ON THE EFFECT OF IMPROPERNESS OF BINORMAL ROC CURVES FOR ESTIMATING FULL AREA UNDER THE CURVE

by

Ben Guo

B Med, Sichuan University, China, 2010

MPH, University of Pittsburgh, 2012

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH
GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented
by
Ben Guo
It was defended on
Dec. 11th, 2014
and approved by

Thesis Advisor:
Bandos, Andriy, Ph.D.
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Member:
Gur, David, Ph.D.
Professor
Department of Radiology
School of Medicine
University of Pittsburgh

Committee Member:
Jeong, Jong Hyeon, Ph. D.
Professor
Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Copyright © by Ben Guo

2014

Bandos, Andriy, Ph.D.

ON THE EFFECT OF IMPROPERNESS OF BINORMAL ROC CURVES FOR ESTIMATING FULL AREA UNDER THE CURVE

Ben Guo, MS

University of Pittsburgh, 2014

ABSTRACT

The “binormal” model is commonly used for evaluating diagnostic performance with smooth Receiver Operating Characteristic (ROC) curves. However, one of the artifacts of the binormal model is the non-concave (improper) shape of the ROC curves, which is sometimes evident as a visible and practically unreasonable “hook”. The artificial hook can often be triggered, when the true ROC curve is concave but has high initial slope. In these scenarios it is natural to be concerned with the bias in the estimates of global summary measures, e.g., in the commonly used area under the ROC curve (AUC). The objective of this study is to evaluate the magnitude of said bias as a function of improperness of the fitted binormal ROC curves. The public health relevance of this work stems from the importance of the ROC methodology for various stages of development and regulatory approval of medical diagnostic systems. This work investigates whether the AUC for a visually improper binormal ROC curve provides an acceptable estimate of the full area under an actually concave ROC curve. For this purpose a simulation study was conducted based on a wide range of scenarios described by the concave bigamma ROC curves. The binormal ROC curves were fitted using the least squares approach. Based on the “mean-to-sigma ratio” criteria proposed in the literature, the fitted binormal curves were divided into the three groups based on the magnitude of their visual improperness. In order to assess bias in these groups of curves the binormal estimates of AUCs were compared with the empirical AUCs (which are unbiased for continuous data). Our results indicate that for continuous data the bias of the binormal estimate of AUC was small regardless of the magnitude of improperness of the fitted curve. Thus, if one is interested only in estimating AUC using continuous diagnostic data, the improper shape of the binormal curve can often be unimportant. We used data from a multireader study with 36 ROC curves, to illustrate the differences between the bigamma and binormal AUC estimates for different shapes of binormal ROC curves fitted to pseudo-continuous data from actual diagnostic accuracy studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Receiver operating characteristic (ROC) analysis is a method to evaluate the diagnostic performance in general and accuracy of diagnostic imaging system in particular. Two quantities summarized by the ROC curve are the true positive fraction (TPF), or sensitivity, which is the probability of a “positive” test when the signal of interest (any well-defined condition of interest) is actually present (e.g., patient has a disease of interest), and the specificity, which is the probability of a “negative” test result when the signal of interest is actually absent (e.g., the patient does not have a disease of interest). The ROC curve describes TPF as a function of 1-specificity, or false positive fraction (FPF), computed at the same threshold. The threshold, ξ, is a value that is used to dichotomize a multi-category diagnostic marker, T, in the “positive” group indicating presence of the “disease” (e.g., ‘+’=$\{T{:}T> \xi$) and “negative” group indicating absence of the disease. In practice, diagnostic markers are imperfect and there is no threshold value that can lead to a complete separation of “diseased” and “non-diseased” sub-populations, hence pairs of TPF and FPF at various threshold values need to be considered in order to determine the most appropriate threshold for practical application (1).

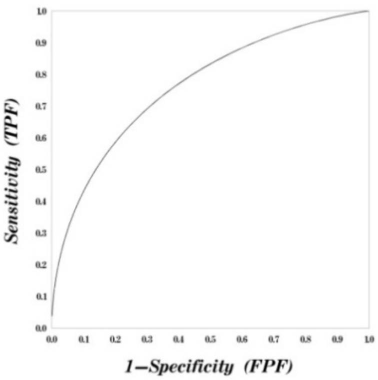

Figure 1.1 ROC Curve

In order to conduct meaningful ROC analyses the "truth", D, has to be established (known) for each subject in terms of actual presence or absence of the signal ("disease") of interest. The "gold standard" for determining the truth is not always known in absolute terms, but this topic is beyond the scope of this work. For the conventional ROC analyses, the truth status can only have two states, for example, "diseased" $(D=1)$ and "non-diseased" $(D=0)$, although generalized forms of ROC analysis have been set up to solve the problem with more than two states (2). Knowing the truth and distribution of diagnostic results the $FPF(\xi)=P(T> \xi|D=0)$ and $TPF(\xi)=P(T>\xi|D=1)$ functions (survival distribution function) can be determined from correspondingly the non-diseased and diseased subpopulations and used to construct the ROC curve, i.e.:

$$ROC(t) = TPF(FPF^{-1}(t))$$

The area under the ROC curve (AUC) is a common summary index for quantifying the overall accuracy of a diagnostic system (marker). It is defined as:

$$A = \int_0^1 ROC(t)dt$$

There are several interpretations of AUC. It can be interpreted as the average value of sensitivity for all possible values of specificity. Also, it is the probability that a randomly selected diseased subject has a test result indicating greater suspicion than a randomly selected non-diseased subject (18). The range of plausible AUC values is typically from 0.5 to 1, with 1 corresponding to perfect accuracy (10). Values of AUC between 0.5-0.7 are generally considered as "low accuracy" level; values between 0.7 and 0.9 as "moderate accuracy" level; and values higher than 0.9 as "high accuracy" level (21).

Empirically the ROC curve can be estimated by computing empirical estimates of *FPF(ξ)* and *TPF(ξ)* for different thresholds and connecting empirical ROC points $(\widehat{fpf}, \widehat{tpf})$ with straight lines. Suppose the sample sizes of non-disease group and disease group are m and n, and X_i and Y_j are the diagnostic results for correspondingly non-diseased and diseased subjects, the empirical sensitivity and 1- specificity can then be computed as:

$$\widehat{FPF}(\xi) = \frac{1}{m}\sum_{i=1}^{m} I(x_i > \xi) \quad \widehat{TPF}(\xi) = \frac{1}{n}\sum_{j=1}^{n} I(y_j > \xi)$$

where for convenience of notations *X* denotes *T|D=0*, and *Y* denotes *T|D=1*. And the empirical estimator of the AUC can be computed by summing the area of trapezoids formed by connecting the points of the empirical ROC curve. The resulting "empirical" estimator is also known to equivalent to the Mann-Whitney statistic (6). This enables formulating the empirical AUC as a U-statistic as follows (22):

$$\hat{A} = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m} \psi(X_i, Y_j) \qquad where\ \psi(X,Y) = \begin{cases} 0 & Y < X \\ \frac{1}{2} & Y = X \\ 1 & Y > X \end{cases}$$

This formulation makes it easy to see that for continuous data the empirical AUC provides an unbiased estimator. Indeed:

$$E(\hat{A}) = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} E(\psi) = P(X < Y) = \int TPF\, dFPF$$

ROC curves can also be estimated using multiple parametric and semiparametric approaches (4). Among parametric approach, the most widely used model is the "binormal" model. Under this model, one assumes that each of the distributions of diagnostic results for non-disease and disease groups follow a normal distribution, or a monotonically increasing transformation thereof (since the ROC curve depends only on ranks of the data). For example, if $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$, then $TPF(\xi) = \Phi\big(a + b\Phi^{-1}(FPF)\big)$ $FPF(\xi) = \Phi\left(\frac{\mu_x - \xi}{\sigma_x}\right)$. As a result the ROC curve could be written as:

$$ROC(fpf) = \Phi(\mathrm{a} + \mathrm{b}\Phi^{-1}(\mathrm{x})), \text{ where } \quad a = \frac{(\mu_y - \mu_x)}{\sigma_y} \quad and \quad b = \frac{\sigma_x}{\sigma_y}$$

Based on the values of 'a' and 'b', the AUC can be written as in a following closed form:

$$AUC = \Phi\left(\frac{a}{\sqrt{1 + b^2}}\right)$$

Since 'b' is always positive, the binormal ROC curve can alternatively be parameterized using parameters AUC and b (which is exploited in some statistical packages, e.g., PASS v.12).

Binormal ROC curves are known to provide a good visual fit to points from other types of ROC curves, as long as points are located in the middle of the range. For example, Hanley et al. generated 5 discrete points with FPF from 0.09 to 0.81 from several types of ROC curves based on various distributions, such as, Binomial, Poisson, Chi-squared distribution, and several others and fitted the corresponding ROC curves using the binormal model. Based on comparison of the original ROC point to the binormal fitted ROC curves using the goodness-of-fit criteria the binormal curves appear adequate for a wide range of ROC shapes (3). However the robustness of binormal ROC curve was also demonstrated to diminish when the discrete operating points are less well distributed (8). For example, for bi-logistic ROC curve, the statistical properties of the binormal estimates deteriorate, and the estimated AUC and its variance increase as the ROC points get less well dispersed (8).

One of the deficiencies of the binormal ROC curves is the presence of the non-concave regions. If parameter b<1 then binormal ROC curve has a non-concave region (“hook”) for high values of FPF (low values of specificity). If b>1 then non-concave region occurs for low values of FPF (Figure 1.2,)

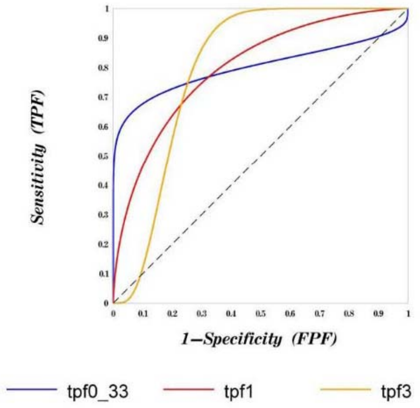

Figure 1.2 Three different shapes of bionormal ROC curves with AUC of 0.8 and b of 0.33, 1, and 3

The binormal ROC curve with non-concave region is often termed “improper”. The incorrectness (improperness) of the hook follows the fact that non-concave regions on the ROC curve correspond to the level of diagnostic performance that is worse than that of chance. In particular, by connecting the (TPF, FPF) point to the right-upper corner with a straight line, one can obtain a “chance line” based on the point from which the curve is extrapolated. This line represents the results of randomly guessing (26). A “hook” on the binormal ROC curve lies below this chance line, and hence describes poorer performance than performance of random guessing in the specific region. Therefore, improper binromal ROC curves are conceptually unreasonable. However, when b is close to 1, the improperness although exist, might not be visible.

Hillis formalized an approach for evaluating the visual "improperness" of binormal ROC curves by using the "mean to sigma ratio" defined as:

$$r = \frac{\mu_y - \mu_x}{\sigma_y - \sigma_x} \; or \; r = \frac{a}{1-b}$$

Using this quantity the binormal ROC curves can be classified into three improperness groups. When the absolute value of the ratio is less than or equal to 2, the curve would exhibit a "noticeable" hook. When the absolute value of the ratio is between 2 and 3, the curve would exhibit a "slight" hook. Otherwise, the curve can be considered "visually proper" (20).

As an alternative to the binormal model, some investigators consider the constant-shape bi-gamma ROC curves which are always concave (9). Under the bigamma model, we assume that diagnostic results (or monotonically increasing transformation thereof) is gamma-distributed in the "diseased" and "non-diseased" subpopulations, i.e., $X \sim GAM(\theta_1, \kappa_1)$ and $Y \sim GAM(\theta_2, \kappa_2)$, then the ROC curve is :

$$ROC(fpf) = S_y\left(S_x^{-1}(fpf)\right)$$

where S denotes the survival function of gamma distribution (the density function of gamma distribution is expressed as $f(x; \kappa, \theta) = \frac{1}{\theta^\kappa}\frac{1}{\Gamma(\kappa)}x^{\kappa-1}e^{-\frac{x}{\theta}}$). The AUC for the bi-gamma ROC curve can be computed as follows (23):

$$AUC = 1 - F_F\left(\left(\frac{\kappa_2}{\kappa_1}\right) * \left(\frac{\theta_1}{\theta_2}\right); 2\kappa_2, 2\kappa_1\right)$$

where $F_F(\,\cdot\,; 2\kappa_2, 2\kappa_1\,)$ is the cumulative distribution function of an F random variable with $2\kappa_2$ numerator and $2\kappa_1$ denominator degrees of freedom. Restriction of κ_1 to be equal to κ_2 leads to concavity of the bigamma ROC curve (9). The corresponding family is called “constant-shape” family bigamma ROC curves, but for brevity I will refer to that as simply “bi-gamma” ROC curves.

Another well-known alternative for the conventional binormal model is provided “Proper binormal ROC model” (PBM model), which always fits a concave ROC curve. The family of the ROC curves provided under this model can be thought of as binormal ROC curve pieces of which were reordered according to their likelihood ratios. For that reason the resulting ROC curves are sometimes called binormal-LR curves. Rather than the two conventional parameters ‘a’ and ‘b’, the binormal-LR curves are defined by new parameters d_a and c which can be computed from the original binormal parameters as follows:

$$d_a \equiv \frac{\sqrt{2}a}{\sqrt{1+b^2}} \quad and \quad c \equiv \frac{b-1}{b+1}$$

When compared with the conventional binormal model, the proper binormal model provides a visually more plausible results if binormal ROC curve has a hook. Otherwise, two binormal models tend to provide very similar results (11, 12). The authors later improved the computational approach and showed that their newer implementation of the PBM was more reliable in a very broad variety of datasets. (19). Recently, the binormal-LR ROC curves were shown to be related to the ROC curves based on the Gamma family of probability distribution with specific parameterization (27).

Another modification of binormal model, the “contaminated binormal model” was introduced in 2000. This model attempted to deal with datasets that results in empirical points with TPF>0 and FPF=0 using a mixture of two normal distributions. The fitted ROC curves provided by this model are also always proper (14, 15).

Thus, many different approaches exist to correct deficiencies of the conventional binormal model. However, none of them approach the level of simplicity of binormal model for planning studies and data analysis. Commercial statistical packages (e.g., PASS v.12) use sample size estimation approaches based on binormal model for planning ROC studies. And unlike other models the binormal ROC curve can be easily fitted to the categorical data using standard statistical software packages (29) and allows for a simple estimation using generalized linear models (4). Therefore, until other approaches are developed to the similar extent, it is crucial to understand the properties of the binormal model, and the effect of its artifacts on statistical inferences.

2.0 PROBLEM STATEMENT

The "binormal" model is frequently used to fit a smooth ROC curve to empirical points. Under this model a "hook" (the non-concavity region) is always present in the fitted ROC curve, but frequently, the hook is too small to be visually seen in the plot, while in other conditions, the hook is quite noticeable. The hook can often be caused by a high slope of empirical curve in the left corner, since for the fixed area under the ROC curve (AUC) the slope at '(0,0)' is directly related to the shape parameter 'b' thereby impacting the degree of improperness. Indeed:

$$FPF(\xi) = \Phi\left(\frac{\mu_x - \xi}{\sigma_x}\right) = \Phi(-\xi) \quad TPF(\xi) = \Phi\left(\frac{\mu_y - \xi}{\sigma_y}\right) = \Phi(a - b\xi)$$

$$slope = \frac{dROC(FPF)}{dFPF} = \frac{\frac{dTPF(\xi)}{d\xi}}{\frac{dFPF(\xi)}{d\xi}} = \frac{-bf(a-b\xi)}{-f(\xi)} = b\frac{f(a-b\xi)}{f(\xi)}$$

$$= b\ exp\left\{-\frac{(a-b\xi)^2}{2} + \frac{\xi^2}{2}\right\} = b\ exp\left\{\frac{\xi^2 - a^2 + 2ab\xi - b^2\xi^2}{2}\right\}$$

$$= b\ exp\left\{\frac{\xi^2 - b^2\xi^2 - a\xi + ab\xi + a\xi + ab\xi - a^2}{2}\right\}$$

$$= b\ exp\left\{\frac{(1-b)(1+b)\xi - a(1-b)\xi + a(1+b)\xi - a^2}{2}\right\}$$

$$= b\ exp\left\{\frac{[(1-b)\xi + a] * [(1+b)\xi - a]}{2}\right\}$$

Based on the formula above, if b is in the range of (0, 1), as ξ sufficiently large, both $[(1-b)\xi+a]$ and$[(1+b)\xi-a]$ will be positive, hence the product will be positive. Thus, with ξ is approaching infinity, the slope will approach infinity. However, if b is larger than 1, term $[(1-b)\xi+a]$ will be negative, and the product of the two terms will be negative for sufficiently large ξ. Thus, in this case, as ξ is approaching infinity, the slope will approach '0'. Moreover it is straightforward to demonstrate that for ROC points with fixed TPF=tpf_0, the slope increases with b approaching 0.

When the original ROC curve is concave, the presence of a hook could generate a significant and systematic discrepancy between fitted and the true curves. In particular, rather than converging toward the true ROC curve with increasing sample size, the fitted binormal ROC curve would approach the "best" binormal approximation to the true curve. For example, figure 2.1 illustrates the "closest" binormal ROC curve for a bi-gamma ROC curve with κ=0.1 and AUC of 0.8. (Black line corresponds to the bigamma ROC, and red line shows the ROC curve based on the binormal model fitting; binormal curve on the left figure is obtained by minimizing the horizontal distance between the curves, and the binormal curve on the right figure is obtained by minimizing the vertical distance.)

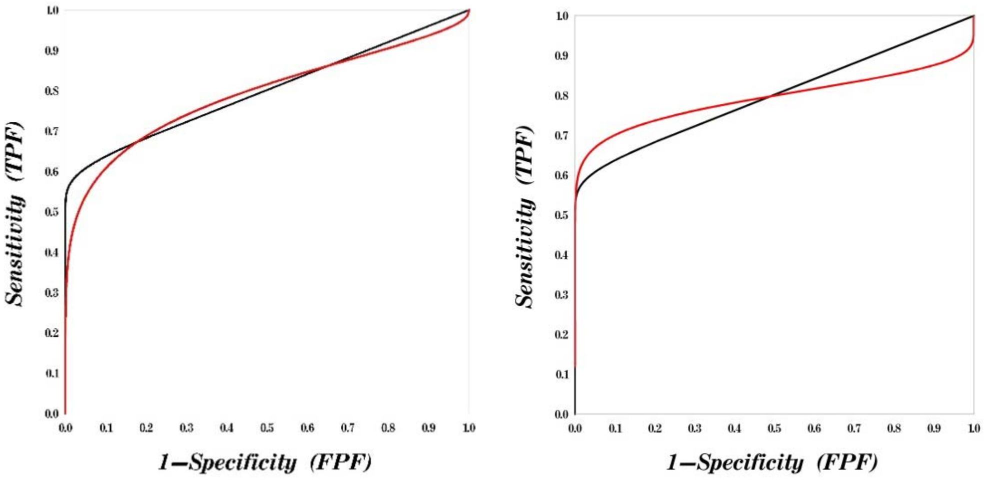

Figure 2.1 True bigamma curve and the fitted binormal curve (optimization of horizontal and vertical distance)

When there is a systematic difference between the ROC curves inferences based on the binormal model, such as the estimated AUCs and its variance, may be severely unreliable. The objective of this study is to evaluate the appropriateness of inferences from the artificially hooked binormal ROC curves in practically reasonable scenarios. In particular, in this study we will focus on bias of the estimated binormal AUC.

Some facts related to our work were previously evaluated by other researchers. In 1995, Hajian-Tilaki et al. (24) discussed the robustness of the conventional binormal model when used to fit the data corresponding to non-binormal ROC curves. In their study, in addition to generating data sets from binormal model (G), data were generated from several mixtures of gaussian (MG) distributions. The authors concluded that for all cases in which the original distributions were {G, MG} pairs, binormal model fitting provided nearly-unbiased estimates of AUCS. For {MG, MG} pairs of distribution, the bias was small for all practical purposes (24). Metz et al (11) in their paper on the “proper” ROC model, compared estimates based on the binormal and proproc ROC curves and commented on similarity among the estimates in several specific scenarios while indicating that sever improperness of the binormal ROC curve might substantially alter the inferences. In contrast to all these studies we focus specifically on scenarios that have high probability of severe improperness in the fitted binormal ROC curve, yet the true binormal ROC is concave.

3.0 METHODS

Scenarios leading to improper shape of the fitted ROC curve were constructed by exploiting the relationship between the slope of the binormal ROC curve at ‘0’ and the degree of the curve improperness. Concave curve with high slope at ‘0’ were modeled using bi-gamma ROC curves, which have higher initial slope at lower values of κ. In 1996, Dorfman proposed the constant-shape bigamma model and in their paper they also generated data sets from the binormal and bigamma models. However, in their study only one specific combination of parameter values were used, specifically parameter κ was set at 4.391, and the scale parameter was set at 0.439. By comparing the estimated AUCs under bigamma and binormal fitting models, they concluded that two models provided similar results for the large sample (9). In order to focus on the improper binormal ROC curves in my study, I consider a substantially more extreme and wider range of parameters with κ ranging from 0.1 to 10. In figure 3.1, the left side shows curves with shape parameter equals 1/3, the right shows curves with shape parameter equals 3, and on both sides different colors indicate the different AUCs (ranging from 0.6 to 0.9). Curves with smaller shape parameter have higher curvature in the range of low FPFs and are nearly straight-line in the range of high FPF. This property cannot be reflected by the binormal model. In this situation, the best fitted conventional binormal curve could have a hook on the right side of the curve (e.g., left side of figure 2.1). On the other hand, when the shape parameter equals 3, the curve is smoother, and has a smaller slope at high specificity regions thereby allowing for a better fitting binormal ROC curve.

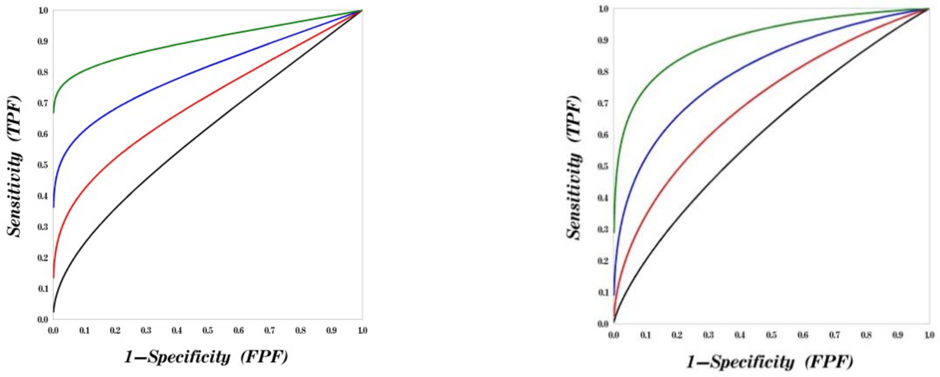

Figure 3.1 Bigamma ROC curves for different values of kappa and AUCs

In my simulation study, 10,000 Monte Carlo samples with sample sizes 20;20, 50:50, and 100:100 were generated from the constant-shape bigamma family of the ROC curve with AUCs of 0.6, 0.7, 0.8 and 0.9 and κ of 1/10, 1/3,3 and 10. For each simulation, we computed all the empirical points and the empirical estimator of AUC. Then, by minimizing the sum of squared horizontal differences between the empirical FPF point and the corresponding estimated FPF point based on the binormal model, we estimated the parameters of the best fitted binormal ROC curve and calculated the corresponding binormal AUCs. In particular, for a given sensitivity level tpf_0 corresponding to the selected empirical point (fpf_0, tpf_0), the binormal FPF was computed as follow:

$$FPF \; = \Phi((\Phi^{-1}(tpf_0) - \hat{a})/\hat{b})$$

where Φ is the cumulative density function of the standard normal distribution (24). The least-square approach was selected due to ability to use it for fitting both binormal and bigamma ROC curves using continuous data. Although offering consistent estimates (7) the least-square approach is not as efficient as the maximum-likelihood approach. However, true maximum likelihood approaches are not well developed for non-binormal ROC curves, and even for binormal ROC curve constitute computation difficulties for continuous data (16). The horizontal differences were used specifically to address the problems of ROC points with fpf=0. The optimization was implemented in SAS using “call nlpnms”, which provides nonlinear optimization by Nelder-Mead simplex method that attempts to minimize a scalar-valued nonlinear function of real variables using only function values.

Knowing the area under the true ROC curve for a given simulation scenario, the bias for the AUC estimators was estimated by comparison with the true AUC. Using the 10,000 independent datasets generated from the same distribution we tested equality of bias to ‘0’, and estimated the equal-tail 95% range of the sampling distribution of individual differences between the true and estimated AUCs.

Furthermore, based on the “mean to sigma ratio” criteria (20), all fitted ROC curves were separated into the following three groups: “proper”, “slightly hooked” and “noticeably hooked”. Within “slightly hooked” and “noticeable hooked” groups, based on the value of b, curves were also separated into two subgroups corresponding to “b<1” and “b>1”. For individual groups, we do not know the true value of the AUC (=P(X<Y|group)), but we know the empirical AUC estimates which are unbiased in presence of continuous data. This enables estimation of bias of binormal AUC estimator by considering the differences between the empirical and binormal estimates, indeed:

$$E\left(\hat{A}_{binormal} - \hat{A}_{empirical}\right) = E\left(\hat{A}_{binormal}\right) - A_{true}$$

Thus, for each of the improperness group the bias of the binormal AUC estimator was estimated as the difference between the empirical AUC and binormal AUC.

We also considered examples based on three experimental data sets. These three data sets came from the study by Herron, et al. 2000 (13). Each dataset contains confidence ratings (0-100 scale) provided 6 readers, for every case under two imaging modalities. The "interstitial" dataset contains confidence ratings regarding the presence of the interstitial disease 84 actually diseased and 223 non-diseased cases. The "nodules" dataset contains confidence ratings regarding the presence of lung nodules for 103 diseased (with verified nodules) and 204 non-diseased cases. The "pneumothorax" dataset contains confidence ratings regarding the presence of pneumothorax for 50 diseased and 200 non-diseased cases. For each individual combination of reader, disease, and modality (36 scenarios in total) we compared the empirical AUCs, the bigamma AUC and binormal AUC.

4.0 RESULTS

Table 4.1 summarizes the types of the binormal ROC curve obtained as a result of fitting simulated data (10,000 resamples), which were generated from various bi-gamma distributions and different sample sizes (20:20, 50:50; 100:100). Different bi-gamma scenarios are parameterized using AUC and shape parameter κ. Based on the parameters of fitted binormal ROC curve results were divided into five categories, namely “visually proper”, “slightly hooked” with b<1, “slightly hooked” with b>1, “noticeably hooked” with b<1 and “noticeably hooked” with b>1.

Table 4.1 Frequency of scenarios with different type and degrees of improperness of the fitted binormal ROC curve

True bigamma AUC	Improperness groups for fitted binormal ROC	κ=1/10			κ=1/3			κ=3			κ=10		
		No. of “diseased”			No. of “diseased”			No. of “diseased”			No. of “diseased”		
		20	50	100	20	50	100	20	50	100	20	50	100
0.6	V P	2560	2304	1669	2861	3110	2892	3217	4448	5638	3286	4596	5900
	S & b<1	837	1122	1359	860	1181	1683	787	1059	1241	692	940	1123
	S & b>1	248	98	19	306	204	46	504	500	384	515	607	524
	N & b<1	5612	6310	6941	4868	5184	5332	3306	2936	2300	3090	2465	1757
	N & b>1	743	166	12	1105	321	47	2186	1057	437	2417	1392	696
0.7	V P	2411	1588	796	3150	2795	1983	4797	6721	7710	5030	7403	8767
	S & b<1	1323	1826	2013	1501	2223	2963	1251	1517	1604	1091	1141	862
	S & b>1	62	1	--	127	5	--	364	210	23	472	360	101
	N & b<1	6086	6585	7191	4969	4976	5054	2612	1459	661	1965	883	260
	N & b>1	118	--	--	253	1	--	976	93	2	1442	213	10
0.8	V P	2177	1649	819	2961	2680	1857	5247	7694	8678	5550	8513	9555
	S & b<1	2201	2714	3239	2285	3141	4265	1625	1563	1202	1262	865	379
	S & b>1	239	2	--	272	10	--	644	99	9	757	219	25
	N & b<1	5320	5635	5942	4350	4169	3878	1486	552	111	1021	175	27
	N & b>1	63	--	--	132	--	--	998	92	--	1410	228	14
0.9	V P	2537	2417	1685	2935	3360	2951	4525	7293	9057	4502	7181	9106
	S & b<1	3445	4155	5409	3288	4222	5387	1836	1284	679	1272	567	129
	S & b>1	769	63	2	1026	158	3	2621	1266	252	3164	2110	762
	N & b<1	3239	3365	2904	2733	2259	1659	592	117	12	247	28	--
	N & b>1	10	--	--	18	1	--	426	40	--	815	114	3

As expected, we observe that the frequency of the improper binormal ROC curve increases for bigamma scenarios with low κ (hence high initial slope). Also due to the relatively high initial slope in all bigamma scenarios the frequency of the binormal ROC curves with b>1 is relatively small. In results presented later in the main manuscript we will focus on the most practically relevant scenarios, namely groups of curves where binormal ROC curves are visually proper (VP) and where they have noticeable improperness in the high-FPF region (N&b<1). Tables with results for all improperness groups are presented in Appendix A.

Table 4.2 illustrates that fitted binormal ROC curves in the “noticeably improper” category have values of ‘b’ parameter as low as 0.1, whereas the “visually proper” ROC curves rarely have b less than 0.4. Figure 4.1, provides an example of the empirical ROC curve and a “noticeably improper” fitted binormal ROC curve for a dataset generated from a bi-gamma scenario with AUC = 0.8 and κ = 0.1. The fitted binormal ROC curve has a high slope in the low FPF range, a noticeable hook in the range of high FPF.

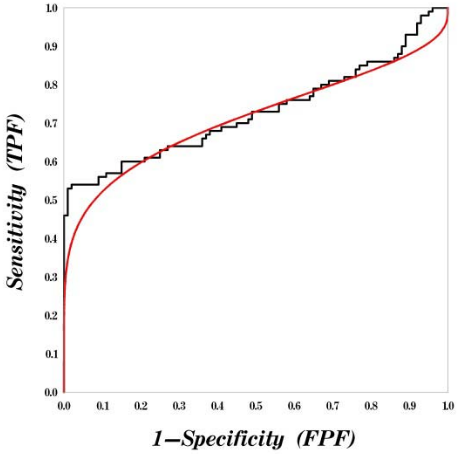

Figure 4.1 Empirical bigamma ROC curve (auc=0.8, κ=0.1) and fitted binormal ROC curve

Table 4.2 The ranges of values of parameter ‘b’ of fitted binormal ROC curves for bi-gamma scenarios and for subsets of generated datasets corresponding to different types and degree of improperness of the fitted binormal ROC curves

Bigamma ROC scenario		**Improperness groups for fitted**	**No. of “non-diseased” and “diseased”**					
κ	**AUC**	**binormal ROC**	**20:20**		**50:50**		**100:100**	
1/10	**0.6**	**VP**	0.61	1.46	0.69	1.24	0.77	1.15
		N & b<1	0.12	1.00	0.31	0.99	0.41	0.98
		All	0.12	3.48	0.31	1.52	0.41	1.30
	0.7	**VP**	0.43	2.35	0.60	1.30	0.65	1.04
		N & b<1	0.12	0.98	0.19	0.87	0.32	0.81
		All	0.12	3.38	0.19	1.33	0.32	1.04
	0.8	**VP**	0.44	2.38	0.51	1.16	0.54	0.99
		N & b<1	0.10	0.87	0.14	0.72	0.19	0.66
		All	0.10	3.07	0.14	2.10	0.19	0.99
	0.9	**VP**	0.43	2.32	0.34	2.21	0.36	1.97
		N & b<1	0.10	0.74	0.11	0.57	0.13	0.50
		All	0.10	2.72	0.11	2.30	0.13	2.16
1/3	**0.6**	**VP**	0.56	1.38	0.71	1.33	0.76	1.23
		N & b<1	0.13	0.99	0.35	1.00	0.44	0.99
		All	0.13	3.72	0.35	1.66	0.44	1.31
	0.7	**VP**	0.51	1.45	0.58	1.43	0.65	1.16
		N & b<1	0.11	0.97	0.21	0.88	0.35	0.82
		All	0.11	3.51	0.21	1.43	0.35	1.16
	0.8	**VP**	0.43	2.33	0.47	1.30	0.53	1.20
		N & b<1	0.11	0.84	0.15	0.74	0.21	0.72
		All	0.11	3.11	0.15	2.40	0.21	1.20
	0.9	**VP**	0.43	2.35	0.35	2.20	0.37	1.87
		N & b<1	0.11	0.70	0.13	0.56	0.16	0.49
		All	0.11	2.83	0.13	2.46	0.16	2.02
3	**0.6**	**VP**	0.62	1.49	0.71	1.34	0.77	1.24
		N & b<1	0.13	1.00	0.39	0.99	0.52	0.99
		All	0.13	3.80	0.39	2.52	0.52	1.54
	0.7	**VP**	0.47	1.45	0.61	1.40	0.66	1.33
		N & b<1	0.17	0.98	0.26	0.96	0.46	0.90
		All	0.17	3.72	0.26	2.72	0.46	1.44
	0.8	**VP**	0.45	2.42	0.51	1.61	0.56	1.49
		N & b<1	0.16	0.92	0.25	0.76	0.39	0.66
		All	0.16	3.79	0.25	2.76	0.39	2.19
	0.9	**VP**	0.43	2.44	0.38	2.30	0.42	2.11
		N & b<1	0.16	0.70	0.14	0.55	0.23	0.43
		All	0.16	3.30	0.14	2.69	0.23	2.39
10	**0.6**	**VP**	0.58	1.49	0.74	1.37	0.75	1.30
		N & b<1	0.15	1.00	0.40	0.99	0.56	0.99
		All	0.15	3.75	0.40	2.01	0.56	1.53
	0.7	**VP**	0.51	1.48	0.63	1.47	0.67	1.41
		N & b<1	0.12	0.99	0.27	0.93	0.50	0.85
		All	0.12	3.93	0.27	2.62	0.50	1.61
	0.8	**VP**	0.43	2.37	0.51	1.82	0.58	1.71
		N & b<1	0.16	0.89	0.30	0.83	0.44	0.67
		All	0.16	4.13	0.30	2.80	0.42	2.58
	0.9	**VP**	0.43	2.47	0.36	2.37	0.38	2.23
		N & b<1	0.18	0.75	0.22	0.48	.	.
		All	0.18	3.53	0.22	2.95	0.28	2.49

Table 4.3 summarizes estimates of the AUC for different bigamma scenarios and subsets of simulated data leading to fitted binormal ROC curves with different degree of improperness.

As expected, the empirical AUC leads to unbiased estimates in all considered scenarios. The binormal AUC exhibits a slight negative bias, which is more pronounced scenarios with low κ (high initial slope of the bi-gamma ROC curve). As the sample size increases, the bias somewhat decreases, but remains noticeable, especially for low value of κ. Interestingly, the magnitude of true bias in the "visually proper" (VP) and "noticeably hooked" (N &b<1) groups is very similar.

Table 4.4 demonstrates that, as expected, the observed bias for the empirical AUC is not statistically different from 0. The bias of the binormal AUC is statistically significantly different from 0 even for large sample size. However, the magnitude of bias is always lower than 0.02, which is less than minimally meaningful difference for many practical applications. Because of the unbiased nature of the empirical AUC, the differences between binormal and empirical AUC estimates provides another approach for inferring about the bias of the binormal AUC estimator. The last two columns of Table 4.4 demonstrate that the corresponding confidence interval is very similar to the confidence interval based on the differences between the true AUC and binormal AUC estimates.

Consideration of the differences between the empirical and binormal AUC estimates was used to evaluate bias of the binormal AUC in the different improperness subgroups. Table 4.5 shows that the negative bias of binormal AUC estimator is statistically significant not only in the subgroup of noticeably improper fitted binormal curves (N&b<1), but also in the subgroup of visually proper (VP) curves. Interestingly, when κ is small, the confidence intervals for the bias in the "N&b<1" subgroup are often tighter than the confidence intervals for VP subgroup.

Table 4.3 Estimated empirical and binormal AUC for different bi-gamma scenarios, sample sizes, and for subsets of generated dataset corresponding to different types and degree of improperness of the fitted binormal ROC curve (results are based on 10,000 simulations per bi-gamma scenario; frequencies of individual subsets are reported in Table 4.1)

Bigamma ROC scenario		Improperness groups for fitted binormal ROC	Number of “non-diseased” and “diseased”								
			20:20			50:50			100:100		
κ	AUC		Emp	Bin	Bias	Emp	Bin	Bias	Emp	Bin	Bias
1/10	0.6	**V P**	0.652	0.645	*-0.007*	0.636	0.630	*-0.006*	0.626	0.620	*-0.006*
		N & b<1	0.574	0.565	*-0.009*	0.583	0.575	*-0.008*	0.589	0.581	*-0.008*
		All	**0.600**	**0.593**	***-0.007***	**0.600**	**0.593**	***-0.007***	**0.600**	**0.593**	***-0.007***
	0.7	**V P**	0.745	0.733	*-0.012*	0.737	0.726	*-0.011*	0.733	0.721	*-0.012*
		N & b<1	0.668	0.654	*-0.014*	0.682	0.670	*-0.012*	0.690	0.678	*-0.012*
		All	**0.698**	**0.685**	***-0.013***	**0.699**	**0.687**	***-0.012***	**0.700**	**0.688**	***-0.012***
	0.8	**V P**	0.834	0.818	*-0.016*	0.835	0.821	*-0.014*	0.831	0.818	*-0.013*
		N & b<1	0.765	0.751	*-0.014*	0.781	0.767	*-0.014*	0.788	0.775	*-0.013*
		All	**0.800**	**0.785**	***-0.015***	**0.800**	**0.787**	***-0.013***	**0.800**	**0.788**	***-0.012***
	0.9	**V P**	0.939	0.917	*-0.022*	0.922	0.908	*-0.014*	0.919	0.908	*-0.011*
		N & b<1	0.850	0.837	*-0.013*	0.875	0.865	*-0.010*	0.882	0.872	*-0.010*
		All	**0.899**	**0.883**	***-0.016***	**0.900**	**0.889**	***-0.011***	**0.900**	**0.890**	***-0.010***
1/3	0.6	**V P**	0.648	0.645	*-0.003*	0.629	0.627	*-0.002*	0.621	0.618	*-0.003*
		N & b<1	0.572	0.567	*-0.005*	0.579	0.576	*-0.003*	0.583	0.580	*-0.003*
		All	**0.601**	**0.597**	***-0.004***	**0.600**	**0.597**	***-0.003***	**0.600**	**0.597**	***-0.003***
	0.7	**V P**	0.738	0.729	*-0.009*	0.727	0.720	*-0.007*	0.722	0.716	*-0.006*
		N & b<1	0.664	0.657	*-0.007*	0.677	0.670	*-0.007*	0.684	0.678	*-0.006*
		All	**0.700**	**0.693**	***-0.007***	**0.700**	**0.693**	***-0.007***	**0.700**	**0.694**	***-0.006***
	0.8	**V P**	0.826	0.813	*-0.013*	0.826	0.816	*-0.010*	0.822	0.812	*-0.010*
		N & b<1	0.758	0.748	*-0.010*	0.775	0.767	*-0.008*	0.781	0.773	*-0.008*
		All	**0.799**	**0.787**	***-0.012***	**0.800**	**0.792**	***-0.008***	**0.800**	**0.792**	***-0.008***
	0.9	**V P**	0.930	0.911	*-0.019*	0.918	0.906	*-0.012*	0.915	0.906	*-0.009*
		N & b<1	0.848	0.838	*-0.010*	0.871	0.863	*-0.008*	0.877	0.870	*-0.007*
		All	**0.899**	**0.884**	***-0.015***	**0.900**	**0.891**	***-0.009***	**0.900**	**0.893**	***-0.007***
3	0.6	**V P**	0.644	0.642	*-0.002*	0.625	0.625	*0.000*	0.613	0.613	*0.000*
		N & b<1	0.570	0.570	*0.000*	0.570	0.569	*-0.001*	0.574	0.573	*-0.001*
		All	**0.601**	**0.600**	***-0.001***	**0.600**	**0.599**	***-0.001***	**0.600**	**0.599**	***-0.001***
	0.7	**V P**	0.723	0.719	*-0.004*	0.709	0.708	*-0.001*	0.705	0.704	*-0.001*
		N & b<1	0.650	0.648	*-0.002*	0.660	0.660	*0.000*	0.667	0.667	*0.000*
		All	**0.700**	**0.697**	***-0.003***	**0.699**	**0.698**	***-0.001***	**0.700**	**0.699**	***-0.001***
	0.8	**V P**	0.799	0.793	*-0.006*	0.804	0.801	*-0.003*	0.802	0.800	*-0.002*
		N & b<1	0.747	0.745	*-0.002*	0.758	0.758	*0.000*	0.767	0.769	*0.002*
		All	**0.800**	**0.793**	***-0.007***	**0.800**	**0.797**	***-0.003***	**0.800**	**0.798**	***-0.002***
	0.9	**V P**	0.899	0.887	*-0.012*	0.899	0.894	*-0.005*	0.900	0.897	*-0.003*
		N & b<1	0.840	0.836	*-0.004*	0.861	0.863	*0.002*	0.882	0.886	*0.004*
		All	**0.899**	**0.885**	***-0.014***	**0.900**	**0.892**	***-0.008***	**0.900**	**0.897**	***-0.003***
10	0.6	**V P**	0.644	0.643	*-0.001*	0.624	0.624	*0.000*	0.611	0.611	*0.000*
		N & b<1	0.565	0.565	*0.000*	0.568	0.567	*-0.001*	0.571	0.571	*0.000*
		All	**0.599**	**0.598**	***-0.001***	**0.599**	**0.599**	***0.000***	**0.599**	**0.599**	***0.000***
	0.7	**V P**	0.719	0.716	*-0.003*	0.707	0.706	*-0.001*	0.703	0.703	*0.000*
		N & b<1	0.647	0.646	*-0.001*	0.657	0.658	*0.001*	0.661	0.661	*0.000*
		All	**0.700**	**0.697**	***-0.003***	**0.699**	**0.698**	***-0.001***	**0.701**	**0.700**	***-0.001***
	0.8	**V P**	0.795	0.790	*-0.005*	0.801	0.799	*-0.002*	0.800	0.799	*-0.001*
		N & b<1	0.744	0.743	*-0.001*	0.750	0.753	*0.003*	0.755	0.759	*0.004*
		All	**0.799**	**0.792**	***-0.007***	**0.800**	**0.797**	***-0.003***	**0.800**	**0.799**	***-0.001***
	0.9	**V P**	0.897	0.887	*-0.010*	0.896	0.892	*-0.004*	0.899	0.896	*-0.003*
		N & b<1	0.838	0.835	*-0.003*	0.869	0.876	*0.007*	--	--	--
		All	**0.900**	**0.886**	***-0.014***	**0.900**	**0.891**	***-0.009***	**0.900**	**0.895**	***-0.005***

Table 4.4 95% Confidence interval for bias of empirical and binormal estimates of the AUC (Wald-type confidence intervals based on 10,000 independent observations)

κ	No. of “diseased”	True bigamma AUC	Empirical - AUC		Binormal - AUC		Binormal-Empricial	
1/10	**20**	0.6	-0.001	0.002	-0.009	-0.005	-0.008	-0.007
		0.7	-0.004	0.000	-0.017	-0.013	-0.013	-0.013
		0.8	-0.002	0.001	-0.017	-0.014	-0.015	-0.015
		0.9	-0.002	0.000	-0.018	-0.016	-0.016	-0.016
	50	0.6	-0.001	0.001	-0.008	-0.006	-0.007	-0.007
		0.7	-0.002	0.000	-0.014	-0.012	-0.012	-0.012
		0.8	-0.001	0.001	-0.014	-0.012	-0.013	-0.013
		0.9	-0.001	0.000	-0.012	-0.010	-0.011	-0.011
	100	0.6	-0.001	0.001	-0.008	-0.007	-0.007	-0.007
		0.7	-0.001	0.001	-0.012	-0.011	-0.012	-0.012
		0.8	0.000	0.001	-0.013	-0.012	-0.013	-0.013
		0.9	-0.001	0.000	-0.010	-0.009	-0.009	-0.009
1/3	**20**	0.6	-0.001	0.002	-0.005	-0.001	-0.004	-0.004
		0.7	-0.001	0.002	-0.009	-0.006	-0.008	-0.007
		0.8	-0.003	0.000	-0.014	-0.011	-0.012	-0.011
		0.9	-0.002	0.000	-0.017	-0.015	-0.016	-0.015
	50	0.6	-0.001	0.001	-0.004	-0.002	-0.003	-0.003
		0.7	-0.001	0.001	-0.008	-0.006	-0.007	-0.007
		0.8	-0.001	0.001	-0.009	-0.007	-0.009	-0.008
		0.9	0.000	0.001	-0.010	-0.008	-0.009	-0.009
	100	0.6	-0.001	0.001	-0.004	-0.002	-0.003	-0.003
		0.7	0.000	0.001	-0.007	-0.005	-0.006	-0.006
		0.8	-0.001	0.000	-0.009	-0.008	-0.008	-0.008
		0.9	0.000	0.001	-0.008	-0.007	-0.008	-0.007
3	**20**	0.6	0.000	0.003	-0.002	0.002	-0.001	-0.001
		0.7	-0.001	0.002	-0.005	-0.002	-0.004	-0.003
		0.8	-0.002	0.001	-0.009	-0.006	-0.007	-0.007
		0.9	-0.002	0.000	-0.016	-0.014	-0.015	-0.014
	50	0.6	-0.001	0.001	-0.002	0.000	-0.001	0.000
		0.7	-0.002	0.000	-0.003	-0.001	-0.001	-0.001
		0.8	-0.001	0.001	-0.004	-0.002	-0.003	-0.002
		0.9	0.000	0.001	-0.008	-0.007	-0.008	-0.008
	100	0.6	-0.001	0.000	-0.002	0.000	0.000	0.000
		0.7	-0.001	0.001	-0.002	0.000	-0.001	-0.001
		0.8	-0.001	0.000	-0.002	-0.001	-0.002	-0.001
		0.9	0.000	0.000	-0.004	-0.003	-0.004	-0.003
10	**20**	0.6	-0.002	0.001	-0.004	0.000	-0.001	-0.001
		0.7	-0.002	0.002	-0.005	-0.002	-0.004	-0.003
		0.8	-0.002	0.000	-0.010	-0.007	-0.007	-0.007
		0.9	-0.001	0.001	-0.015	-0.014	-0.015	-0.015
	50	0.6	-0.002	0.000	-0.002	0.000	0.000	0.000
		0.7	-0.002	0.000	-0.003	-0.001	-0.001	-0.001
		0.8	-0.001	0.001	-0.003	-0.002	-0.003	-0.002
		0.9	-0.001	0.001	-0.010	-0.009	-0.010	-0.009
	100	0.6	-0.002	0.000	-0.002	0.000	0.000	0.000
		0.7	0.000	0.001	-0.001	0.001	-0.001	0.000
		0.8	-0.001	0.000	-0.002	-0.001	-0.001	-0.001
		0.9	0.000	0.001	-0.005	-0.004	-0.005	-0.004

Table 4.5 95% Confidence interval for bias of binormal AUC estimates for subsets of generated dataset corresponding to different types and degree of improperness of the fitted binormal ROC curve (Wald-type confidence intervals based on 10,000 independent observations)

Improperness group of the fitted binormal ROC curve	**Bigamma ROC scenario**		**Number of “non-diseased” and “diseased”**					
	κ	**AUC**	**20:20**		**50:50**		**100:100**	
V P	**1/10**	**0.6**	-0.0072	-0.0065	-0.0065	-0.0060	-0.0063	-0.0060
		0.7	-0.0129	-0.0122	-0.0119	-0.0114	-0.0117	-0.0111
		0.8	-0.0163	-0.0156	-0.0139	-0.0134	-0.0137	-0.0131
		0.9	-0.0218	-0.0207	-0.0142	-0.0135	-0.0113	-0.0109
		All	**-0.0143**	**-0.0139**	**-0.0113**	**-0.0110**	**-0.0100**	**-0.0097**
	1/3	**0.6**	-0.0039	-0.0032	-0.0031	-0.0027	-0.0029	-0.0026
		0.7	-0.0089	-0.0083	-0.0074	-0.0070	-0.0069	-0.0065
		0.8	-0.0129	-0.0123	-0.0102	-0.0097	-0.0097	-0.0093
		0.9	-0.0191	-0.0182	-0.0121	-0.0114	-0.0092	-0.0089
		All	**-0.0111**	**-0.0106**	**-0.0081**	**-0.0078**	**-0.0069**	**-0.0067**
	3	**0.6**	-0.0018	-0.0011	-0.0006	-0.0003	-0.0004	-0.0002
		0.7	-0.0041	-0.0035	-0.0019	-0.0016	-0.0012	-0.0010
		0.8	-0.0066	-0.0060	-0.0032	-0.0029	-0.0021	-0.0018
		0.9	-0.0121	-0.0114	-0.0057	-0.0052	-0.0033	-0.0030
		All	**-0.0063**	**-0.0060**	**-0.0030**	**-0.0028**	**-0.0018**	**-0.0017**
	10	**0.6**	-0.0014	-0.0007	-0.0002	0.0002	-0.0001	0.0001
		0.7	-0.0037	-0.0031	-0.0014	-0.0010	-0.0008	-0.0006
		0.8	-0.0057	-0.0051	-0.0023	-0.0019	-0.0012	-0.0009
		0.9	-0.0106	-0.0099	-0.0042	-0.0037	-0.0027	-0.0023
		All	**-0.0054**	**-0.0051**	**-0.0021**	**-0.0019**	**-0.0012**	**-0.0011**
N & b<1	**1/10**	**0.6**	-0.0087	-0.0080	-0.0082	-0.0078	-0.0078	-0.0076
		0.7	-0.0138	-0.0131	-0.0128	-0.0125	-0.0122	-0.0120
		0.8	-0.0146	-0.0140	-0.0132	-0.0129	-0.0130	-0.0127
		0.9	-0.0126	-0.0120	-0.0102	-0.0098	-0.0096	-0.0093
		All	**-0.0123**	**-0.0119**	**-0.0111**	**-0.0109**	**-0.0107**	**-0.0106**
	1/3	**0.6**	-0.0048	-0.0039	-0.0038	-0.0033	-0.0036	-0.0033
		0.7	-0.0077	-0.0069	-0.0070	-0.0065	-0.0066	-0.0063
		0.8	-0.0103	-0.0096	-0.0083	-0.0078	-0.0082	-0.0078
		0.9	-0.0109	-0.0102	-0.0080	-0.0075	-0.0071	-0.0067
		All	**-0.0079**	**-0.0075**	**-0.0063**	**-0.0061**	**-0.0059**	**-0.0058**
	3	**0.6**	-0.0010	0.0000	-0.0011	-0.0004	-0.0008	-0.0003
		0.7	-0.0029	-0.0016	-0.0004	0.0006	-0.0005	0.0006
		0.8	-0.0027	-0.0010	-0.0005	0.0013	0.0003	0.0030
		0.9	-0.0052	-0.0031	0.0005	0.0043	-0.0002	0.0089
		All	**-0.0019**	**-0.0012**	**-0.0006**	**-0.0001**	**-0.0005**	**-0.0001**
	10	**0.6**	-0.0010	0.0001	-0.0008	-0.0001	-0.0007	-0.0002
		0.7	-0.0022	-0.0008	0.0006	0.0019	-0.0008	0.0008
		0.8	-0.0027	-0.0004	0.0005	0.0040	-0.0001	0.0073
		0.9	-0.0049	-0.0011	0.0022	0.0108	--	--
		All	**-0.0015**	**-0.0007**	**-0.0001**	**0.0005**	**-0.0006**	**-0.0001**

Tables 4.6 and 4.7 summarize the differences that could be observed between true and estimated AUCs and between empirical and binormal AUC estimates for individual datasets generated from bigamma ROC scenarios. The summaries are based on the equal-tail 95% range of the observed differences, which represents the values that are plausible (non-unlikely) to observe in individual experiments.

Table 4.6 summarizes differences between the empirical and true, binormal and true, and binormal and empirical estimates of the AUC. As evident from the results, with either empirical or binormal approach it is very possible to obtain the AUC estimates that are substantially smaller and substantially larger than the true AUC (+/-0.15). Naturally, the absolute differences decrease with increasing sample size, but remain greater than 0.05 even for sample as high as 100:100. However, the differences between the empirical and binormal estimates are generally very close, with the plausible absolute differences within 0.05 overall, and for large samples are usually within 0.02.

Table 4.7 summarizes differences between the empirical and binormal estimates of AUC in different improperness subgroups (datasets leading to difference degree of improperness of the fitted ROC curve). The results demonstrate that the differences between the binormal and empirical AUC are within 0.05 from each other for both visually proper (VP) and noticeably improper (N&b<1) fitted binormal ROC curves. The ranges of plausible differences are similar for the “VP” and “N&b<1” groups, with values for the latter subgroup being slightly larger but within 0.01 for sample size 100:100.

Table 4.6 95% equal-tail range of the sampling distribution of differences in AUC values (based on 10,000 resamples for each bigamma scenario)

No. of “non-diseased” & “disesed”	Bigamma ROC scenario							
	κ	**AUC**	**Empirical-True**		**Binormal-True**		**Binormal-Empirical**	
20:20	**1/10**	**0.6**	-0.1825	0.1725	-0.1937	0.1683	-0.0330	0.0167
		0.7	-0.1813	0.1525	-0.1994	0.1422	-0.0378	0.0099
		0.8	-0.1550	0.1275	-0.1763	0.1112	-0.0383	0.0046
		0.9	-0.1250	0.0925	-0.1440	0.0834	-0.0495	-0.0007
	1/3	**0.6**	-0.1850	0.1750	-0.1929	0.1735	-0.0285	0.0231
		0.7	-0.1725	0.1575	-0.1836	0.1509	-0.0304	0.0193
		0.8	-0.1550	0.1275	-0.1692	0.1118	-0.0345	0.0123
		0.9	-0.1200	0.0900	-0.1361	0.0783	-0.0477	0.0027
	3	**0.6**	-0.1800	0.1700	-0.1819	0.1708	-0.0236	0.0248
		0.7	-0.1700	0.1550	-0.1762	0.1488	-0.0270	0.0251
		0.8	-0.1475	0.1250	-0.1546	0.1096	-0.0323	0.0241
		0.9	-0.1100	0.0825	-0.1178	0.0687	-0.0444	0.0138
	10	**0.6**	-0.1800	0.1725	-0.1827	0.1725	-0.0229	0.0255
		0.7	-0.1700	0.1550	-0.1737	0.1502	-0.0257	0.0257
		0.8	-0.1538	0.1250	-0.1614	0.1096	-0.0316	0.0244
		0.9	-0.1100	0.0800	-0.1205	0.0686	-0.0429	0.0157
50:50	**1/10**	**0.6**	-0.1136	0.1086	-0.1251	0.1049	-0.0220	0.0068
		0.7	-0.1064	0.0986	-0.1228	0.0907	-0.0263	0.0019
		0.8	-0.0952	0.0836	-0.1127	0.0749	-0.0260	0.0002
		0.9	-0.0728	0.0596	-0.0873	0.0496	-0.0297	-0.0003
	1/3	**0.6**	-0.1148	0.1088	-0.1199	0.1088	-0.0173	0.0124
		0.7	-0.1060	0.0996	-0.1165	0.0958	-0.0205	0.0095
		0.8	-0.0940	0.0828	-0.1084	0.0792	-0.0218	0.0079
		0.9	-0.0710	0.0592	-0.0830	0.0497	-0.0393	0.0037
	3	**0.6**	-0.1144	0.1098	-0.1169	0.1124	-0.0133	0.0142
		0.7	-0.1060	0.0956	-0.1101	0.0981	-0.0148	0.0163
		0.8	-0.0916	0.0808	-0.0959	0.0789	-0.0189	0.0178
		0.9	-0.0658	0.0544	-0.0731	0.0459	-0.0404	0.0159
	10	**0.6**	-0.1148	0.1076	-0.1172	0.1096	-0.0127	0.0147
		0.7	-0.1064	0.0964	-0.1095	0.0980	-0.0141	0.0167
		0.8	-0.0896	0.0812	-0.0934	0.0796	-0.0210	0.0181
		0.9	-0.0648	0.0524	-0.0720	0.0453	-0.0395	0.0186
100:100	**1/10**	**0.6**	-0.0803	0.0778	-0.0894	0.0733	-0.0171	0.0024
		0.7	-0.0751	0.0709	-0.0901	0.0628	-0.0218	-0.0019
		0.8	-0.0661	0.0618	-0.0823	0.0523	-0.0216	-0.0036
		0.9	-0.0512	0.0447	-0.0637	0.0384	-0.0172	-0.0017
	1/3	**0.6**	-0.0795	0.0798	-0.0849	0.0786	-0.0126	0.0072
		0.7	-0.0745	0.0717	-0.0842	0.0677	-0.0160	0.0046
		0.8	-0.0658	0.0604	-0.0775	0.0550	-0.0179	0.0030
		0.9	-0.0502	0.0434	-0.0606	0.0390	-0.0159	0.0019
	3	**0.6**	-0.0797	0.0773	-0.0818	0.0782	-0.0092	0.0099
		0.7	-0.0738	0.0696	-0.0765	0.0712	-0.0104	0.0111
		0.8	-0.0626	0.0576	-0.0665	0.0601	-0.0122	0.0121
		0.9	-0.0462	0.0398	-0.0516	0.0392	-0.0333	0.0128
	10	**0.6**	-0.0796	0.0772	-0.0811	0.0791	-0.0088	0.0097
		0.7	-0.0726	0.0720	-0.0751	0.0734	-0.0097	0.0110
		0.8	-0.0643	0.0575	-0.0667	0.0588	-0.0118	0.0137
		0.9	-0.0450	0.0388	-0.0496	0.0384	-0.0358	0.0149

Table 4.7 95% equal-tail range for sampling distribution of individual differences between binormal and empirical estimates of AUCs for subsets of generated datasets corresponding to different types and degree of improperness of the fitted binormal ROC curves

Improperness group of fitted binormal ROC curve	**Bigamma ROC scenario**		**Number of "non-diseased" and "diseased"**					
	κ	**AUC**	**20:20**		**50:50**		**100:100**	
V P	**1/10**	**0.6**	-0.0246	0.0122	-0.0163	0.0045	-0.0133	0.0017
		0.7	-0.0279	0.0041	-0.0222	-0.0003	-0.0189	-0.0039
		0.8	-0.0328	-0.0017	-0.0237	-0.0018	-0.0201	-0.0049
		0.9	-0.0532	-0.0057	-0.0460	-0.0021	-0.0184	-0.0036
		All	***-0.0450***	***0.0057***	***-0.0257***	***0.0018***	***-0.0186***	***-0.0002***
	1//3	**0.6**	-0.0211	0.0167	-0.0142	0.0096	-0.0103	0.0060
		0.7	-0.0257	0.0115	-0.0182	0.0053	-0.0151	0.0033
		0.8	-0.0295	0.0058	-0.0210	0.0038	-0.0178	0.0010
		0.9	-0.0514	-0.0029	-0.0441	0.0013	-0.0168	0.0001
		All	***-0.0397***	***0.0111***	***-0.0220***	***0.0062***	***-0.0163***	***0.0039***
	3	**0.6**	-0.0204	0.0206	-0.0119	0.0123	-0.0083	0.0089
		0.7	-0.0226	0.0193	-0.0141	0.0132	-0.0101	0.0099
		0.8	-0.0255	0.0186	-0.0165	0.0151	-0.0123	0.0109
		0.9	-0.0450	0.0100	-0.0373	0.0142	-0.0165	0.0116
		All	***-0.0284***	***0.0182***	***-0.0180***	***0.0139***	***-0.0128***	***0.0105***
	10	**0.6**	-0.0197	0.0222	-0.0114	0.0137	-0.0081	0.0093
		0.7	-0.0230	0.0210	-0.0136	0.0146	-0.0096	0.0103
		0.8	-0.0254	0.0207	-0.0157	0.0166	-0.0116	0.0131
		0.9	-0.0402	0.0149	-0.0363	0.0186	-0.0304	0.0148
		All	***-0.0273***	***0.0203***	***-0.0174***	***0.0162***	***-0.0131***	***0.0124***
	All		**-0.0336**	**0.0164**	**-0.0203**	**0.0137**	**-0.0152**	**0.0107**
N & b<1	**1/10**	**0.6**	-0.0380	0.0186	-0.0237	0.0076	-0.0179	0.0027
		0.7	-0.0414	0.0114	-0.0279	0.0024	-0.0223	-0.0016
		0.8	-0.0387	0.0061	-0.0279	0.0006	-0.0223	-0.0035
		0.9	-0.0318	0.0019	-0.0221	-0.0003	-0.0176	-0.0018
		All	***-0.0385***	***0.0123***	***-0.0263***	***0.0040***	***-0.0213***	***0.0001***
	1/3	**0.6**	-0.0331	0.0275	-0.0190	0.0144	-0.0139	0.0077
		0.7	-0.0329	0.0233	-0.0221	0.0111	-0.0171	0.0053
		0.8	-0.0347	0.0154	-0.0232	0.0089	-0.0189	0.0037
		0.9	-0.0308	0.0055	-0.0206	0.0050	-0.0158	0.0025
		All	***-0.0330***	***0.0212***	***-0.0215***	***0.0113***	***-0.0169***	***0.0059***
	3	**0.6**	-0.0310	0.0332	-0.0169	0.0181	-0.0110	0.0116
		0.7	-0.0322	0.0330	-0.0177	0.0222	-0.0122	0.0135
		0.8	-0.0305	0.0376	-0.0193	0.0256	-0.0139	0.0165
		0.9	-0.0258	0.0258	-0.0156	0.0261	-0.0045	0.0200
		All	***-0.0312***	***0.0333***	***-0.0172***	***0.0205***	***-0.0111***	***0.0122***
	10	**0.6**	-0.0294	0.0333	-0.0160	0.0171	-0.0107	0.0115
		0.7	-0.0320	0.0372	-0.0163	0.0247	-0.0120	0.0181
		0.8	-0.0334	0.0368	-0.0177	0.0271	-0.0168	0.0191
		0.9	-0.0298	0.0312	-0.0102	0.0402	--	--
		All	***-0.0308***	***0.0352***	***-0.0161***	***0.0200***	***-0.0107***	***0.0122***
	All		**-0.0352**	**0.0241**	**-0.0240**	**0.0124**	**-0.0199**	**0.0062**
All			**-0.0335**	**0.0219**	**-0.0245**	**0.0120**	**-0.0198**	**0.0061**

Figure 4.2 summarizes the differences between the binormal and empirical AUCs against the log of mean-to-sigma ratio for all bigamma ROC scenarios. All generated datasets were grouped into “visually proper”, “slightly hooked” and “noticeable hooked” categories regardless of the location of hook on the fitted binornal ROC curve (the value of ‘b’). In agreement with Table 4.6, in most cases, the two estimates were within 0.05 from each other. In the “visually proper” group, as the mean-to-sigma ratio increases (curves become more proper), the range of differences approaches ‘0’. In the “slightly hooked” group, the range of differences stays approximately the same regardless of the value of log of mean-to-sigma ratio. And in the “noticeable hooked” group, as the mean-to-sigma ratio decreases (curve become more improper), the range of difference between two estimates approaches ‘0’.

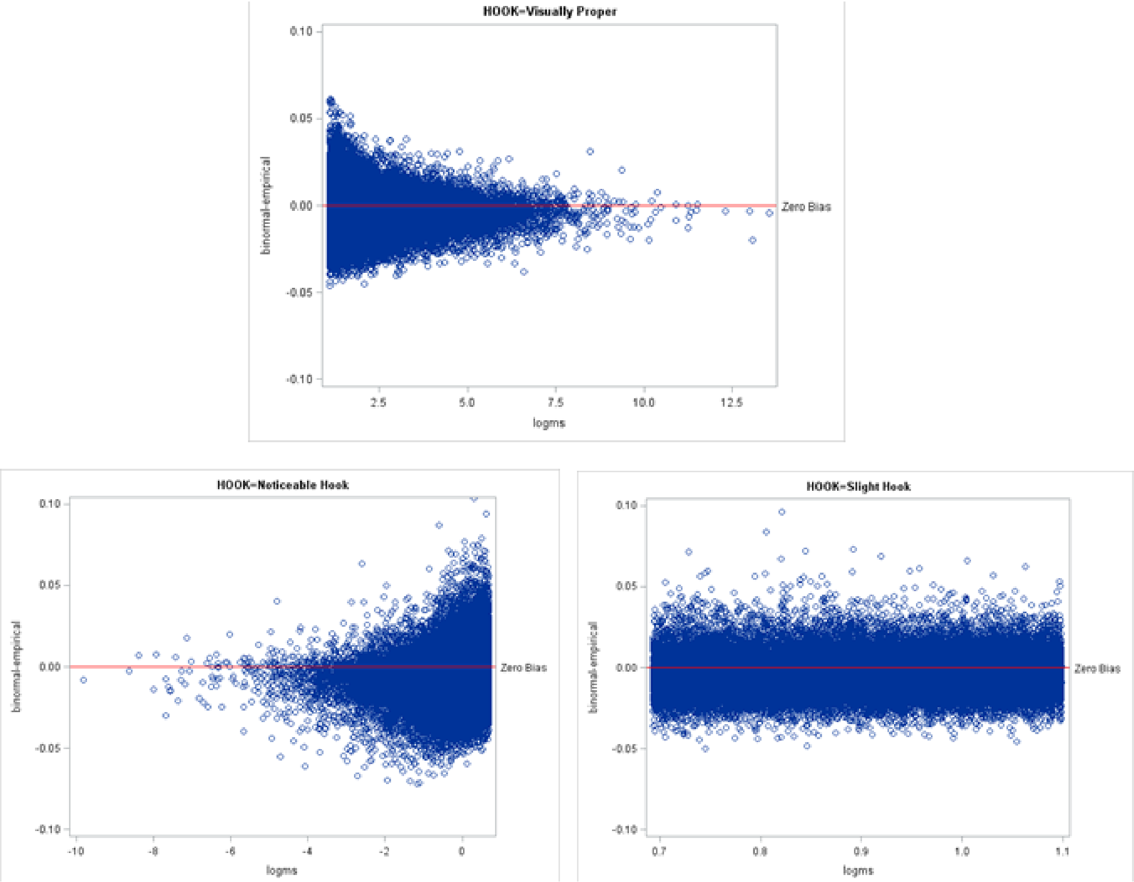

Figure 4.2 Difference between estimated binormal AUCs and estimated empirical AUCs for different values of “mean-to-sigma ratio”

Table 4.8 shows the results of the simulations of partially grouped data (10,000 datasets of size 100:100 for each scenario). The data were initially generated from the bigamma ROC scenarios, followed by grouping ratings that were below the 90th percentile of the true distribution for "non-diseased" subjects'. Such grouping effectively forces empirical ROC points to reside in the range of fpf-values from 0 to 0.1. Unlike the case of completely continuous data, results for partial grouped data indicate that large biases could result from using either empirical or binormal estimates even for large sample sizes. For bigamma scenarios with low κ (high initial slope and nearly-flat shape in the high-FPF range), the empirical AUC tends to have smaller bias than the binormal AUC estimate. However, for bigamma scenarios with larger κ (higher curvature in the high-FPF range), the empirical AUC tends to have a much larger bias than the binormal AUC estimate.

The difference between the empirical and binormal AUC estimates is on average rather substantial both overall and in the different improperness subgroups. In the subgroup of visually proper fitted binormal ROC curve (VP) the empirical AUC is on average substantially larger than the binormal AUC estimate (in many scenarios as high as 0.1 on average). Whereas in the subgroup of visually improper fitted binormal ROC curves the empirical AUC is noticeably smaller.

Table 4.8 Estimated empirical AUC and estimated AUC under binormal model with different value of kappa based on the partially grouped data (Sample Size = 100:100, grouping below 90% percentile of ratings for non-disease cases, 10,000 simulations)

True bigamma scenario		**Improperness**	**Emp**	**Bin**				
κ	**AUC**	**group**	**AUC**	**AUC**	**Emp-AUC**	**Bin-AUC**	**Bin-Emp**	**N**
1/10	**0.6**	**V P**	0.610	0.721			0.111	772
		N & b<1	0.592	0.486			-0.106	8935
		All	0.594	0.510	-0.006	-0.090	-0.084	10000
	0.7	**V P**	0.715	0.815			0.100	67
		N & b<1	0.695	0.595			-0.100	9823
		All	0.695	0.599	-0.005	-0.101	-0.096	10000
	0.8	**V P**	0.843	0.903			0.060	21
		N & b<1	0.796	0.726			-0.070	9874
		All	0.797	0.728	-0.003	-0.072	-0.069	10000
	0.9	**V P**	0.934	0.956			0.022	70
		N & b<1	0.895	0.871			-0.024	8932
		All	0.898	0.878	-0.002	-0.022	-0.020	10000
1/3	**0.6**	**V P**	0.584	0.695			0.111	3337
		N & b<1	0.571	0.498			-0.073	6061
		All	0.576	0.572	-0.024	-0.028	-0.004	10000
	0.7	**V P**	0.685	0.788			0.103	1843
		N & b<1	0.668	0.620			-0.048	7251
		All	0.672	0.661	-0.028	-0.039	-0.011	10000
	0.8	**V P**	0.795	0.867			0.072	915
		N & b<1	0.774	0.739			-0.035	7883
		All	0.778	0.761	-0.022	-0.039	-0.017	10000
	0.9	**V P**	0.909	0.940			0.031	902
		N & b<1	0.882	0.862			-0.020	6668
		All	0.778	0.761	-0.022	-0.039	-0.017	10000
3	**0.6**	**V P**	0.561	0.674			0.113	4151
		N & b<1	0.546	0.472			-0.074	4155
		All	0.551	0.583	-0.049	-0.017	0.032	10000
	0.7	**V P**	0.630	0.746			0.116	5514
		N & b<1	0.617	0.582			-0.035	3711
		All	0.625	0.680	-0.075	-0.020	0.055	10000
	0.8	**V P**	0.728	0.824			0.096	6076
		N & b<1	0.715	0.700			-0.015	2656
		All	0.724	0.784	-0.076	-0.016	0.060	10000
	0.9	**V P**	0.852	0.907			0.055	7016
		N & b<1	0.835	0.825			-0.010	1288
		All	0.849	0.890	-0.051	-0.010	0.041	10000
10	**0.6**	**V P**	0.557	0.669			0.112	4055
		N & b<1	0.541	0.463			-0.078	3946
		All	0.546	0.580	-0.054	-0.020	0.034	10000
	0.7	**V P**	0.618	0.736			0.118	6136
		N & b<1	0.607	0.571			-0.036	3138
		All	0.614	0.679	-0.086	-0.021	0.065	10000
	0.8	**V P**	0.712	0.813			0.101	7016
		N & b<1	0.698	0.686			-0.012	1978
		All	0.708	0.782	-0.092	-0.018	0.073	10000
	0.9	**V P**	0.837	0.900			0.063	8201
		N & b<1	0.823	0.813			-0.010	723
		All	0.835	0.889	-0.065	-0.011	0.053	10000

Table 4.9 and 4.10 summarize the estimates for 36 ROC datasets sampled from study by Herron et al. (2000). Table 4.9 shows the results for datasets resulting in “visually proper” fitted binormal ROC curves. The largest absolute differences between the empirical AUC and two parametric estimates of AUC under the bigamma and binormal models are 0.073 and 0.078. At the same time, the largest absolute difference between the bigamma and binormal AUC estimates is 0.035.

Table 4.10 summarizes results for the “visually improper” group which includes the “slightly hooked” group (the upper part of the table) and “noticeably hooked” group (the lower part of the table). For the slightly hooked group, the largest absolute differences between the estimated empirical AUCs and two parametric estimates AUC under the bigamma and binormal models are 0.023 and 0.018. The largest absolute difference between the two parametric models is 0.006. For the “noticeably hooked” group, the largest absolute differences between the estimated empirical AUC and two parametric estimate AUCs under the bigamma and binormal models are 0.018 and 0.063. The largest absolute difference between the two parametric models is 0.064.

The two parametric estimated AUCs are farther from the empirical AUCs than in the simulation results for continuous data. This may be due to the fact, that the data in the experimental datasets are not completely continuous (e.g., multiple ties at ‘0’ and ‘100’), therefore there are more opportunities for the differences in the extrapolated parts of the curves. Similar results were observed in the partially grouped data (Table 4.8).

Table 4.9 Estimates for the real data sets (with visually proper fitted ROC curves)

Data sets	Modality	Reader	Emp	Big	Bin	κ	Big-Emp	Bin-Emp	Bin-Big
Interstitial	**1**	**2**	0.630	0.667	0.632	14710	0.037	0.002	-0.035
		4	0.751	0.755	0.753	2.2	0.004	0.002	-0.002
223:84		**5**	0.758	0.761	0.761	11.6	0.003	0.003	0.000
		6	0.761	0.765	0.764	2.9	0.004	0.003	-0.001
	2	**1**	0.751	0.783	0.783	226.6	0.032	0.032	0.000
		2	0.644	0.669	0.649	18627	0.025	0.005	-0.020
		4	0.789	0.807	0.806	6.3	0.018	0.017	-0.001
		5	0.758	0.755	0.755	938.8	-0.003	-0.003	0.000
Nodules	**1**	**2**	0.751	0.821	0.820	5.2	0.070	0.069	-0.001
203:104	**2**	**1**	0.842	0.845	0.842	1.3	0.003	0.000	-0.003
		2	0.751	0.816	0.813	2.9	0.065	0.062	-0.003
Pneumothorax	**1**	**1**	0.876	0.945	0.944	12.1	0.069	0.068	-0.001
200:50		**2**	0.841	0.914	0.919	2947	0.073	0.078	0.005
		4	0.952	0.963	0.963	2.4	0.011	0.011	0.000
		5	0.957	0.957	0.956	4.7	0.000	-0.001	-0.001
	2	**2**	0.968	0.995	1.000	8889.1	0.027	0.032	0.005
		4	1.000	0.999	0.998	109	-0.001	-0.002	-0.001
		5	0.993	0.996	0.996	4	0.003	0.003	0.000

Table 4.10 Estimates for the real data sets (with visually improper fitted ROC curves)

Hook	Data sets	Modality	Reader	Emp	Big	Bin	κ	Big-Emp	Bin-Emp	Bin-Big
S	**Intersitial**	**1**	**1**	0.702	0.725	0.720	1.3	0.023	0.018	-0.005
	223:84	**2**	**6**	0.756	0.762	0.756	0.4	0.006	0.000	-0.006
	Nodules	**1**	**1**	0.843	0.850	0.846	0.7	0.007	0.003	-0.004
	204:103		**4**	0.845	0.866	0.861	0.3	0.021	0.016	-0.005
			6	0.822	0.827	0.822	0.5	0.005	0.000	-0.005
		2	**4**	0.836	0.850	0.845	0.4	0.014	0.009	-0.005
	Pneumothorax	**2**	**1**	0.978	0.985	0.981	0.7	0.007	0.003	-0.004
	200:50		**6**	0.957	0.964	0.959	0.06	0.007	0.002	-0.005
N	**Intersitial**	**1**	**7**	0.596	0.611	0.620	0.2	0.015	0.024	0.009
	223:84	**2**	**7**	0.619	0.633	0.621	0.8	0.014	0.002	-0.012
	Nodules	**1**	**5**	0.797	0.801	0.795	0.4	0.004	-0.002	-0.006
	204:103		**7**	0.663	0.674	0.610	0.1	0.011	-0.053	-0.064
		2	**5**	0.805	0.808	0.803	0.4	0.003	-0.002	-0.005
			6	0.775	0.773	0.768	0.2	-0.002	-0.007	-0.005
			7	0.663	0.645	0.600	0.01	-0.018	-0.063	-0.045
	Pneumothorax	**1**	**6**	0.824	0.828	0.823	0.3	0.004	-0.001	-0.005
	200:50		**7**	0.726	0.743	0.735	0.3	0.017	0.009	-0.008
		2	**7**	0.833	0.843	0.854	0.06	0.010	0.021	0.011

5.0 CONCLUSIONS

Binormal ROC model has an artifact of improper shape of the ROC curves, which imposes restrictions on the proximity of by-point approximation provided by binormal model to concave ROC curves. This is especially evident for true concave ROC curve with the initially high slope. However, for the fitted binormal ROC curve, the non-concave extrapolation in the high-FPF region caused by the high initial slope of the curve can be compensated by the overestimation of the curve in the low-FPF region. As a result the effect of the improperness on the overall area under the ROC curve (AUC) can be substantially reduced.

In scenarios that frequently lead to improper fitted curves the binormal AUC is slightly biased downward, however the magnitude of bias is negligible for common tasks in diagnostic performance evaluation. The individual estimates of binormal and empirical AUCs are not unlikely to be as far as 0.1 from the true value, especially for low sample sizes (20:20). However, the difference between empirical and binormal AUCs is rather small, and in 95% of cases the estimates fall within 0.05 from each other.

Moreover, in the subsets data leading to notably improper fitted binormal ROC curves, the difference between the empirical and binormal AUC estimates appears no worse than in cases of visually proper fitted binormal ROC curves. Analysis of data from the previously conducted observer performance study leads to similar conclusions. Results based fitting binormal ROC curve by optimizing vertical, instead of horizontal, differences (provided in Appendix B), further support these conclusions.

Thus, although inadequate as meaningful extrapolation of diagnostic performance, the improperness of binormal ROC curves often has little effect on the estimates of the overall area under the ROC curve.

6.0 DISCUSSION

This work is focused on the assessment of the estimated area under the complete ROC curve under the binormal model with emphasis on scenarios that are likely to lead to improper ROC curves, but correspond to actually concave ROC curve. To model the concave truth we used a constant-shape family of bigamma ROC curves. The rationale for the focus specifically on bigamma family is three-fold. First, this is a family of concave ROC curves that have been described and advocated by several investigators (9, 28). Second, as a two parameter family of curves on a (0, 1) support it provides a variety of shapes of the curves. Finally, this family is important in practical applications, e.g., it provides an important model for planning inferences based on the partial area under the ROC curve (30).

In our study we considered bigamma scenarios that cover a wide range of AUCs that are commonly observed in practice and the curve shapes that range from almost binormal to highly non-binormal. Scenarios where bi-gamma ROC curves have highly non-binormal shape correspond to curves with very high initial slope (i.e. curves with shape parameter kappa<<1). Naturally, in these scenarios the fitted binormal ROC curves also tend to have an initial high slope (slope at ROC points with fpf near ‘0’). The initial slope of the binormal ROC curve is inversely related to the binormal shape parameter ‘b’, and the smaller the values of b<1 the larger is the improperness, or the “hook”, of the binormal ROC curve in the high-fpf range. Results of the presented simulation study have confirmed that the frequency of improper shape of the fitted binormal ROC curves increases with an increasing value of the initial slope of the true bi-gamma ROC curve (i.e., increasing κ).

Continuous data considered in our study often lead to empirical ROC points that are scattered throughout the range, rather than concentrated in the region of low fpf values. When present, multiple points in the range of medium and high fpf-values can alleviate the effect of the points with low fpf-values, sometimes leading to lower values of the initial slope for binormal ROC curves. Our investigation of the more extreme shapes of the ROC curves was based on the subset of fitted curves with "noticeable improperness" as defined by several investigators (11, 20). The bias of estimated binormal AUC in the corresponding subsets was evaluated using the property of bias of the empirical AUC for continuous data. The corresponding simulation also allowed us to demonstrate a range of plausible differences between the empirical and binormal estimates of the AUC. We found that in many instances the observed differences are less than changes that are typically considered clinically important.

This work has several limitations. First, a specific parametric family, -- constant-shape bi-gamma ROC curves – was used to model the underlying truth. Whereas the constant-shape bigamma ROC curves constitute a flexible family of concave ROC curves there are other types of concave ROC curves that were also used in literature. In our study we alleviated the parametric restrictions by considering subsets of the generated data (i.e., corresponding to different degree of improperness). However, it is also reasonable to consider other types of the underlying ROC curves which could trigger improperness in the fitted binormal curves (e.g., Dorfman, et al (14)).

Furthermore, in estimating bi-normal ROC curves we used a somewhat simplistic approach based on the least square fit. This method generally provides reasonable point estimates (e.g., (7)), but is not as efficient as the maximum likelihood (ML) approach and is not as amenable to statistical inferences. The future studies focused on statistical inferences can be implemented using either ML approach based on the Box-Cox transformation of continuous data (17), combination of data grouping and ML approach for categorical data (11), the ML approach based on ranked data (16), and/or the ROC-GLM approach (5).

In this study we primarily focused on continuous data which ensures the unbiasedness of the empirical area under the ROC curve. In actual studies of diagnostic accuracy it is also frequent to encounter data that are pseudo-continuous (e.g., ties at 0 and 100 for 0-100 ratings of malignancy), or categorical (e.g., 5 ordered test result such as BIRADS scores). For discrete data with operating points scattered in the middle of the range (specifically, with fpf from 0.1 to 0.8) binormal ROC curves are known to provide good approximation for various non-binormal ROC curves (3). However, the properties of the binormal approximation deteriorate for operating points that are poorly distributed (25). The major problem with discrete data with poorly distributed points lies in the multiplicity of possible extrapolations and their possible large impact on the overall indices. This agrees with our results for the partially continuous diagnostic data where all points are concentrated in (0, 0.1). In those results we showed that both empirical (straight-line extrapolations) and the binormal extrapolations could be severely biased and thereby lead to substantially different results.

Finally this study is focused on estimation of a single area under ROC curve. Comparisons of two ROC curve estimated from the same data are often of interest in practice. Assessment of the proximity of the differences in AUCs and their variability between estimates from improper binormal curves and concave curves under the correct model is beyond the scope of this work.

Our investigation demonstrates that with continuous data the improperness of the fitted binormal ROC curve does not have a substantial effect on the estimates of the area under the complete ROC curve (full AUC), when underlying ROC curves have bi-gamma shape. This is an important finding that indicates that for the purpose of estimating full AUC, if the actual shape of the curve is of no interest, the improper shape of the binormal ROC curve could be of little concern. Thus, in practice, for planning of studies and the actual analysis based on the overall indices, such as full AUC improper ROC curves could be used to obtain realistic estimates of the summary indices and their variability. For example, consideration of the improper binormal ROC curve would allow modeling scenarios where analysis of a single partial AUC is more efficient than that based on the full AUC, which is not possible for proper or nearly proper binormal curve. Elimination of the concern that improperness would substantially distort estimates of the full AUC allows for better planning of sample size based on the convenient binormal ROC model. Evaluation of proximity of variability of the AUC estimates from improper binormal curves and concave ROC curves under the correct model is a subject of future research.

APPENDIX A: COMPLETE VERSION OF CORRESPONDING TABLES

This Appendix includes complete version of tables 4.2, 4.3, 4.5, 4.7, and 4.8 presented in the main text. In particular

Table A.1 (complete version of table 4.2).

True bigamma scenario		Improperness	Number of “non-diseased” and “diseased”					
κ	**AUC**	**group**	**20:20**		**50:50**		**100:100**	
1/10	**0.6**	**VP**	0.61	1.46	0.69	1.24	0.77	1.15
		S & b<1	0.39	0.99	0.61	0.99	0.68	0.98
		S & b>1	1.01	2.44	1.02	1.35	1.05	1.26
		N & b<1	0.12	1.00	0.31	0.99	0.41	0.98
		N & b>1	1.02	3.48	1.01	1.52	1.01	1.30
		All	**0.12**	**3.48**	**0.31**	**1.52**	**0.41**	**1.30**
	0.7	**VP**	0.43	2.35	0.60	1.30	0.65	1.04
		S & b<1	0.34	0.93	0.48	0.94	0.56	0.86
		S & b>1	1.04	2.68	1.33	1.33	.	.
		N & b<1	0.12	0.98	0.19	0.87	0.32	0.81
		N & b>1	1.10	3.38	.	.	.	.
		All	**0.12**	**3.38**	**0.19**	**1.33**	**0.32**	**1.04**
	0.8	**VP**	0.44	2.38	0.51	1.16	0.54	0.99
		S & b<1	0.28	0.91	0.33	0.79	0.35	0.75
		S & b>1	1.47	2.76	1.73	2.10	.	.
		N & b<1	0.10	0.87	0.14	0.72	0.19	0.66
		N & b>1	1.82	3.07	.	.	.	.
		All	**0.10**	**3.07**	**0.14**	**2.10**	**0.19**	**0.99**
	0.9	**VP**	0.43	2.32	0.34	2.21	0.36	1.97
		S & b<1	0.25	0.72	0.18	0.63	0.21	0.63
		S & b>1	1.93	2.63	1.91	2.30	1.97	2.16
		N & b<1	0.10	0.74	0.11	0.57	0.13	0.50
		N & b>1	2.18	2.72	.	.	.	.
		All	**0.10**	**2.72**	**0.11**	**2.30**	**0.13**	**2.16**
1/3	**0.6**	**VP**	0.56	1.38	0.71	1.33	0.76	1.23
		S & b<1	0.40	0.99	0.59	0.99	0.68	0.97
		S & b>1	1.01	2.32	1.02	1.45	1.01	1.21
		N & b<1	0.13	0.99	0.35	1.00	0.44	0.99
		N & b>1	1.01	3.72	1.02	1.66	1.01	1.31
		All	**0.13**	**3.72**	**0.35**	**1.66**	**0.44**	**1.31**
	0.7	**VP**	0.51	1.45	0.58	1.43	0.65	1.16
		S & b<1	0.35	0.97	0.43	0.91	0.51	0.88
		S & b>1	1.05	2.65	1.22	1.41	.	.
		N & b<1	0.11	0.97	0.21	0.88	0.35	0.82
		N & b>1	1.06	3.51	1.19	1.19	.	.
		All	**0.11**	**3.51**	**0.21**	**1.43**	**0.35**	**1.16**

Table A.1 Continued

	0.8	**VP**	0.43	2.33	0.47	1.30	0.53	1.20
		S & b<1	0.27	0.85	0.27	0.87	0.37	0.75
		S & b>1	1.18	2.68	1.91	2.40	.	.
		N & b<1	0.11	0.84	0.15	0.74	0.21	0.72
		N & b>1	1.39	3.11	.	.	.	.
		All	**0.11**	**3.11**	**0.15**	**2.40**	**0.21**	**1.20**
	0.9	**VP**	0.43	2.35	0.35	2.20	0.37	1.87
		S & b<1	0.25	0.72	0.21	0.64	0.22	0.64
		S & b>1	1.92	2.70	1.86	2.46	1.91	2.02
		N & b<1	0.11	0.70	0.13	0.56	0.16	0.49
		N & b>1	1.99	2.83	2.28	2.28	.	.
		All	**0.11**	**2.83**	**0.13**	**2.46**	**0.16**	**2.02**
3	**0.6**	**VP**	0.62	1.49	0.71	1.34	0.77	1.24
		S & b<1	0.47	1.00	0.61	0.99	0.72	0.99
		S & b>1	1.01	1.93	1.00	1.47	1.01	1.34
		N & b<1	0.13	1.00	0.39	0.99	0.52	0.99
		N & b>1	1.01	3.80	1.01	2.52	1.01	1.54
		All	**0.13**	**3.80**	**0.39**	**2.52**	**0.52**	**1.54**
	0.7	**VP**	0.47	1.45	0.61	1.40	0.66	1.33
		S & b<1	0.32	0.97	0.46	0.96	0.58	0.91
		S & b>1	1.04	2.79	1.07	2.02	1.16	1.44
		N & b<1	0.17	0.98	0.26	0.96	0.46	0.90
		N & b>1	1.01	3.72	1.07	2.72	1.29	1.40
		All	**0.17**	**3.72**	**0.26**	**2.72**	**0.46**	**1.44**
	0.8	**VP**	0.45	2.42	0.51	1.61	0.56	1.49
		S & b<1	0.27	0.91	0.31	0.84	0.45	0.79
		S & b>1	1.17	2.98	1.30	2.54	1.75	2.19
		N & b<1	0.16	0.92	0.25	0.76	0.39	0.66
		N & b>1	1.17	3.79	1.92	2.76	.	.
		All	**0.16**	**3.79**	**0.25**	**2.76**	**0.39**	**2.19**
	0.9	**VP**	0.43	2.44	0.38	2.30	0.42	2.11
		S & b<1	0.27	0.72	0.23	0.66	0.26	0.62
		S & b>1	1.44	2.93	1.68	2.69	1.76	2.39
		N & b<1	0.16	0.70	0.14	0.55	0.23	0.43
		N & b>1	1.85	3.30	2.02	2.63	.	.
		All	**0.16**	**3.30**	**0.14**	**2.69**	**0.23**	**2.39**
10	**0.6**	**VP**	0.58	1.49	0.74	1.37	0.75	1.30
		S & b<1	0.43	1.00	0.61	0.99	0.71	0.99
		S & b>1	1.01	1.58	1.00	1.57	1.02	1.33
		N & b<1	0.15	1.00	0.40	0.99	0.56	0.99
		N & b>1	1.00	3.75	1.01	2.01	1.01	1.53
		All	**0.15**	**3.75**	**0.40**	**2.01**	**0.56**	**1.53**
	0.7	**VP**	0.51	1.48	0.63	1.47	0.67	1.41
		S & b<1	0.37	0.98	0.50	0.97	0.58	0.90
		S & b>1	1.01	2.80	1.05	1.96	1.12	1.61
		N & b<1	0.12	0.99	0.27	0.93	0.50	0.85
		N & b>1	1.03	3.93	1.06	2.62	1.24	1.61
		All	**0.12**	**3.93**	**0.27**	**2.62**	**0.50**	**1.61**
	0.8	**VP**	0.43	2.37	0.51	1.82	0.58	1.71
		S & b<1	0.28	0.92	0.32	0.88	0.42	0.78
		S & b>1	1.13	2.95	1.24	2.62	1.52	2.17
		N & b<1	0.16	0.89	0.30	0.83	0.44	0.67
		N & b>1	1.19	4.13	1.51	2.80	1.72	2.58
		All	**0.16**	**4.13**	**0.30**	**2.80**	**0.42**	**2.58**
	0.9	**VP**	0.43	2.47	0.36	2.37	0.38	2.23
		S & b<1	0.27	0.76	0.24	0.67	0.28	0.66
		S & b>1	1.32	3.07	1.60	2.73	1.68	2.49
		N & b<1	0.18	0.75	0.22	0.48	.	.
		N & b>1	1.85	3.53	1.79	2.95	2.42	2.44
		All	**0.18**	**3.53**	**0.22**	**2.95**	**0.28**	**2.49**

Table A.2 (complete version of table 4.3)

True bigamma			Number of “non-diseased” and “diseased”								
scenario		Improperness		20:20			50:50			100:100	
κ	AUC	group	Emp	Bin	Bias	Emp	Bin	Bias	Emp	Bin	Bias
1/10	0.6	**V P**	0.652	0.645	*-0.007*	0.636	0.630	*-0.006*	0.626	0.620	*-0.006*
		S & b<1	0.654	0.649	*-0.005*	0.633	0.627	*-0.006*	0.625	0.618	*-0.007*
		S & b>1	0.608	0.601	*-0.007*	0.602	0.596	*-0.006*	0.597	0.590	*-0.007*
		N & b<1	0.574	0.565	*-0.009*	0.583	0.575	*-0.008*	0.589	0.581	*-0.008*
		N & b>1	0.562	0.556	*-0.006*	0.545	0.539	*-0.006*	0.544	0.539	*-0.005*
		All	**0.600**	**0.593**	***-0.007***	**0.600**	**0.593**	***-0.007***	**0.600**	**0.593**	***-0.007***
	0.7	**V P**	0.745	0.733	*-0.012*	0.737	0.726	*-0.011*	0.733	0.721	*-0.012*
		S & b<1	0.748	0.737	*-0.011*	0.727	0.716	*-0.011*	0.723	0.712	*-0.011*
		S & b>1	0.728	0.713	*-0.015*	0.717	0.706	*-0.011*	--	--	--
		N & b<1	0.668	0.654	*-0.014*	0.682	0.670	*-0.012*	0.690	0.678	*-0.012*
		N & b>1	0.730	0.714	*-0.016*	--	--	--	--	--	--
		All	**0.698**	**0.685**	***-0.013***	**0.699**	**0.687**	***-0.012***	**0.700**	**0.688**	***-0.012***
	0.8	**V P**	0.834	0.818	*-0.016*	0.835	0.821	*-0.014*	0.831	0.818	*-0.013*
		S & b<1	0.833	0.820	*-0.013*	0.820	0.809	*-0.011*	0.815	0.803	*-0.012*
		S & b>1	0.926	0.894	*-0.032*	0.884	0.853	*-0.031*	--	--	--
		N & b<1	0.765	0.751	*-0.014*	0.781	0.767	*-0.014*	0.788	0.775	*-0.013*
		N & b>1	0.856	0.831	*-0.025*	--	--	--	--	--	--
		All	**0.800**	**0.785**	***-0.015***	**0.800**	**0.787**	***-0.013***	**0.800**	**0.788**	***-0.012***
	0.9	**V P**	0.939	0.917	*-0.022*	0.922	0.908	*-0.014*	0.919	0.908	*-0.011*
		S & b<1	0.908	0.897	*-0.011*	0.906	0.897	*-0.009*	0.904	0.895	*-0.009*
		S & b>1	0.941	0.904	*-0.037*	0.942	0.898	*-0.044*	0.938	0.893	*-0.045*
		N & b<1	0.850	0.837	*-0.013*	0.875	0.865	*-0.010*	0.882	0.872	*-0.010*
		N & b>1	0.869	0.839	*-0.030*	--	--	--	--	--	--
		All	**0.899**	**0.883**	***-0.016***	**0.900**	**0.889**	***-0.011***	**0.900**	**0.890**	***-0.010***
1/3	0.6	**V P**	0.648	0.645	*-0.003*	0.629	0.627	*-0.002*	0.621	0.618	*-0.003*
		S & b<1	0.644	0.643	*-0.001*	0.629	0.627	*-0.002*	0.619	0.617	*-0.002*
		S & b>1	0.617	0.612	*-0.005*	0.600	0.596	*-0.004*	0.582	0.579	*-0.003*
		N & b<1	0.572	0.567	*-0.005*	0.579	0.576	*-0.003*	0.583	0.580	*-0.003*
		N & b>1	0.566	0.562	*-0.004*	0.552	0.548	*-0.004*	0.549	0.546	*-0.003*
		All	**0.601**	**0.597**	***-0.004***	**0.600**	**0.597**	***-0.003***	**0.600**	**0.597**	***-0.003***
	0.7	**V P**	0.738	0.729	*-0.009*	0.727	0.720	*-0.007*	0.722	0.716	*-0.006*
		S & b<1	0.734	0.729	*-0.005*	0.717	0.711	*-0.006*	0.713	0.707	*-0.006*
		S & b>1	0.725	0.711	*-0.014*	0.679	0.671	*-0.008*	--	--	--
		N & b<1	0.664	0.657	*-0.007*	0.677	0.670	*-0.007*	0.684	0.678	*-0.006*
		N & b>1	0.733	0.719	*-0.014*	0.580	0.572	*-0.008*	--	--	--
		All	**0.700**	**0.693**	***-0.007***	**0.700**	**0.693**	***-0.007***	**0.700**	**0.694**	***-0.006***
	0.8	**V P**	0.826	0.813	*-0.013*	0.826	0.816	*-0.010*	0.822	0.812	*-0.010*
		S & b<1	0.823	0.814	*-0.009*	0.811	0.804	*-0.007*	0.807	0.799	*-0.008*
		S & b>1	0.918	0.887	*-0.031*	0.903	0.868	*-0.035*	--	--	--
		N & b<1	0.758	0.748	*-0.010*	0.775	0.767	*-0.008*	0.781	0.773	*-0.008*
		N & b>1	0.852	0.828	*-0.024*	--	--	--	--	--	--
		All	**0.799**	**0.787**	***-0.012***	**0.800**	**0.792**	***-0.008***	**0.800**	**0.792**	***-0.008***
	0.9	**V P**	0.930	0.911	*-0.019*	0.918	0.906	*-0.012*	0.915	0.906	*-0.009*
		S & b<1	0.901	0.892	*-0.009*	0.900	0.893	*-0.007*	0.899	0.893	*-0.006*
		S & b>1	0.939	0.903	*-0.036*	0.940	0.900	*-0.040*	0.940	0.896	*-0.044*
		N & b<1	0.848	0.838	*-0.010*	0.871	0.863	*-0.008*	0.877	0.870	*-0.007*

Table A.2 Continued

		N & b>1	0.875	0.849	*-0.026*	0.880	0.846	*-0.034*	--	--	--
		All	**0.899**	**0.884**	***-0.015***	**0.900**	**0.891**	***-0.009***	**0.900**	**0.893**	***-0.007***
3	**0.6**	**V P**	0.644	0.642	*-0.002*	0.625	0.625	*0.000*	0.613	0.613	*0.000*
		S & b<1	0.641	0.643	*0.002*	0.618	0.619	*0.001*	0.606	0.606	*0.000*
		S & b>1	0.618	0.615	*-0.003*	0.605	0.602	*-0.003*	0.592	0.590	*-0.002*
		N & b<1	0.570	0.570	*0.000*	0.570	0.569	*-0.001*	0.574	0.573	*-0.001*
		N & b>1	0.568	0.566	*-0.002*	0.555	0.553	*-0.002*	0.557	0.556	*-0.001*
		All	**0.601**	**0.600**	***-0.001***	**0.600**	**0.599**	***-0.001***	**0.600**	**0.599**	***-0.001***
	0.7	**V P**	0.723	0.719	*-0.004*	0.709	0.708	*-0.001*	0.705	0.704	*-0.001*
		S & b<1	0.719	0.719	*0.000*	0.697	0.698	*0.001*	0.691	0.691	*0.000*
		S & b>1	0.710	0.700	*-0.010*	0.678	0.673	*-0.005*	0.682	0.678	*-0.004*
		N & b<1	0.650	0.648	*-0.002*	0.660	0.660	*0.000*	0.667	0.667	*0.000*
		N & b>1	0.697	0.688	*-0.009*	0.657	0.650	*-0.007*	0.645	0.639	*-0.006*
		All	**0.700**	**0.697**	***-0.003***	**0.699**	**0.698**	***-0.001***	**0.700**	**0.699**	***-0.001***
	0.8	**V P**	0.799	0.793	*-0.006*	0.804	0.801	*-0.003*	0.802	0.800	*-0.002*
		S & b<1	0.799	0.798	*-0.001*	0.789	0.790	*0.001*	0.786	0.787	*0.001*
		S & b>1	0.886	0.862	*-0.024*	0.860	0.835	*-0.025*	0.856	0.833	*-0.023*
		N & b<1	0.747	0.745	*-0.002*	0.758	0.758	*0.000*	0.767	0.769	*0.002*
		N & b>1	0.826	0.807	*-0.019*	0.845	0.821	*-0.024*	--	--	--
		All	**0.800**	**0.793**	***-0.007***	**0.800**	**0.797**	***-0.003***	**0.800**	**0.798**	***-0.002***
	0.9	**V P**	0.899	0.887	*-0.012*	0.899	0.894	*-0.005*	0.900	0.897	*-0.003*
		S & b<1	0.886	0.882	*-0.004*	0.886	0.888	*0.002*	0.888	0.891	*0.003*
		S & b>1	0.929	0.901	*-0.028*	0.924	0.892	*-0.032*	0.922	0.888	*-0.034*
		N & b<1	0.840	0.836	*-0.004*	0.861	0.863	*0.002*	0.882	0.886	*0.004*
		N & b>1	0.865	0.842	*-0.023*	0.876	0.848	*-0.028*	--	--	--
		All	**0.899**	**0.885**	***-0.014***	**0.900**	**0.892**	***-0.008***	**0.900**	**0.897**	***-0.003***
10	**0.6**	**V P**	0.644	0.643	*-0.001*	0.624	0.624	*0.000*	0.611	0.611	*0.000*
		S & b<1	0.632	0.634	*0.002*	0.614	0.615	*0.001*	0.602	0.603	*0.001*
		S & b>1	0.614	0.611	*-0.003*	0.607	0.605	*-0.002*	0.594	0.593	*-0.001*
		N & b<1	0.565	0.565	*0.000*	0.568	0.567	*-0.001*	0.571	0.571	*0.000*
		N & b>1	0.570	0.567	*-0.003*	0.560	0.558	*-0.002*	0.563	0.561	*-0.002*
		All	**0.599**	**0.598**	***-0.001***	**0.599**	**0.599**	***0.000***	**0.599**	**0.599**	***0.000***
	0.7	**V P**	0.719	0.716	*-0.003*	0.707	0.706	*-0.001*	0.703	0.703	*0.000*
		S & b<1	0.716	0.718	*0.002*	0.693	0.694	*0.001*	0.689	0.690	*0.001*
		S & b>1	0.692	0.685	*-0.007*	0.678	0.674	*-0.004*	0.679	0.676	*-0.003*
		N & b<1	0.647	0.646	*-0.001*	0.657	0.658	*0.001*	0.661	0.661	*0.000*
		N & b>1	0.694	0.686	*-0.008*	0.648	0.643	*-0.005*	0.650	0.646	*-0.004*
		All	**0.700**	**0.697**	***-0.003***	**0.699**	**0.698**	***-0.001***	**0.701**	**0.700**	***-0.001***
	0.8	**V P**	0.795	0.790	*-0.005*	0.801	0.799	*-0.002*	0.800	0.799	*-0.001*
		S & b<1	0.798	0.799	*0.001*	0.784	0.787	*0.003*	0.784	0.788	*0.004*
		S & b>1	0.869	0.847	*-0.022*	0.834	0.815	*-0.019*	0.839	0.819	*-0.020*
		N & b<1	0.744	0.743	*-0.001*	0.750	0.753	*0.003*	0.755	0.759	*0.004*
		N & b>1	0.816	0.799	*-0.017*	0.831	0.809	*-0.022*	0.833	0.810	*-0.023*
		All	**0.799**	**0.792**	***-0.007***	**0.800**	**0.797**	***-0.003***	**0.800**	**0.799**	***-0.001***
	0.9	**V P**	0.897	0.887	*-0.010*	0.896	0.892	*-0.004*	0.899	0.896	*-0.003*
		S & b<1	0.884	0.882	*-0.002*	0.887	0.892	*0.005*	0.893	0.902	*0.009*
		S & b>1	0.926	0.900	*-0.026*	0.919	0.889	*-0.030*	0.916	0.883	*-0.033*
		N & b<1	0.838	0.835	*-0.003*	0.869	0.876	*0.007*	--	--	--
		N & b>1	0.864	0.843	*-0.021*	0.871	0.845	*-0.026*	0.877	0.853	*-0.024*
		All	**0.900**	**0.886**	***-0.014***	**0.900**	**0.891**	***-0.009***	**0.900**	**0.895**	***-0.005***

Table A.3 (complete version of table 4.5).

Improperness group	**True bigamma scenario**		**Number of “non-diseased” and “diseased”**					
	κ	**AUC**	**20:20**		**50:50**		**100:100**	
V P	1/10	**0.6**	-0.0072	-0.0065	-0.0065	-0.0060	-0.0063	-0.0060
		0.7	-0.0129	-0.0122	-0.0119	-0.0114	-0.0117	-0.0111
		0.8	-0.0163	-0.0156	-0.0139	-0.0134	-0.0137	-0.0131
		0.9	-0.0218	-0.0207	-0.0142	-0.0135	-0.0113	-0.0109
		All	**-0.0143**	**-0.0139**	**-0.0113**	**-0.0110**	**-0.0100**	**-0.0097**
	1/3	**0.6**	-0.0039	-0.0032	-0.0031	-0.0027	-0.0029	-0.0026
		0.7	-0.0089	-0.0083	-0.0074	-0.0070	-0.0069	-0.0065
		0.8	-0.0129	-0.0123	-0.0102	-0.0097	-0.0097	-0.0093
		0.9	-0.0191	-0.0182	-0.0121	-0.0114	-0.0092	-0.0089
		All	**-0.0111**	**-0.0106**	**-0.0081**	**-0.0078**	**-0.0069**	**-0.0067**
	3	**0.6**	-0.0018	-0.0011	-0.0006	-0.0003	-0.0004	-0.0002
		0.7	-0.0041	-0.0035	-0.0019	-0.0016	-0.0012	-0.0010
		0.8	-0.0066	-0.0060	-0.0032	-0.0029	-0.0021	-0.0018
		0.9	-0.0121	-0.0114	-0.0057	-0.0052	-0.0033	-0.0030
		All	**-0.0063**	**-0.0060**	**-0.0030**	**-0.0028**	**-0.0018**	**-0.0017**
	10	**0.6**	-0.0014	-0.0007	-0.0002	0.0002	-0.0001	0.0001
		0.7	-0.0037	-0.0031	-0.0014	-0.0010	-0.0008	-0.0006
		0.8	-0.0057	-0.0051	-0.0023	-0.0019	-0.0012	-0.0009
		0.9	-0.0106	-0.0099	-0.0042	-0.0037	-0.0027	-0.0023
		All	**-0.0054**	**-0.0051**	**-0.0021**	**-0.0019**	**-0.0012**	**-0.0011**
S & b<1	1/10	**0.6**	-0.0061	-0.0045	-0.0060	-0.0052	-0.0065	-0.0061
		0.7	-0.0112	-0.0101	-0.0112	-0.0106	-0.0112	-0.0108
		0.8	-0.0130	-0.0123	-0.0119	-0.0115	-0.0121	-0.0118
		0.9	-0.0113	-0.0109	-0.0091	-0.0088	-0.0090	-0.0088
		All	**-0.0110**	**-0.0106**	**-0.0098**	**-0.0096**	**-0.0098**	**-0.0097**
	1/3	**0.6**	-0.0020	-0.0003	-0.0023	-0.0015	-0.0026	-0.0021
		0.7	-0.0063	-0.0051	-0.0061	-0.0055	-0.0060	-0.0056
		0.8	-0.0094	-0.0086	-0.0078	-0.0073	-0.0078	-0.0075
		0.9	-0.0101	-0.0096	-0.0071	-0.0068	-0.0068	-0.0066
		All	**-0.0081**	**-0.0076**	**-0.0064**	**-0.0062**	**-0.0064**	**-0.0062**
	3	**0.6**	0.0010	0.0029	0.0004	0.0014	0.0001	0.0007
		0.7	-0.0006	0.0010	0.0002	0.0011	0.0003	0.0009
		0.8	-0.0017	-0.0003	0.0008	0.0018	0.0010	0.0018
		0.9	-0.0048	-0.0038	0.0015	0.0025	0.0029	0.0040
		All	**-0.0018**	**-0.0011**	**0.0010**	**0.0015**	**0.0010**	**0.0013**
	10	**0.6**	0.0012	0.0032	0.0011	0.0021	0.0006	0.0012
		0.7	0.0007	0.0025	0.0009	0.0020	0.0006	0.0015
		0.8	-0.0006	0.0011	0.0019	0.0034	0.0032	0.0047
		0.9	-0.0033	-0.0018	0.0044	0.0063	0.0072	0.0102
		All	**-0.0004**	**0.0005**	**0.0021**	**0.0027**	**0.0016**	**0.0021**
S & b>1	1/10	**0.6**	-0.0076	-0.0059	-0.0071	-0.0056	-0.0075	-0.0054
		0.7	-0.0185	-0.0130	.	.	--	--
		0.8	-0.0339	-0.0317	-0.0869	0.0238	--	--
		0.9	-0.0382	-0.0371	-0.0451	-0.0431	-0.0827	-0.0066
		All	**-0.0307**	**-0.0291**	**-0.0241**	**-0.0183**	**-0.0154**	**-0.0048**
	1/3	**0.6**	-0.0055	-0.0039	-0.0041	-0.0030	-0.0038	-0.0021
		0.7	-0.0154	-0.0118	-0.0098	-0.0054	--	--
		0.8	-0.0318	-0.0297	-0.0377	-0.0328	--	--
		0.9	-0.0365	-0.0355	-0.0414	-0.0397	-0.0498	-0.0377
		All	**-0.0287**	**-0.0273**	**-0.0218**	**-0.0180**	**-0.0084**	**-0.0025**
	3	**0.6**	-0.0041	-0.0028	-0.0027	-0.0020	-0.0020	-0.0013
		0.7	-0.0101	-0.0082	-0.0060	-0.0048	-0.0057	-0.0033
		0.8	-0.0250	-0.0233	-0.0261	-0.0230	-0.0264	-0.0214
		0.9	-0.0284	-0.0276	-0.0331	-0.0323	-0.0353	-0.0342

Table A.3 Continued

		All	**-0.0232**	**-0.0224**	**-0.0229**	**-0.0216**	**-0.0158**	**-0.0133**
	10	**0.6**	-0.0036	-0.0023	-0.0023	-0.0016	-0.0017	-0.0012
		0.7	-0.0080	-0.0064	-0.0049	-0.0040	-0.0042	-0.0030
		0.8	-0.0222	-0.0207	-0.0211	-0.0185	-0.0229	-0.0165
		0.9	-0.0260	-0.0253	-0.0306	-0.0300	-0.0326	-0.0320
		All	**-0.0212**	**-0.0205**	**-0.0220**	**-0.0211**	**-0.0194**	**-0.0177**
N & b<1	**1/10**	**0.6**	-0.0087	-0.0080	-0.0082	-0.0078	-0.0078	-0.0076
		0.7	-0.0138	-0.0131	-0.0128	-0.0125	-0.0122	-0.0120
		0.8	-0.0146	-0.0140	-0.0132	-0.0129	-0.0130	-0.0127
		0.9	-0.0126	-0.0120	-0.0102	-0.0098	-0.0096	-0.0093
		All	**-0.0123**	**-0.0119**	**-0.0111**	**-0.0109**	**-0.0107**	**-0.0106**
	1/3	**0.6**	-0.0048	-0.0039	-0.0038	-0.0033	-0.0036	-0.0033
		0.7	-0.0077	-0.0069	-0.0070	-0.0065	-0.0066	-0.0063
		0.8	-0.0103	-0.0096	-0.0083	-0.0078	-0.0082	-0.0078
		0.9	-0.0109	-0.0102	-0.0080	-0.0075	-0.0071	-0.0067
		All	**-0.0079**	**-0.0075**	**-0.0063**	**-0.0061**	**-0.0059**	**-0.0058**
	3	**0.6**	-0.0010	0.0000	-0.0011	-0.0004	-0.0008	-0.0003
		0.7	-0.0029	-0.0016	-0.0004	0.0006	-0.0005	0.0006
		0.8	-0.0027	-0.0010	-0.0005	0.0013	0.0003	0.0030
		0.9	-0.0052	-0.0031	0.0005	0.0043	-0.0002	0.0089
		All	**-0.0019**	**-0.0012**	**-0.0006**	**-0.0001**	**-0.0005**	**-0.0001**
	10	**0.6**	-0.0010	0.0001	-0.0008	-0.0001	-0.0007	-0.0002
		0.7	-0.0022	-0.0008	0.0006	0.0019	-0.0008	0.0008
		0.8	-0.0027	-0.0004	0.0005	0.0040	-0.0001	0.0073
		0.9	-0.0049	-0.0011	0.0022	0.0108	--	--
		All	**-0.0015**	**-0.0007**	**-0.0001**	**0.0005**	**-0.0006**	**-0.0001**
N & b>1	**1/10**	**0.6**	-0.0061	-0.0052	-0.0066	-0.0053	-0.0070	-0.0028
		0.7	-0.0181	-0.0149	--	--	--	--
		0.8	-0.0265	-0.0237	--	--	--	--
		0.9	-0.0332	-0.0263	--	--	--	--
		All	**-0.0092**	**-0.0080**	**-0.0066**	**-0.0053**	**-0.0070**	**-0.0028**
	1/3	**0.6**	-0.0044	-0.0036	-0.0037	-0.0029	-0.0038	-0.0022
		0.7	-0.0154	-0.0132	.	.	--	--
		0.8	-0.0251	-0.0232	--	--	--	--
		0.9	-0.0279	-0.0245	.	.	--	--
		All	-0.0082	-0.0073	-0.0039	-0.0029	-0.0038	-0.0022
	3	**0.6**	-0.0029	-0.0024	-0.0021	-0.0016	-0.0017	-0.0012
		0.7	-0.0096	-0.0086	-0.0073	-0.0048	-0.0287	0.0176
		0.8	-0.0187	-0.0178	-0.0247	-0.0229	--	--
		0.9	-0.0234	-0.0223	-0.0294	-0.0269	--	--
		All	**-0.0096**	**-0.0090**	**-0.0050**	**-0.0041**	**-0.0017**	**-0.0012**
	10	**0.6**	-0.0027	-0.0022	-0.0019	-0.0015	-0.0018	-0.0014
		0.7	-0.0086	-0.0077	-0.0056	-0.0044	-0.0056	-0.0024
		0.8	-0.0174	-0.0166	-0.0223	-0.0210	-0.0263	-0.0198
		0.9	-0.0208	-0.0198	-0.0266	-0.0249	-0.0279	-0.0211
		All	**-0.0098**	**-0.0093**	**-0.0062**	**-0.0054**	**-0.0025**	**-0.0018**

Table A.4 (complete version of table 4.7).

Improperness	True bigamma scenario		Number of “non-diseased” and “diseased”					
group	κ	AUC	20:20		50:50		100:100	
V P	1/10	**0.6**	-0.0246	0.0122	-0.0163	0.0045	-0.0133	0.0017
		0.7	-0.0279	0.0041	-0.0222	-0.0003	-0.0189	-0.0039
		0.8	-0.0328	-0.0017	-0.0237	-0.0018	-0.0201	-0.0049
		0.9	-0.0532	-0.0057	-0.0460	-0.0021	-0.0184	-0.0036
		All	***-0.0450***	***0.0057***	***-0.0257***	***0.0018***	***-0.0186***	***-0.0002***
	1/3	**0.6**	-0.0211	0.0167	-0.0142	0.0096	-0.0103	0.0060
		0.7	-0.0257	0.0115	-0.0182	0.0053	-0.0151	0.0033
		0.8	-0.0295	0.0058	-0.0210	0.0038	-0.0178	0.0010
		0.9	-0.0514	-0.0029	-0.0441	0.0013	-0.0168	0.0001
		All	***-0.0397***	***0.0111***	***-0.0220***	***0.0062***	***-0.0163***	***0.0039***
	3	**0.6**	-0.0204	0.0206	-0.0119	0.0123	-0.0083	0.0089
		0.7	-0.0226	0.0193	-0.0141	0.0132	-0.0101	0.0099
		0.8	-0.0255	0.0186	-0.0165	0.0151	-0.0123	0.0109
		0.9	-0.0450	0.0100	-0.0373	0.0142	-0.0165	0.0116
		All	***-0.0284***	***0.0182***	***-0.0180***	***0.0139***	***-0.0128***	***0.0105***
	10	**0.6**	-0.0197	0.0222	-0.0114	0.0137	-0.0081	0.0093
		0.7	-0.0230	0.0210	-0.0136	0.0146	-0.0096	0.0103
		0.8	-0.0254	0.0207	-0.0157	0.0166	-0.0116	0.0131
		0.9	-0.0402	0.0149	-0.0363	0.0186	-0.0304	0.0148
		All	***-0.0273***	***0.0203***	***-0.0174***	***0.0162***	***-0.0131***	***0.0124***
	All		**-0.0336**	**0.0164**	**-0.0203**	**0.0137**	**-0.0152**	**0.0107**
S & b<1	1/10	**0.6**	-0.0257	0.0204	-0.0167	0.0071	-0.0144	0.0024
		0.7	-0.0304	0.0120	-0.0230	0.0019	-0.0193	-0.0021
		0.8	-0.0317	0.0049	-0.0224	-0.0001	-0.0201	-0.0035
		0.9	-0.0266	-0.0005	-0.0187	0.0005	-0.0161	-0.0014
		All	***-0.0290***	***0.0075***	***-0.0212***	***0.0021***	***-0.0185***	***-0.0011***
	1/3	**0.6**	-0.0264	0.0271	-0.0157	0.0133	-0.0114	0.0079
		0.7	-0.0275	0.0212	-0.0186	0.0096	-0.0147	0.0043
		0.8	-0.0281	0.0130	-0.0203	0.0088	-0.0172	0.0029
		0.9	-0.0260	0.0039	-0.0177	0.0045	-0.0149	0.0025
		All	***-0.0268***	***0.0154***	***-0.0188***	***0.0085***	***-0.0156***	***0.0043***
	3	**0.6**	-0.0235	0.0314	-0.0124	0.0174	-0.0089	0.0110
		0.7	-0.0247	0.0340	-0.0148	0.0194	-0.0106	0.0133
		0.8	-0.0260	0.0320	-0.0158	0.0229	-0.0107	0.0166
		0.9	-0.0250	0.0214	-0.0141	0.0238	-0.0096	0.0199
		All	***-0.0251***	***0.0296***	***-0.0144***	***0.0211***	***-0.0102***	***0.0151***
	10	**0.6**	-0.0237	0.0326	-0.0118	0.0176	-0.0088	0.0116
		0.7	-0.0248	0.0345	-0.0152	0.0219	-0.0104	0.0143
		0.8	-0.0241	0.0332	-0.0155	0.0288	-0.0102	0.0225
		0.9	-0.0251	0.0288	-0.0137	0.0310	-0.0055	0.0262
		All	***-0.0245***	***0.0318***	***-0.0142***	***0.0240***	***-0.0096***	***0.0165***
	All		**-0.0269**	**0.0234**	**-0.0192**	**0.0156**	**-0.0168**	**0.0092**
S & b>1	1/10	**0.6**	-0.0189	0.0073	-0.0147	0.0007	-0.0097	-0.0026
		0.7	-0.0405	-0.0013	-0.0110	-0.0110	--	--
		0.8	-0.0484	-0.0163	-0.0359	-0.0272	--	--
		0.9	-0.0516	-0.0199	-0.0499	-0.0333	-0.0477	-0.0417
		All	***-0.0512***	***0.0001***	***-0.0490***	***-0.0008***	***-0.0477***	***-0.0026***
	1/3	**0.6**	-0.0183	0.0103	-0.0110	0.0043	-0.0072	0.0032
		0.7	-0.0365	0.0024	-0.0096	-0.0049	--	--
		0.8	-0.0443	-0.0123	-0.0395	-0.0306	--	--
		0.9	-0.0508	-0.0183	-0.0487	-0.0282	-0.0466	-0.0422
		All	***-0.0502***	***0.0023***	***-0.0477***	***0.0027***	***-0.0426***	***0.0032***
	3	**0.6**	-0.0168	0.0106	-0.0109	0.0068	-0.0074	0.0049
		0.7	-0.0331	0.0052	-0.0131	0.0032	-0.0116	0.0005

Table A.4 Continued

		0.8	-0.0424	-0.0037	-0.0363	-0.0091	-0.0277	-0.0192
		0.9	-0.0460	-0.0091	-0.0433	-0.0162	-0.0424	-0.0256
		All	***-0.0449***	***0.0032***	***-0.0428***	***0.0032***	***-0.0404***	***0.0041***
	10	**0.6**	-0.0167	0.0120	-0.0103	0.0069	-0.0069	0.0043
		0.7	-0.0310	0.0071	-0.0127	0.0039	-0.0105	0.0022
		0.8	-0.0417	-0.0005	-0.0352	-0.0029	-0.0305	-0.0065
		0.9	-0.0448	-0.0063	-0.0420	-0.0141	-0.0403	-0.0224
		All	***-0.0437***	***0.0038***	***-0.0411***	***0.0033***	***-0.0397***	***0.0032***
	All		**-0.0466**	**0.0031**	**-0.0434**	**0.0032**	**-0.0399**	**0.0033**
All			**-0.0428**	**0.0199**	**-0.0370**	**0.0144**	**-0.0275**	**0.0089**
N & b<1	**1/10**	**0.6**	-0.0380	0.0186	-0.0237	0.0076	-0.0179	0.0027
		0.7	-0.0414	0.0114	-0.0279	0.0024	-0.0223	-0.0016
		0.8	-0.0387	0.0061	-0.0279	0.0006	-0.0223	-0.0035
		0.9	-0.0318	0.0019	-0.0221	-0.0003	-0.0176	-0.0018
		All	***-0.0385***	***0.0123***	***-0.0263***	***0.0040***	***-0.0213***	***0.0001***
	1/3	**0.6**	-0.0331	0.0275	-0.0190	0.0144	-0.0139	0.0077
		0.7	-0.0329	0.0233	-0.0221	0.0111	-0.0171	0.0053
		0.8	-0.0347	0.0154	-0.0232	0.0089	-0.0189	0.0037
		0.9	-0.0308	0.0055	-0.0206	0.0050	-0.0158	0.0025
		All	***-0.0330***	***0.0212***	***-0.0215***	***0.0113***	***-0.0169***	***0.0059***
	3	**0.6**	-0.0310	0.0332	-0.0169	0.0181	-0.0110	0.0116
		0.7	-0.0322	0.0330	-0.0177	0.0222	-0.0122	0.0135
		0.8	-0.0305	0.0376	-0.0193	0.0256	-0.0139	0.0165
		0.9	-0.0258	0.0258	-0.0156	0.0261	-0.0045	0.0200
		All	***-0.0312***	***0.0333***	***-0.0172***	***0.0205***	***-0.0111***	***0.0122***
	10	**0.6**	-0.0294	0.0333	-0.0160	0.0171	-0.0107	0.0115
		0.7	-0.0320	0.0372	-0.0163	0.0247	-0.0120	0.0181
		0.8	-0.0334	0.0368	-0.0177	0.0271	-0.0168	0.0191
		0.9	-0.0298	0.0312	-0.0102	0.0402	--	--
		All	***-0.0308***	***0.0352***	***-0.0161***	***0.0200***	***-0.0107***	***0.0122***
	All		**-0.0352**	**0.0241**	**-0.0240**	**0.0124**	**-0.0199**	**0.0062**
N & b>1	**1/10**	**0.6**	-0.0197	0.0063	-0.0137	0.0019	-0.0130	-0.0009
		0.7	-0.0307	0.0004	--	--	--	--
		0.8	-0.0330	-0.0125	--	--	--	--
		0.9	-0.0352	-0.0206	--	--	--	--
		All	***-0.0305***	***0.0053***	***-0.0137***	***0.0019***	***-0.0130***	***-0.0009***
	1/3	**0.6**	-0.0186	0.0084	-0.0113	0.0039	-0.0091	0.0011
		0.7	-0.0290	0.0009	-0.0087	-0.0087	--	--
		0.8	-0.0334	-0.0111	--	--	--	--
		0.9	-0.0323	-0.0190	-0.0344	-0.0344	--	--
		All	***-0.0291***	***0.0070***	***-0.0113***	***0.0039***	***-0.0091***	***0.0011***
	3	**0.6**	-0.0169	0.0106	-0.0097	0.0057	-0.0067	0.0041
		0.7	-0.0254	0.0042	-0.0242	0.0035	-0.0074	-0.0038
		0.8	-0.0305	-0.0033	-0.0308	-0.0163	--	--
		0.9	-0.0328	-0.0103	-0.0332	-0.0177	--	--
		All	***-0.0293***	***0.0075***	***-0.0289***	***0.0051***	***-0.0067***	***0.0041***
	10	**0.6**	-0.0169	0.0099	-0.0099	0.0062	-0.0075	0.0040
		0.7	-0.0257	0.0056	-0.0172	0.0021	-0.0075	-0.0002
		0.8	-0.0302	-0.0020	-0.0307	-0.0124	-0.0325	-0.0097
		0.9	-0.0318	-0.0069	-0.0328	-0.0140	-0.0258	-0.0231
		All	***-0.0288***	***0.0072***	***-0.0296***	***0.0054***	***-0.0097***	***0.0040***
	All		**-0.0293**	**0.0072**	**-0.0287**	**0.0052**	**-0.0086**	**0.0040**
All			**-0.0335**	**0.0219**	**-0.0245**	**0.0120**	**-0.0198**	**0.0061**

Table A.5 (complete version of table 4.8).

True bigamma scenario		Improperness						
κ	AUC	group	Emp	Bin	Emp-AUC	Bin-AUC	Bin-Emp	N
1/10	0.6	V P	0.610	0.721			0.111	772
		S & b<1	0.610	0.667			0.058	293
		N & b<1	0.592	0.486			-0.106	8935
		All	0.594	0.510	-0.006	-0.090	-0.084	10000
	0.7	V P	0.715	0.815			0.100	67
		S & b<1	0.725	0.786			0.061	110
		N & b<1	0.695	0.595			-0.099	9823
		All	0.695	0.599	-0.005	-0.101	-0.096	10000
	0.8	V P	0.843	0.903			0.061	21
		S & b<1	0.829	0.862			0.033	105
		N & b<1	0.796	0.726			-0.070	9874
		All	0.797	0.728	-0.003	-0.072	-0.069	10000
	0.9	V P	0.934	0.956			0.022	70
		S & b<1	0.926	0.932			0.006	998
		N & b<1	0.895	0.871			-0.024	8932
		All	0.898	0.878	-0.002	-0.022	-0.020	10000
1/3	0.6	V P	0.584	0.695			0.111	3337
		S & b<1	0.585	0.631			0.046	452
		S & b>1	0.550	0.665			0.115	124
		N & b<1	0.571	0.498			-0.074	6061
		N & b>1	0.523	0.611			0.088	26
		All	0.576	0.572	-0.024	-0.028	-0.004	10000
	0.7	V P	0.685	0.788			0.103	1843
		S & b<1	0.681	0.733			0.052	906
		N & b<1	0.668	0.620			-0.048	7251
		All	0.672	0.661	-0.028	-0.039	-0.011	10000
	0.8	V P	0.795	0.867			0.072	915
		S & b<1	0.791	0.828			0.037	1202
		N & b<1	0.774	0.739			-0.036	7883
		All	0.778	0.761	-0.022	-0.039	-0.017	10000
	0.9	V P	0.909	0.940			0.031	902
		S & b<1	0.899	0.912			0.012	2430
		N & b<1	0.882	0.862			-0.020	6668
		All	0.889	0.881	-0.011	-0.019	-0.008	10000
3	0.6	V P	0.561	0.674			0.113	4151
		S & b<1	0.556	0.591			0.035	326
		S & b>1	0.540	0.656			0.116	880
		N & b<1	0.546	0.472			-0.074	4155
		N & b>1	0.517	0.609			0.092	488
		All	0.551	0.583	-0.049	-0.017	0.032	10000
	0.7	V P	0.630	0.746			0.116	5514
		S & b<1	0.629	0.676			0.047	747
		S & b>1	0.568	0.682			0.114	27
		N & b<1	0.617	0.582			-0.035	3711
		N & b>1	0.562	0.675			0.113	1
		All	0.625	0.680	-0.075	-0.020	0.055	10000
	0.8	V P	0.728	0.824			0.096	6076
		S & b<1	0.724	0.767			0.043	1268
		N & b<1	0.715	0.700			-0.015	2656
		All	0.724	0.784	-0.076	-0.016	0.060	10000

Table A.5 Continued

	0.9	**V P**	0.852	0.907			0.055	7016
		S & b<1	0.844	0.867			0.023	1696
		N & b<1	0.835	0.825			-0.011	1288
		All	0.849	0.890	-0.051	-0.010	0.041	10000
10	**0.6**	**V P**	0.557	0.669			0.111	4055
		S & b<1	0.549	0.579			0.030	281
		S & b>1	0.539	0.657			0.119	1015
		N & b<1	0.541	0.463			-0.078	3946
		N & b>1	0.516	0.609			0.093	703
		All	0.546	0.580	-0.054	-0.020	0.034	10000
	0.7	**V P**	0.618	0.736			0.118	6136
		S & b<1	0.616	0.663			0.047	605
		S & b>1	0.574	0.690			0.116	111
		N & b<1	0.607	0.571			-0.036	3138
		N & b>1	0.540	0.638			0.098	10
		All	0.614	0.679	-0.086	-0.021	0.065	10000
	0.8	**V P**	0.712	0.813			0.101	7016
		S & b<1	0.707	0.750			0.044	1006
		N & b<1	0.698	0.686			-0.012	1978
		All	0.708	0.782	-0.092	-0.018	0.073	10000
	0.9	**V P**	0.837	0.900			0.063	8201
		S & b<1	0.829	0.855			0.025	1076
		N & b<1	0.823	0.813			-0.010	723
		All	0.835	0.889	-0.065	-0.011	0.053	10000

APPENDIX B: VERTICAL FITTED RESULTS FOR THE CORRESPONDING TABLES

This Appendix provides tables summarizing results for the vertically fitted binormal ROC curve for the bigamma scenarios with κ=0.1. The corresponding tables in the main text are indicates in the headings.

Table B.1 The range of values of parameter b of fitted binormal ROC curve (supplement to table 4.2)

	Improperness	**Number of “non-diseased” and “diseased”**					
AUC	**group**	**20/20**		**50/50**		**100/100**	
0.6	**VP**	0.68	1.65	0.73	1.23	0.77	1.16
	S & b<1	0.38	0.99	0.57	0.99	0.66	0.98
	S & b>1	1.01	1.79	1.01	1.44	1.05	1.14
	N & b<1	0.2	0.99	0.3	0.99	0.33	0.99
	N & b>1	1.01	3.3	1.02	1.91	1.01	1.36
	All	**0.2**	**3.3**	**0.3**	**1.91**	**0.33**	**1.36**
0.7	**VP**	0.45	1.51	0.71	1.16	0.72	0.9
	S & b<1	0.35	0.98	0.4	0.91	0.63	0.83
	S & b>1	1.02	1.44	.	.	.	.
	N & b<1	0.19	0.98	0.24	0.89	0.28	0.8
	N & b>1	1.06	1.81	1.2	1.2	.	.
	All	**0.19**	**1.81**	**0.24**	**1.2**	**0.28**	**0.9**
0.8	**VP**	0.4	4.2	.	.	.	.
	S & b<1	0.33	0.82	0.34	0.44	0.35	0.4
	S & b>1	.	.	.	.	.	.
	N & b<1	0.18	0.82	0.22	0.69	0.25	0.4
	N & b>1	.	.	.	.	.	.
	All	**0.18**	**4.2**	**0.22**	**0.69**	**0.25**	**0.4**
0.9	**VP**	0.4	4.22	0.4	0.59	0.41	0.55
	S & b<1	0.31	0.43	0.32	0.43	0.32	0.42
	S & b>1	.	.	.	.	.	.
	N & b<1	0.2	0.42	0.24	0.37	0.27	0.36
	N & b>1	.	.	.	.	.	.
	All	**0.2**	**4.22**	**0.24**	**0.59**	**0.27**	**0.55**

Table B.2 Estimated empirical and binormal AUC (supplement to table 4.3)

		Number of “non-diseased” and “diseased”											
	Improperness	**20:20**				**50:50**				**100:100**			
AUC	**group**	**Emp**	**Bin**	**Bias**	**N**	**Emp**	**Bin**	**Bias**	**N**	**Emp**	**Bin**	**Bias**	**N**
0.6	**V P**	0.645	0.639	*-0.006*	2167	0.636	0.630	*-0.006*	1537	0.626	0.621	*-0.005*	802
	S & b<1	0.632	0.626	*-0.006*	710	0.639	0.634	*-0.005*	975	0.630	0.626	*-0.004*	931
	S & b>1	0.607	0.599	*-0.008*	196	0.581	0.573	*-0.008*	67	0.570	0.561	*-0.009*	7
	N & b<1	0.587	0.584	*-0.003*	6197	0.589	0.585	*-0.004*	7245	0.594	0.590	*-0.004*	8244
	N & b>1	0.549	0.538	*-0.011*	730	0.542	0.531	*-0.011*	176	0.546	0.536	*-0.010*	16
	All	**0.600**	**0.595**	***-0.005***	**10000**	**0.600**	**0.596**	***-0.004***	**10000**	**0.600**	**0.596**	***-0.004***	**10000**
0.7	**V P**	0.714	0.701	*-0.013*	935	0.720	0.710	*-0.010*	193	0.729	0.719	*-0.010*	16
	S & b<1	0.711	0.699	*-0.012*	392	0.707	0.697	*-0.010*	289	0.716	0.708	*-0.008*	78
	S & b>1	0.620	0.601	*-0.019*	32	--	--	--	0	--	--	--	0
	N & b<1	0.699	0.695	*-0.004*	8597	0.699	0.696	*-0.003*	9517	0.700	0.696	*-0.004*	9906
	N & b>1	0.573	0.549	*-0.024*	44	0.536	0.523	*-0.013*	1	--	--	--	0
	All	**0.700**	**0.695**	***-0.005***	**10000**	**0.700**	**0.696**	***-0.004***	**10000**	**0.700**	**0.696**	***-0.004***	**10000**
0.8	**V P**	0.866	0.856	*-0.010*	169	--	--	--	0	--	--	--	0
	S & b<1	0.912	0.904	*-0.008*	765	0.905	0.899	*-0.006*	149	0.895	0.888	*-0.007*	18
	S & b>1	--	--	--	0	--	--	--	0	--	--	--	0
	N & b<1	0.790	0.785	*-0.005*	9066	0.798	0.794	*-0.004*	9851	0.800	0.796	*-0.004*	9982
	N & b>1	--	--	--	0	--	--	--	0	--	--	--	0
	All	**0.800**	**0.796**	***-0.004***	**10000**	**0.800**	**0.796**	***-0.004***	**10000**	**0.800**	**0.796**	***-0.004***	**10000**
0.9	**V P**	0.971	0.970	*-0.001*	1803	0.960	0.960	*0.000*	645	0.954	0.955	*0.001*	149
	S & b<1	0.926	0.921	*-0.005*	3438	0.921	0.919	*-0.002*	4804	0.915	0.914	*-0.001*	5527
	S & b>1	--	--	--	0	--	--	--	--	--	--	--	0
	N & b<1	0.854	0.849	*-0.005*	4759	0.870	0.867	*-0.003*	4551	0.878	0.876	*-0.002*	4324
	N & b>1	--	--	--	0	--	--	--	--	--	--	--	0
	All	**0.899**	**0.895**	***-0.004***	**10000**	**0.900**	**0.898**	***-0.002***	**10000**	**0.900**	**0.898**	***-0.002***	**10000**

Table B.3 95% confidence interval for bias of empirical and binormal estimates of the AUC (supplement to table 4.4)

No. of “non-d” & “d”	**AUC**	**Empirical - True**		**Binormal - True**		**Binormal-Empricial**	
20:20	0.6	-0.001	0.002	-0.006	-0.003	-0.005	-0.005
	0.7	-0.002	0.001	-0.007	-0.003	-0.005	-0.005
	0.8	-0.001	0.002	-0.006	-0.003	-0.005	-0.004
	0.9	-0.002	0.001	-0.006	-0.003	-0.004	-0.004
50:50	0.6	-0.001	0.001	-0.006	-0.003	-0.005	-0.004
	0.7	-0.001	0.001	-0.005	-0.003	-0.004	-0.004
	0.8	-0.001	0.000	-0.005	-0.003	-0.004	-0.004
	0.9	-0.001	0.001	-0.003	-0.002	-0.002	-0.002
100:100	0.6	-0.001	0.001	-0.005	-0.003	-0.004	-0.004
	0.7	-0.001	0.001	-0.004	-0.003	-0.004	-0.003
	0.8	-0.001	0.000	-0.004	-0.003	-0.004	-0.003
	0.9	-0.001	0.000	-0.002	-0.001	-0.002	-0.001

Table B.4 95% confidence interval for bias between empirical and binormal estimates of the AUC (supplement of table 4.5)

Improperness group	**AUC**	**Number of “non-diseased” and “diseased”**					
		20:20		**50:50**		**100:100**	
V P	**0.6**	-0.007	-0.006	-0.006	-0.005	-0.005	-0.005
	0.7	-0.013	-0.012	-0.011	-0.009	-0.013	-0.009
	0.8	-0.013	-0.008	--	--	--	--
	0.9	-0.002	0.000	0.000	0.000	0.001	0.002
	All	**-0.006**	**-0.005**	**-0.005**	**-0.004**	**-0.005**	**-0.004**
S & b<1	**0.6**	-0.007	-0.006	-0.005	-0.005	-0.005	-0.005
	0.7	-0.013	-0.011	-0.010	-0.009	-0.009	-0.008
	0.8	-0.009	-0.008	-0.007	-0.006	-0.010	-0.006
	0.9	-0.005	-0.005	-0.002	-0.002	-0.001	-0.001
	All	**-0.006**	**-0.006**	**-0.003**	**-0.003**	**-0.002**	**-0.002**
S & b>1	**0.6**	-0.009	-0.006	-0.010	-0.006	-0.013	-0.006
	0.7	-0.023	-0.016	--	--	--	--
	0.8	--	--	--	--	--	--
	0.9	--	--	--	--	--	--
	All	**-0.011**	**-0.007**	**-0.010**	**-0.006**	**-0.013**	**-0.006**
N & b<1	**0.6**	-0.004	-0.003	-0.004	-0.004	-0.004	-0.004
	0.7	-0.004	-0.003	-0.004	-0.003	-0.003	-0.003
	0.8	-0.004	-0.004	-0.004	-0.004	-0.004	-0.003
	0.9	-0.005	-0.004	-0.003	-0.003	-0.002	-0.002
	All	**-0.004**	**-0.004**	**-0.004**	**-0.004**	**-0.003**	**-0.003**
N & b>1	**0.6**	-0.012	-0.010	-0.012	-0.010	-0.014	-0.006
	0.7	-0.029	-0.019	--	--	--	--
	0.8	--	--	--	--	--	--
	0.9	--	--	--	--	--	--
	All	**-0.013**	**-0.011**	**-0.012**	**-0.010**	**-0.014**	**-0.006**

Table B.5 95% equal-tail range of the sampling distribution of differences in AUC values (supplement to table 4.6)

No. of “non-d” & “d”	**AUC**	**Empirical-True**		**Binormal-True**		**Binormal-Empirical**	
20:20	**0.6**	-0.183	0.173	-0.188	0.165	-0.026	0.012
	0.7	-0.175	0.153	-0.179	0.140	-0.024	0.012
	0.8	-0.156	0.133	-0.151	0.123	-0.019	0.008
	0.9	-0.123	0.093	-0.125	0.091	-0.013	0.006
50:50	**0.6**	-0.114	0.109	-0.120	0.104	-0.015	0.003
	0.7	-0.109	0.100	-0.110	0.090	-0.015	0.007
	0.8	-0.097	0.085	-0.094	0.078	-0.013	0.004
	0.9	-0.075	0.061	-0.078	0.060	-0.008	0.004
100:100	**0.6**	-0.080	0.078	-0.086	0.074	-0.011	0.001
	0.7	-0.076	0.071	-0.077	0.062	-0.011	0.005
	0.8	-0.066	0.061	-0.065	0.055	-0.010	0.002
	0.9	-0.050	0.044	-0.053	0.045	-0.006	0.004

Table B.6 95% equal-tail range for sampling distribution of individual differences between binormal and empirical estimates of AUCs (supplement to table 4.7)

Improperness group		**95% Range of (Binormal - Empirical)**					
	AUC	**20/20**		**50/50**		**100/100**	
V P	**0.6**	-0.027	0.016	-0.017	0.006	-0.013	0.003
	0.7	-0.032	0.009	-0.021	0.001	-0.017	-0.004
	0.8	-0.032	0.008	--	--	--	--
	0.9	-0.012	0.070	-0.007	0.009	-0.005	0.009
	All	***-0.028***	***0.015***	***-0.017***	***0.008***	***-0.013***	***0.005***
S & b<1	**0.6**	-0.024	0.008	-0.014	0.004	-0.012	0.002
	0.7	-0.027	0.002	-0.020	-0.001	-0.014	-0.003
	0.8	-0.025	0.004	-0.019	0.003	-0.013	0.002
	0.9	-0.014	0.005	-0.009	0.005	-0.006	0.004
	All	***-0.021***	***0.005***	***-0.012***	***0.005***	***-0.009***	***0.004***
S & b>1	**0.6**	-0.034	0.024	-0.028	0.007	-0.015	-0.004
	0.7	-0.044	0.006	--	--	--	--
	0.8	--	--	--	--	--	--
	0.9	--	--	--	--	--	--
	All	***-0.034***	***0.023***	***-0.028***	***0.007***	***-0.015***	***-0.004***
N & b<1	**0.6**	-0.019	0.010	-0.012	0.002	-0.010	0.000
	0.7	-0.020	0.012	-0.014	0.007	-0.011	0.005
	0.8	-0.017	0.009	-0.013	0.004	-0.010	0.002
	0.9	-0.011	0.002	-0.008	0.002	-0.006	0.002
	All	***-0.018***	***0.010***	***-0.013***	***0.005***	***-0.010***	***0.003***
N & b>1	**0.6**	-0.040	0.020	-0.027	0.008	-0.022	0.010
	0.7	-0.052	0.007	*-0.013*	-0.013	--	--
	0.8	--	--	--	--	--	--
	0.9	--	--	--	--	--	--
	All	***-0.043***	***0.020***	***-0.027***	***0.008***	***-0.022***	***0.010***

BIBLIOGRAPHY

1. Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 1978; 8: 283-298.
2. Metz CE. ROC methodology in radiologic imaging. *Investigative Radiology.* 1986; 21: 720-733.
3. Hanley JA. The robustness of the "Binormal" assumptions used in fitting ROC curves. *Medical Decision Making.* 1988; 8: 197-203.
4. Pepe MS. Statistical evaluation of medical tests for classification and prediction. *Oxford University Press: Oxford.* 2003.
5. Pepe MS. An interpretation for the ROC curve and inference using GLM procedures. *Biometrics.* 2000; 56: 352-359.
6. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982; 143: 745-750.
7. Hsieh F, Turnbull BW. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics.* 1996; 24: 25-40.
8. Walsh SJ. Limitations to the robustness of binormal ROC curves: effects of model misspecification and location of decision thresholds on bias, precision, size and power. *Statistics in Medicine.*1997; 16: 669-679.
9. Dorfman DD, Berbaum KS, Metz CE, et al. Proper receiver operating characteristic analysis: the bigamma model. *Academic Radiology.* 1997; 4: 138-149.
10. Grzybowski M, Younger JG. Statistical methodology: III. Receiver operating characteristic (ROC) curves. *Academic Emergency Medicine.* 1997; 4: 818-826.
11. Pan X, Metz CE. The "proper" binormal model: parametric receiver operating characteristic curve estimation with degenerate data. *Academic Radiology.*1997; 4: 380-389.
12. Metz CE, Pan X. "Proper" binormal ROC curves: theory and maximum-likelihood estimation. *Journal of Mathematical Psychology.* 1999; 43: 1-33.

13. Herron JM, Bender TM, Campbell WL, et al. Effects of luminance and resolution on observer performance with chest rediographs. *Radiology*. 2000; 4: 1.

14. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data: Part II. A formal model. *Academic Radiology.* 2000; 7: 427-437.

15. Dorfman DD, Berbaum KS. A contaminated binormal model for ROC data: Part III. Initial evaluation with detection ROC data. *Academic Radiology.* 2000; 7: 438-447.

16. Zou KH, Hall WJ. Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics.*2000; 27: 621-631.

17. Hall WJ, Zou KH. Semiparametric and parametric transformation models for comparing diagnostic markers with paired design. *Journal of Applied Statistics.* 2002; 29: 803-816.

18. Zhou XH, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. *Wiley.* 2011.

19. Pesce LL, Metz CE. Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. *Academic Radiology.* 2007; 14: 814-829.

20. 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.

21. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988; 240: 1285-1293.

22. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44: 837-845.

23. Constantine K, Karson M, Tse SK. Estimation of P(Y<X) in the gamma case. *Communications in Statistics - Simulation and Computation.* 1986; 15: 365-388.

24. Hajian-Tilaki K. Methodologic contributions to ROC analysis: a study of the robustness of the binormal model for quantitative data and methods for studies involving multiple signals. *McGill University (Canada).* 1995.

25. Walsh SJ. Diagnostic accuracy and deep venous thrombosis: a biostatistician's perspective. *Journal of Nuclear Medicine*. 1991; 32: 2328-2331.

26. Wagner RF, Metz CE, Campbell G. Assessment of medical imaging systems and computer aids: a tutorial review. *Academic Radiology*. 2007; 14: 723-748.

27. Hillis SL. A simplifying reformulation of the binormal likelihood-ratio model. *ENAR*. 2014.

28. Egan JP. Signal detection theory and ROC analysis. *Academic Press, New York.* 1975.

29. Gonen M. Analyzing receiver operating characteristic curves with SAS. *SAS Institute*. 2007.

30. Ma H, Bandos AI, Gur D. On use of partial area under the ROC curve for comparison of two diagnostic tests. *Biometrical Journal (in press).* 2014 or 2015.