \sigma_2^2} \quad (\text{B.4})$$

Define

$$c = \Phi^{-1}(\text{AUC}) \quad (\text{B.5})$$

From Eqn. B.1-B.5 we have

$$\Delta m = c \sqrt{\sigma_1^2 + (\sigma_1 + \Delta \sigma)^2} = c \sqrt{\sigma_1^2 + \left(\sigma_1 + \frac{\Delta m}{r} \right)^2}$$

Solving for Δm yields the following quadratic equation in Δm :

$$\tilde{a}(\Delta m)^2 + \tilde{b}(\Delta m) + \tilde{c} = 0$$

where

$$\tilde{a} = r^2 - c^2, \tilde{b} = -2c^2 r \sigma_1^2, \tilde{c} = -2c^2 r^2 \sigma_1^2$$

Assuming $\mu_2 \neq \mu_1$, the solution for Δm in terms r and AUC, using the quadratic formula, is given by

$$\Delta m = \frac{-\tilde{b} \pm \sqrt{\tilde{b}^2 - 4\tilde{a}\tilde{c}}}{2\tilde{a}}$$

I can then solve for b using the relationship

$$b = \frac{r}{\frac{\Delta m}{\sigma_1} + r}$$

which follows from Eqs. B.1-B.2.

References

- [1]. Roe CA, Metz CE. Dorfman-Berbaum-Metz method for statistical analysis of multireader, multimodality receiver operating characteristic data: Validation with computer simulation. *Academic Radiology*. 1997; 4:298–303. doi: 10.1016/S1076-6332(97)80032-3. [PubMed: 9110028]
- [2]. Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA. Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. *Academic Radiology*. 1998; 5:591–602. [PubMed: 9750888]
- [3]. Beiden SV, Wagner RF, Campbell G. Components-of-variance models and multiple-bootstrap experiments: An alternative method for random-effects, receiver operating characteristic analysis. *Academic Radiology*. 2000; 7(5):341–9. [PubMed: 10803614]
- [4]. Beiden SV, Wagner RF, Campbell G, Chan HP. Analysis of uncertainties in estimates of components of variance in multivariate ROC analysis. *Academic Radiology*. 2001; 8(7):616–22. [PubMed: 11450962]
- [5]. Wagner RF, Beiden SV, Metz CE. Continuous versus categorical data for ROC analysis: some quantitative considerations. *Academic Radiology*. 2001; 8(4):328–34. [PubMed: 11293781]
- [6]. Song X, Zhou XH. A marginal model approach for analysis of multi-reader multi-test receiver operating characteristic (ROC) data. *Biostat*. 2005; 6(2):303–12.
- [7]. Hillis SL, Berbaum KS. Monte Carlo validation of the Dorfman-Berbaum-Metz method using normalized pseudovalues and less data-based model simplification. *Academic Radiology*. 2005; 12:1534–1541. DOI:10.1016/j.acra.2005.07.012. [PubMed: 16321742]
- [8]. Hillis SL. A comparison of denominator degrees of freedom methods for multiple observer ROC analysis. *Statistics in Medicine*. 2007; 26:596–619. DOI:10.1002/sim.2532. [PubMed: 16538699]
- [9]. Hillis SL, Berbaum KS, Metz CE. Recent developments in the Dorfman-Berbaum-Metz procedure for multireader ROC study analysis. *Academic Radiology*. 2008; 15:647–661. DOI:10.1016/j.acra.2007.12.015. [PubMed: 18423323]
- [10]. Hillis SL, Obuchowski NA, Berbaum KS. Power estimation for multireader ROC methods an updated and unified approach. *Acad Radiol*. Feb; 2011 18(2):129–42. [PubMed: 21232681]
- [11]. Chakraborty DP. Prediction Accuracy of a Sample-size Estimation Method for ROC Studies. *Academic Radiology*. 2010; 17:628–638. doi: 10.1016/j.acra.2010.01.007. [PubMed: 20380980]
- [12]. Pepe, M. The statistical evaluation of medical tests for classification and prediction. Oxford University Press; New York: 2003.
- [13]. Swets JA, Tanner WP, Birdsall TG. Decision processes in perception. *Psychological Review*. 1961; 68:301–340. [PubMed: 13774292]
- [14]. Green, DM.; Swets, JA. Signal detection theory and psychophysics. Peninsula Publishing; Los Altos: 1988. Green, DM.; Swets, JA. Signal detection theory and psychophysics. Wiley; New York: 1966. Original work:
- [15]. Egan, JP. Signal detection theory and ROC analysis. Academic Press; New York: 1974.
- [16]. Hillis SL, Berbaum KS. Using the mean-to-sigma ratio as a measure of the improperness of binormal ROC curves. *Academic Radiology*. 2011; 18:143–154. [PubMed: 21232682]
- [17]. Van Dyke, CW.; White, RD.; Obuchowski, NA.; Geisinger, MA.; Lorig, RJ.; Mezziane, MA. Cine MRI in the diagnosis of thoracic aortic dissection. 79th RSNA Meetings; Chicago, IL. November 28 - December 3, 1993;
- [18]. Dorfman DD, Alf E Jr. Maximum likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating method data. *Journal of Mathematical Psychology*. 1969; 6:487–496.
- [19]. Dorfman, DD.; RSCORE, II. Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. Swets, JA.; Pickett, RM., editors. Academic Press; San Diego, CA: 1982. p. 212-232.
- [20]. Dorfman DD, Berbaum KS. Degeneracy and discrete receiver operating characteristic rating data. *Academic Radiology*. 1995; 2:907–915. [PubMed: 9419659]
- [21]. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982; 143:29–36. [PubMed: 7063747]

- [22]. Dorfman DD, Berbaum KS, Metz CE. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Investigative Radiology*. 1992; 27:723–731. [PubMed: 1399456]
- [23]. Obuchowski NA, Rockette HE. Hypothesis testing of the diagnostic accuracy for multiple diagnostic tests: an ANOVA approach with dependent observations. *Communications in Statistics: Simulation and Computation*. 1995; 24:285–308.
- [24]. Quenoille MH. Approximate tests of correlation in time series. *Journal of the Royal Statistical Society, series B*. 1949; 11:68–84.
- [25]. Quenoille MH. Notes on bias in estimation. *Biometrika*. 1956; 43:353–360.
- [26]. Tukey JW. Bias and confidence in not quite large samples, (abstract). *Annals of Mathematical Statistics*. 1958; 29:614.
- [27]. SAS for Windows. Version 9.3. Cary, NC, USA: copyright (c) 2002-2010 by SAS Institute Inc.
- [28]. Swets JA. Form of empirical ROCS in discrimination and diagnostic tasks: implications for theory and measurement of performance. *Psychological Bulletin*. 1986; 99:181–198. [PubMed: 3515382]
- [29]. Swets JA. Indices of discrimination or diagnostic accuracy: their ROCs and implied models. *Psycho- logical Bulletin*. 1986; 99:100–117.

Figure 1.
Binormal ROC curves for Van Dyke et al [17] data by reader.

Figure 2.

Plots of trapezoid and semiparametric (latent binormal) AUC empirical type I error rates ($\alpha = .05$) versus correlation-and-variance structure and sample size for testing the null hypothesis of equal modality AUCs. HL = high data correlation, low reader variance; LL = low data correlation, low reader variance; HH = high data correlation, high reader variance; LH = low data correlation, high reader variance. Case sample sizes are indicated by 25+/25-, 50+/50-, 100+/100-, where “+” indicates a diseased case and “-” indicates a nondiseased case.

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

Null simulation parameter values used by Roe and Metz [1]

Structure*	μ	AUC	Correlations			Variance Components						
			ρ_{WR}	ρ_{BR}	σ_C^2	σ_C^2	σ_C^2	σ_{RC}^2	σ_e^2	σ_R^2	σ_R^2	$\sigma_{\pi R}^2$
HL	0.75	0.702	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.0055	0.0055	0.0055
HL	1.50	0.856	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.0055	0.0055	0.0055
HL	2.50	0.961	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.0055	0.0055	0.0055
LL	0.75	0.702	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.0055	0.0055	0.0055
LL	1.50	0.856	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.0055	0.0055	0.0055
LL	2.50	0.961	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.0055	0.0055	0.0055
HH	0.75	0.702	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.011	0.011	0.011
HH	1.50	0.856	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.030	0.030	0.030
HH	2.50	0.961	0.8	0.6	0.3	0.3	0.3	0.2	0.2	0.056	0.056	0.056
LH	0.75	0.702	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.011	0.011	0.011
LH	1.50	0.856	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.030	0.030	0.030
LH	2.50	0.961	0.4	0.2	0.1	0.1	0.1	0.2	0.6	0.056	0.056	0.056

Notes: This table is reprinted and adapted with permission from Roe and Metz [1]. For these null simulations $\mu_- = \tau_{i+} = \tau_{i-} = 0$ for all i . Thus μ_+ is the median separation of the normal and abnormal decision variable distributions across readers. This table contains slight corrections (e.g., AUC = .961 instead of .962) from Reference [1] so that AUC corresponds to μ_+ . In Reference [1], “ A_Z ” is used instead of “AUC.” All other notation is the same as in Reference [1]. Structure: HH = high data correlation, high reader variance; HL = high data correlation, low reader variance; LH = low data correlation, high reader variance; LL = low data correlation, low reader variance.

Table 2

Null simulation parameter values for unequal-variance Roe and Metz model simulations with the median mean-to-sigma ratio set to 4.5, based on values from Table 1.

Structure	AUC	μ_+	b	Correlations		Variance Components										$r_{.025}$
						Normal cases					Abnormal cases					
				ρ_{WR}	ρ_{BR}	$\sigma_{C(-)}^2$	$\sigma_{\pi(-)}^2$	$\sigma_{RC(-)}^2$	$\sigma_{\varepsilon(-)}^2$	$\sigma_{C(+)}^2$	$\sigma_{\pi(+)}^2$	$\sigma_{RC(+)}^2$	$\sigma_{\varepsilon(+)}^2$	σ_R^2	σ_{π}^2	
HL	0.702	0.821	0.846	0.8	0.6	0.3	0.3	0.2	0.2	0.42	0.42	0.28	0.28	0.0066	0.0066	2.76
HL	0.856	1.831	0.711	0.8	0.6	0.3	0.3	0.2	0.2	0.59	0.59	0.40	0.40	0.0082	0.0082	3.63
HL	0.961	3.661	0.551	0.8	0.6	0.3	0.3	0.2	0.2	0.99	0.99	0.66	0.66	0.0118	0.0118	3.98
LL	0.702	0.821	0.846	0.4	0.2	0.1	0.1	0.2	0.6	0.14	0.14	0.28	0.84	0.0066	0.0066	2.76
LL	0.856	1.831	0.711	0.4	0.2	0.1	0.1	0.2	0.6	0.20	0.20	0.40	1.19	0.0082	0.0082	3.63
LL	0.961	3.661	0.551	0.4	0.2	0.1	0.1	0.2	0.6	0.33	0.33	0.66	1.97	0.0118	0.0118	3.98
HH	0.702	0.821	0.846	0.8	0.6	0.3	0.3	0.2	0.2	0.42	0.42	0.28	0.28	0.0132	0.0132	2.03
HH	0.856	1.831	0.711	0.8	0.6	0.3	0.3	0.2	0.2	0.59	0.59	0.40	0.40	0.0447	0.0447	2.46
HH	0.961	3.661	0.551	0.8	0.6	0.3	0.3	0.2	0.2	0.99	0.99	0.66	0.66	0.1201	0.1201	2.83
LH	0.702	0.821	0.846	0.4	0.2	0.1	0.1	0.2	0.6	0.14	0.14	0.28	0.84	0.0132	0.0132	2.03
LH	0.856	1.831	0.711	0.4	0.2	0.1	0.1	0.2	0.6	0.20	0.20	0.40	1.19	0.0447	0.0447	2.46
LH	0.961	3.661	0.551	0.4	0.2	0.1	0.1	0.2	0.6	0.33	0.33	0.66	1.97	0.1201	0.1201	2.83

Notes: For these null simulations $\mu_- = \tau_{j-} = \tau_{j+} = 0$ for all j ; μ_+ is the median separation of the normal and abnormal decision variable distributions across readers; AUC is the median AUC across readers; the median mean-to-sigma ratio across readers for each test is given by $t = \mu_+ / (\sigma_+ - \sigma_-)$, where $\sigma_- = \sqrt{\sigma_{C(-)}^2 + \sigma_{\pi(-)}^2 + \sigma_{RC(-)}^2 + \sigma_{\varepsilon(-)}^2}$ and $\sigma_+ = \sqrt{\sigma_{C(+)}^2 + \sigma_{\pi(+)}^2 + \sigma_{RC(+)}^2 + \sigma_{\varepsilon(+)}^2}$; variance components involving case the abnormal cases were computed by multiplying the corresponding variance components for the normal cases by b^{-2} ; more precise values of b are 0.84566, 0.71082, and 0.55140; structure is defined as in Table 1; $r_{.025}$ is the 2.5th percentile of the distribution of the mean-to-sigma ratio across readers.

Table 3

Latent-binormal model parameter estimates for Van Dyke et al [17] data

Modality	Reader	μ_2	σ_2	r
Cine	1	3.17	1.86	3.67
	2	2.50	1.78	3.19
	3	2.74	1.58	4.76
	4	9.56	4.96	2.41
	5	2.29	2.16	1.98
Spin echo	1	3.68	1.99	3.72
	2	3.70	2.24	2.99
	3	3.32	2.05	3.17
	4	6.93	1.21	33.19
	5	4.11	2.37	3.00

Notes: The mean and standard deviation of the latent nondiseased decision-variable distribution are set to zero and one, respectively; μ_2 and σ_2 are the estimated mean and standard deviation for the latent diseased decision-variable distribution; $r = \mu_2/(\sigma_2 - 1)$ is the estimated mean-to-sigma ratio.