

Inference about binocular sensitivity and specificity of screening tests for paired organs

Tsung-Shan Tsou¹, Wei-Cheng Hsiao²

Abstract

Recently Perera, *et. al.*¹ introduced two new binocular accuracy measures to evaluate diagnostic tests for paired organs. They adopted the Gaussian copula model to account for correlation between fellow eyes. As the measures are functions of several joint probabilities and due to the nature of the joint models, variations of the estimates for the two new measures were assessed via bootstrapping.

We provide a different approach to inference about the two interesting and innovative measures. In our opinion, when patients are independent, the binomial models suffice for inference about the parameters of interest. Inference becomes simple and straightforward. We perform numerical studies and analyze the data set as of Perera, *et. al.*¹ for illustration. Also, we investigate thru simulations the issue of robustness of the GC and the binomial models under model misspecification.

Keywords: Ophthalmology; Pairing; Binocular sensitivity; Binocular specificity; Gaussian copula

¹Institute of Statistics, National Central University, Taiwan

Corresponding author: Tsung-Shan Tsou, **Email:** chopinmozart0422@gmail.com

²Department of Financial Engineering and Actuarial Mathematics, Soochow University, Taiwan

Email: hsiaowc@scu.edu.tw

1 Introduction

Sensitivity and specificity are two typical accuracy measures for evaluating how well diagnostic procedures provide correct diagnosis for diseases. Sensitivity is the probability that the diagnostic test signals the presence of disease when the gold standard shows that disease is indeed present.

Specificity is the probability that the test gives correct disease free diagnosis as the gold standard.

Medical research in Ophthalmology very often takes observations from fellow eyes. The value of a medical device for the Ophthalmological diseases diagnosis is usually gauged via pairs of eyes. Paired organs are valuable homogeneous units for biomedical researches with which one could draw conclusions less contaminated by the between-subject heterogeneity. Nevertheless, pairing introduces correlation between fellow organs. One would have to deal with correlation parametrically in order to draw likelihood inference.

In Ophthalmology studies, de Leon *et al.*² proposed using the common correlation model to cope with the correlation between fellow eyes. They incorporated random effects that follow beta distributions to induce correlations. As a result, three correlation coefficients emerge, one for the two diagnostic responses, one for the two disease statuses and one for the diagnoses and disease statuses. Numerical integration is then required to derive the marginal bivariate distribution. Other schemes for modeling correlation can be found in de Leon *et al.*²

In order to jointly evaluate the effectiveness of the diagnostic procedures in such binocular settings, Perera, *et. al.*¹ proposed two new accuracy measures, the binocular sensitivity, *bsen*, and the binocular specificity, *bsp*. The former is the probability that the test gives at least one diseased organ correct prediction, given that at least one organ is truly diseased. In other words,

$$bsen = P\left(\begin{array}{c|c} \text{at least one organ is diagnosed} & \text{at least one organ is} \\ \text{correctly as diseased} & \text{truly diseased} \end{array}\right).$$

While, *bsp* is the probability of correct diagnosis for both non-diseased organs, given that both organs are indeed disease free, namely,

$$bsp = P\left(\begin{array}{c|c} \text{both organs are diagnosed} & \text{both organs are} \\ \text{correctly as non-diseased} & \text{truly non-diseased} \end{array}\right).$$

In order to account for correlation between fellow organs, they used the extended common correlation model (ECCM) as the joint model in the case of exchangeability, where two organs are indistinguishable. For non-exchangeable organs, the Gaussian copula (GC) is entertained to describe the joint probabilities. Inference using ECCM or Gaussian copula requires less pleasant computation tasks, such as numerical integration and bootstrapping.

Here, we provide another viewpoint about how to make inference about the parameters of interest. Our view is that, basically, correlation/joint probabilities are not directly relevant for inference about the two new measures. This suggests that one might not need to deal with the joint distribution whose intricate feature attributes only to the correlation. This may sound dubious, as b_{sen} and b_{sp} are functions of joint probabilities that depend on correlation. However, bear in mind that the units for estimation of b_{sen} and b_{sp} are “patients”, instead of “eye/eyes”. Consider estimating b_{sen} for exemplification. By definition, b_{sen} is a conditional probability with the conditioning event being the population that includes patients with at least one diseased eye, and the numerator of b_{sen} concerns the patients who have at least one diseased eye that is also correctly diagnosed. This clearly reveals that the estimation of b_{sen} uses the numbers of patients, not the numbers of eyes. The same applies to the estimation of b_{sp} as well.

The fact is that joint probabilities govern how marginal probabilities are distributed into combinations of responses from fellow eyes. Yet, it is the numbers of eligible patients contributing to the estimation of the new measures. Therefore, our view is that correlations or joint probabilities play no or lesser roles in inference for b_{sen} and b_{sp} .

In this paper, we intend to clarify the facts just alleged and propose that, if patients are independent, two separate simple binomial models will suffice for the inference about the two binocular accuracy measures. In so doing, one can apply familiar statistical procedures for inference without resorting to models for which one would have to undergo numerical integration and bootstrapping.

2 Binocular sensitivity and binocular specificity

We distinguish the responses from the left and the right eyes with subscripts L (left) and R (right) and

let $Y_{iL} = l_1$ and $Y_{iR} = r_1, l_1, r_1 = 0, 1$ denote the diagnosis results (0: negative, 1: positive). Denote by $\delta_{iL} = l_2$ and $\delta_{iR} = r_2, l_2, r_2 = 0, 1$, the true disease statuses (0: absence, 1: presence) of the left and the right eyes. We use $n_{l_1 r_1 l_2 r_2}$ to denote the number of patients with diagnoses l_1, r_1 and disease statuses l_2, r_2 on fellow eyes, respectively. Similarly, $p_{l_1 r_1 l_2 r_2}$ denotes the joint probability for a patient to have responses l_1, r_1, l_2 and r_2 at the corresponding eyes. The sensitivities of the left and right eyes are $sen_L = p_{1+1+} / p_{++1+}$ and $sen_R = p_{+1+1} / p_{++1+}$, respectively; specificities of the two eyes are $sp_L = p_{0+0+} / p_{++0+}$ and $sp_R = p_{+0+0} / p_{++0+}$, where “+” indicates summing over the subscript that it replaces. The marginal characteristics are denoted by $\pi_{1L} = P(Y_{iL} = 1), \pi_{1R} = P(Y_{iR} = 1), \pi_{2L} = P(\delta_{iL} = 1)$ and $\pi_{2R} = P(\delta_{iR} = 1)$.

Recall that $p_{l_1 r_1 l_2 r_2} = P(Y_{iL} = l_1, Y_{iR} = r_1, \delta_{iL} = l_2, \delta_{iR} = r_2), l_1, r_1, l_2, r_2 = 0, 1$. Hence,

$$bsen = (p_{1111} + p_{1011} + p_{0111} + p_{1010} + p_{0101} + p_{1110} + p_{1101}) / (p_{++11} + p_{++10} + p_{++01}).$$

Notice that $p_{1011} + p_{1010} + p_{1110} = p_{1+1+} - p_{1111}$ and $p_{0111} + p_{0101} + p_{1101} = p_{+1+1} - p_{1111}$. Therefore,

$$bsen = \frac{p_{1111} + p_{1+1+} - p_{1111} + p_{+1+1} - p_{1111}}{p_{++11} + p_{++10} + p_{++01}} = \frac{p_{1+1+} + p_{+1+1} - p_{1111}}{p_{++11} + p_{++10} + p_{++01}}. \quad (1)$$

Since $sen_L = P(Y_{iL} = 1 | \delta_{iL} = 1) = p_{1+1+} / p_{++1+}$ and $sen_R = P(Y_{iR} = 1 | \delta_{iR} = 1) = p_{+1+1} / p_{++1+}$,

$$bsen = \frac{sen_L p_{++1+} + sen_R p_{++1+} - p_{1111}}{p_{++11} + p_{++10} + p_{++01}}.$$

Notice that p_{1111} is to be subtracted out of $p_{1+1+} + p_{+1+1}$ in the numerator. But, equations (6) and (8) of Perera, *et. al.*¹ have a plus sign in front of p_{1111} that might not be correct. The binocular specificity is simply p_{0000} / p_{++00} .

Perera, *et. al.*¹ studied the properties of the two new binocular measures under the cases of exchangeability, namely, $p_{l_1 r_1 l_2 r_2} = p_{r_1 l_1 r_2 l_2}$ and non-exchangeability. For the former, they adopted the

ECCM that was espoused in de Leon *et. al.*² and formed estimates

$$\widehat{bsen} = (\widehat{p}_{1111} + \widehat{p}_{1011} + \widehat{p}_{0111} + \widehat{p}_{1010} + \widehat{p}_{0101} + \widehat{p}_{1110} + \widehat{p}_{1101}) / (\widehat{p}_{++11} + \widehat{p}_{++10} + \widehat{p}_{++01})$$

and

$$\widehat{bsp} = \widehat{p}_{0000} / \widehat{p}_{++00}$$

by plugging in the moment estimates $\hat{p}_{l_1 r_1 l_2 r_2}$. The variabilities of the two estimates were then evaluated by resorting to bootstrapping.

For the case of non-exchangeability Perera, *et. al.*¹ used Gaussian copula (GC) as the model and derived the ML estimates for $p_{l_1 r_1 l_2 r_2}$. The plug-in estimates for the binocular measures were correspondingly formed and their variations similarly assessed. They found that ECCM and Gaussian copula provide satisfactory estimates for the binocular measures in terms of bias and efficiency in the exchangeable and non-exchangeable scenarios, respectively.

Notice that de Leon *et al.*² employed moment estimates, namely, the sample proportions as the estimates for the joint probabilities (page 840, de Leon *et al.*²) The resulting plug-in estimates \widehat{bsen} and \widehat{bsp} are simply sample proportions that are the ML estimates based on our proposed binomial models (see Appendix). It is also worthy of mentioning that the correlation estimates under ECCM are meaningful or legitimate only if the distributional assumptions underlying the adopted joint model, such as the beta assumption, are correct.

3 Binomial models for $bsen$ and bsp

Consider the binomial model for inference about $bsen$

$$\prod_{i=1}^N \{bsen^{z_i} (1-bsen)^{1-z_i}\}^{\xi_i} = bsen^{\sum_{i=1}^N z_i \xi_i} (1-bsen)^{\sum_{i=1}^N (1-z_i) \xi_i}, i = 1, \dots, N,$$

where $\xi_i = 1$ if the i th patient has at least one diseased eye, and $z_i = 1$ if the i th patient has at least one diseased eye correctly diagnosed. The score function for $bsen$ is

$$\frac{\sum_{i=1}^N z_i \xi_i}{bsen} - \frac{\sum_{i=1}^N \xi_i - \sum_{i=1}^N z_i \xi_i}{1-bsen}$$

that gives rise to $\widehat{bsen} = \sum_{i=1}^N z_i \xi_i / \sum_{i=1}^N \xi_i = m_{bsen} / m_1$ as the ML estimate for $bsen$, where

$m_1 = N - n_{++00}$ and $m_{bsen} = m_1 - (n_{1001} + n_{0001} + n_{0110} + n_{0010} + n_{0011})$. It is clear that the number of patients

who contribute to inference about $bsen$ is $\sum_{i=1}^N \xi_i = N - n_{++00} = m_1$. Note that m_1 is the number of

patients with at least one diseased eye and m_{bsen} is the number of patients, out of m_1 , having at least

one diseased eye correctly diagnosed. By taking expectation of minus the derivative of the score

function, one obtains the Fisher information

$$-E \left\{ \frac{\sum_{i=1}^N z_i \xi_i}{bsen^2} + \frac{\sum_{i=1}^N \xi_i - \sum_{i=1}^N z_i \xi_i}{(1-bsen)^2} \right\}.$$

Now $E(z_i \xi_i) = E(z_i | \xi_i = 1)P(\xi_i = 1) + E(z_i | \xi_i = 0)P(\xi_i = 0)$. Note that $E(z_i | \xi_i = 1) = bsen$ and $E(z_i | \xi_i = 0) = 0$, so that $E(z_i \xi_i) = bsenP(\xi_i = 1)$, where $P(\xi_i = 1) = E(\xi_i)$. Hence, the Fisher

Information equals

$$\frac{NbsenP(\xi_i = 1)}{bsen^2} + \frac{N(1-bsen)P(\xi_i = 1)}{(1-bsen)^2} = \frac{NP(\xi_i = 1)}{bsen(1-bsen)}.$$

Note that $P(\xi_i = 1)$ is the probability that a patient has at least one diseased eye that is estimable

by $\widehat{P}(\xi_i = 1) = m_1 / N$. Consequently, one has available the variance estimate for \widehat{bsen} ,

$\widehat{bsen}(1 - \widehat{bsen}) / \{\widehat{P}(\xi_i = 1)N\}$, where $\widehat{P}(\xi_i = 1)N$ is simply m_1 . Now, it is conceivable that the above binomial model is equivalent to $bsen^{m_{bsen}}(1-bsen)^{m_1-m_{bsen}}$.

By the same token, let $m_2 = n_{++00}$ and $m_{bsp} = n_{0000}$ then the model for inference about bsp is simply $bsp^{m_{bsp}}(1-bsp)^{(m_2-m_{bsp})}$. If patients are independent, the binomial models are legitimate.

Therefore, $\widehat{bsp} = n_{0000} / n_{++00}$ and its variance estimate is $\widehat{bsp}(1 - \widehat{bsp}) / n_{++00}$.

Evidently, the binomial models apply to both the cases of exchangeability and non-exchangeability as well. Again, exchangeable or not, eventually, it is the patients that contribute to the information for inference about the two binocular accuracy measures. The joint models, such as the ECCM and the Gaussian copula, prescribe how fellow eyes are correlated. In the end, when it comes to inference about $bsen$ and bsp , only the numbers of eligible patients are included for estimation and inference.

Here, we take a detour to comment on the contrast between our binomial model and the so-called marginal model. A marginal model might be one that incorporates the correlation (for example by integrating out the random effects common to a cluster) or ignores the correlation. Anyhow, the idea underlying the marginal approach is that any one of the correlated responses contributes to inference about the parameter of interest. This, however, is not the case we are

dealing with. Statuses of two eyes of a patient form the unit for information about $bsen$ and bsp .

Observations from one single eye do not provide information for the parameter of interest. Hence, our binomial model is not a marginal model in the sense commonly perceived.

4 Simulations

Obviously, if one adopts binomial models inference for the binocular measures becomes that about the success probabilities with independent and identically distributed Bernoulli responses. It becomes a straightforward and easy procedure and bootstrapping is definitely not needed.

We first intended to reconstruct Table II of Perera, *et. al.*¹ by using exactly the same parameter setting including the same values of $\pi_1, \pi_2, \rho, \rho_1, \rho_2$ and their joint model (11). Here,

$\pi_1 = p_{1+++}, \pi_2 = p_{++1+}, \rho$ is the correlation between $Y_{iL} + Y_{iR}$ and $\delta_{iL} + \delta_{iR}$, ρ_1 is the correlation between Y_{iL} and Y_{iR} , and ρ_2 is the correlation between δ_{iL} and δ_{iR} . It turns out that we are able to derive exactly the same sen and sp as they did, except the two parameters of interest $bsen$ and bsp . As a reminder, we calculate $bsen$ according to our formula (1). Recall that sen and sp are the common sensitivity and common specificity under exchangeability. Table 1 shows the discrepancy between the true $bsen$ and bsp values we derived and those in Table II of Perera, *et. al.*¹

<Table 1>

Next, data conforming ECCM with the above parameter settings are simulated. Two thousand simulation runs with sample sizes $N = 100$ and 200 are executed. Empirical statistics resulting from the binomial models are tabulated in Tables 2.1 to 2.3. The relative bias (RB) is defined as

$RB = \{\text{average}(\hat{\theta}) - \theta\} / \theta$, where θ is the true value of the parameter of interest. Unlike that

calculated via bootstrapping in Perera, *et. al.*¹, our exhibited $SE(\hat{\theta})$ is $\{\hat{\theta}(1 - \hat{\theta}) / N^*\}^{1/2}$ that is easily

calculable from binomial, where N^* is either $N - n_{++00}$ or n_{++00} as explained earlier. We also display the

sample variance of all $\hat{\theta}$ from simulation repeats, denoted by $SD(\hat{\theta})$. The empirical 95% confidence

interval is constructed and CI (the average lower and upper bounds), AW (the average width) and CP

(the empirical coverage probability) are tabulated. **Recall that estimates from both ECCM and binomial**

model are identical; hence, results from ECCM are not reported.

<Tables 2.1-2.3>

Tables 3.1 and 3.2 show results of our analysis for the nonexchangeable data generated from the GC model. The following three different parameter settings are considered:

Case I: $(\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}) = (0.4, 0.36, 0.32, 0.36), (\delta_1, \delta_2, \delta_3, \delta_4) = (0.65, 0.37, 0.28, 0.17)$;

Case II: $(\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}) = (0.35, 0.44, 0.33, 0.38), (\delta_1, \delta_2, \delta_3, \delta_4) = (0.7, 0.5, 0.15, 0.2)$ and

Case III: $(\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}) = (0.35, 0.44, 0.27, 0.38), (\delta_1, \delta_2, \delta_3, \delta_4) = (0.7, 0.52, 0.12, 0.21)$.

Note that $\delta_1, \delta_2, \delta_3$ and δ_4 are elements of the correlation matrix D of the GC model (Perera, *et. al.*¹) In order to implement the GC model, one thousand bootstrapping for each of the one thousand simulation repeats are carried out.

<Tables 3.1-3.2>

The contents of Table 3.1 indicate that the ML estimate of b_{sen} from the binomial model, a sample proportion, in general, has a slightly larger bias compared to the ML estimate from the GC model. It appears that the standard error/standard deviation estimate of $\widehat{b_{sen}}$ from GC is smaller. This is anticipated since the data are generated from the GC model. The ideal situation warrants higher efficiency for inference from the method of maximum likelihood. In light of the empirical coverage probability of the confidence intervals from the two models, the binomial approach is a legitimate alternative for inference for b_{sen} . Bear in mind, however, that the statistics are straightforward and easily obtainable from binomial. One needs to perform numerical maximization of the GC likelihood and further bootstrapping for variation assessment. Findings for b_{sp} from Table 3.2 are similar.

Contrasts between the GC and the binomial models in situations where the nonexchangeable data do not conform to the GC assumption are also provided. We generate the paired data from the t copula (TC) with $\nu = 4$ degrees of freedom. We use the same notation as of Perera, *et. al.*¹

letting $Y_{iL} \sim \text{Bernoulli}(\pi_{1L}), \delta_{iL} \sim \text{Bernoulli}(\pi_{2L}), Y_{iR} \sim \text{Bernoulli}(\pi_{1R})$ and $\delta_{iR} \sim \text{Bernoulli}(\pi_{2R})$. The t copula probit model (Demarta and McNell⁵) gives rise to

$$P_{l_1 r_1 l_2 r_2} = \sum_{j_1=0}^1 \sum_{j_2=0}^1 \sum_{j_3=0}^1 \sum_{j_4=0}^1 (-1)^{4+\sum_{h=1}^4 j_h} F_{\Sigma, \nu} \left(\begin{array}{c} T_{\nu}^{-1}(u_{1Lj_1}), T_{\nu}^{-1}(u_{1Rj_2}), \\ T_{\nu}^{-1}(u_{2Lj_3}), T_{\nu}^{-1}(u_{2Rj_4}) \end{array} \right).$$

Here $F_{\Sigma, \nu}(\cdot)$ is the multivariate t cumulative distribution function, Σ is the correlation matrix and

$T_{\nu}^{-1}(\cdot)$ denotes the quantile function of the univariate t distribution, where ν is the degrees of freedom.

One can consult Kotz *et. al.*⁶ for the properties of the multivariate t distribution. Definitions of

u_{mkj_h} , $m = 1, 2$, $k = L, R$, $j_h = 0, 1$ and $h = 1, 2, 3, 4$, the cumulative probabilities of the diagnostic results

and the disease statuses of the fellow organs can be found in Perera, *et. al.*¹

We sample the TC data with three settings that incorporate the same $\pi_{1L}, \pi_{1R}, \pi_{2L}$ and π_{2R} values as of Tables 3.1 and 3.2 but with nonzero values δ_5 and δ_6 as the (1,4)th and the (2,3)th entries of the correlation matrix Σ . In so doing, one might have a glimpse of the impact of the misspecification of D of the GC model. More specifically, the settings of Σ include

Case I: $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (0.65, 0.37, 0.28, 0.17, 0.65, 0.37)$;

Case II: $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (0.7, 0.5, 0.15, 0.2, 0.7, -0.3)$ and

Case III: $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6) = (0.7, 0.52, 0.12, 0.21, 0.7, -0.3)$.

<Tables 4.1-4.2>

Tables 4.1 and 4.2 show results of the binomial and the GC models. Notice that the GC model employs ML estimates under the assumption that δ_5 and δ_6 are both zero. Table 4.1 suggests that the ML estimate of b_{sen} from the binomial model has a much smaller bias than the ML estimate from the GC model. As a reminder, the GC likelihood is now no longer valid for the data. Still, $\widehat{b_{sen}}$ from GC is less variable compared to that from binomial. However, the empirical coverage probability of the confidence interval from GC is way below the 95% nominal level. Obviously, inference for b_{sen} drawn from the GC model is mistaken if the GC assumption is false. Similarly, when the GC assumption fails, deviation of inference for b_{sp} from the required nominal standard is also clearly manifested in Table 4.2. In contrast, our proposed binomial model is able to deliver satisfactory statistics for inference for the binocular measures. The advantage of the simple binomial model being

robust against the specification of the underlying joint distribution is quite evident.

5 Examples

We use as illustrative examples the data analyzed by Tsou⁴ for assessing the value of the screening device high-resolution stereoscopic digital photography. Those diagnosed as “positive” with clinical conditions could be eligible for early treatment of diabetic Retinopathy. The presence/absence of the abnormal conditions diagnosed by the contact len biomicroscopy (CLBM) is taken as the gold standard. Hence, the outcome of CLBM is labeled as the disease status. We analyze the data on 104 patients with complete information on the two conditions macular oedema and hard exudate for illustration. Detailed about the study could be found in Rudnisky, *et al.*³

Example 1

Table 5 contains the macular oedema data and Table 6 exhibits estimates and their standard errors $\{\widehat{bsen}(1 - \widehat{bsen}) / m_1\}^{1/2}$ (SEs) using the binomial model. Note that macular oedema pertains to the thickening and swelling of the eye’s macula due to fluid and protein deposits. In addition to the statistics for the binocular parameters, we also present estimates and their standard errors for $\pi_{1L}, \pi_{1R}, \pi_{2L}, \pi_{2R}, sen_L, sen_R, sp_L$ and sp_R . The statistics for all parameters are ML estimates based on the binomial model with the success probability, such as p_{1+++} for π_{1L} and p_{+1++} / p_{++++} for sen_R , respectively. Hence, the estimate of π_{1L} is n_{1+++} / N and that of sen_R equals n_{+1++} / n_{++++} , and their standard errors are, respectively, $\{n_{1+++}(N - n_{1+++}) / N^3\}^{1/2}$ and $\{n_{+1+}(n_{++++} - n_{+1+}) / n_{++++}^3\}^{1/2}$.

<Table 5>

<Table 6>

Example 2

The hard exudate data are displayed in Table 7 and results are also provided in Table 6. Hard exudate involves the leakage of fluid and lipoprotein into the retina of the eyes.

<Table 7>

As a reminder, one can easily construct confidence intervals and perform hypothesis testing using the binomial likelihood functions for $bsen$ and bsp , respectively. For likelihood inference about

sensitivity and specificity in paired scenario, readers are referred to Tsou.⁴

6. Conclusion

Pairing introduces correlation that makes likelihood inference more challenging. However, there are situations where a full probability structure that incorporates all joint probabilities might not be necessary. Inference for the binocular measures, in our opinion, is a scenario that a full joint distribution is not required.

Our theory is founded upon the fact that correlation might affect the tendency of joint occurrences of diseases/diagnostics of fellow organs. Nevertheless, when it comes to estimations of the two parameters, it is the numbers of patients, not the counts of eyes, which contribute to inference. When patients are independent, binomial models suffice for inference about $bsen$ and bsp .

The applicability of our simpler binomial approaches is, certainly, confined to inference about $bsen$ and bsp . The joint distributions such as ECCM and the Gaussian copula provide a full description of the joint probability structure of the data. The full models enable one to draw inference for a variety of scientific questions of interest in addition to $bsen$ and bsp that are the sole targets for the binomial models. Obviously, the effective sample sizes, such as $N - n_{++00}$ and n_{++00} that we employ need to be large to the extent in accord with the binomial model.

Acknowledgements

This work is supported by grant MOST106-2118-M-008-003-MY2 of Ministry of Science and Technology, Taiwan, R.O.C.

References

1. Perera Y, Ren M, Punzalan J, Rudnisky C, de Leon A. Binocular sensitivity and specificity of screening tests in cross-sectional diagnostic studies of paired organs. *Statistics in Medicine* 2017; **36**: 1754-1766.
2. de Leon AR, Soo A, Bonzo D, Rudnisky CJ. Joint estimation of diagnostic accuracy measures for paired organs - application in ophthalmology. *Biometrical Journal* 2009; **51**:837–850.
3. Rudnisky CJ, Hinz BJ, Tennant MTS, de Leon AR, Greve MDJ. High-resolution stereoscopic digital fundus photography versus contact-lens biomicroscopy for the detection of clinically

significant macular edema. *Ophthalmology* 2002; **109**: 267–274.

4. Tsou TS. Robust likelihood inference for diagnostic accuracy measures for paired organs.

Statistical Methods in Medical Research 2018; Online First: DOI: 10.1177/0962280218798170

5. Demarta S, McNell AJ. The t copula and related copulas. *International Statistical Review* 2005; **73**: 111-129.

6. Kotz S, Balakrishnan N, Johnson N. *Continuous multivariate distributions* 2000; New York: Wiley.

Appendix

In the paper of de Leon *et. al.*,² notation n_{xy} , $x, y = 0, 1, 2$ denotes the number of patients having x eyes diagnosed as positive and y diseased eyes. Note that $\sum_0^2 \sum_0^2 n_{xy} = N$. Also, they defined

$p_{xy} = p(Y_{iL} + Y_{iR} = x, \delta_{iL} + \delta_{iR} = y)$. It is easy to establish the relationship between their p_{xy} and

$p_{l_1 r_1 l_2 r_2}$: $p_{00} = p_{0000}$, $p_{10} = p_{1000} + p_{0100}$, $p_{20} = p_{1100}$, $p_{01} = p_{0010} + p_{0001}$, $p_{11} = p_{1010} + p_{1001} + p_{0110} + p_{0101}$,

$p_{21} = p_{1110} + p_{1101}$, $p_{02} = p_{0011}$, $p_{12} = p_{0111} + p_{1011}$ and $p_{22} = p_{1111}$. By assuming exchangeability, namely,

$p_{l_1 r_1 l_2 r_2} = p_{r_1 l_1 r_2 l_2}$, we have $p_{01} = 2p_{0010} = 2p_{0001}$, $p_{11} = 4p_{1010} = 4p_{1001} = 4p_{0110} = 4p_{0101}$,

$p_{10} = 2p_{1000} = 2p_{0100}$, $p_{21} = 2p_{1110} = 2p_{1101}$ and $p_{12} = 2p_{0111} = 2p_{1011}$.

The moment estimates for p_{xy} are sample proportions $\hat{p}_{xy} = n_{xy} / N$, $x, y = 0, 1, 2$. These estimates were then used to get estimates for the three correlation estimates, see p. 840 of de Leon *et al.*² One can easily establish the correspondence between the two sets of moment estimates \hat{p}_{xy} and $\hat{p}_{l_1 r_1 l_2 r_2}$. For example, $\hat{p}_{01} + \hat{p}_{11} + \hat{p}_{21} = \hat{p}_{++10} + \hat{p}_{++01} = 2\hat{p}_{++10}$, $\hat{p}_{02} + \hat{p}_{12} + \hat{p}_{22} = \hat{p}_{++11}$ and $\hat{p}_{22} = \hat{p}_{1111}$. This confirms the fact that \widehat{bsen} and \widehat{bsp} depend on sample proportions only. The three correlation estimates de Leon *et al.*² derived by plug-in estimates \hat{p}_{xy} are irrelevant for inference about $bsen$ and bsp .