Published in final edited form as:

J Biopharm Stat. 2016 ; 26(6): 1111–1117. doi:10.1080/10543406.2016.1226330.

Estimation of an accuracy index of a diagnostic biomarker when the reference biomarker is continuous and measured with error

Mixia Wu^a, Dianchen Zhang^b, and Aiyi Liu^b

^aCollege of Applied Sciences, Beijing University of Technology, Beijing, China

^bBiostatistics and Bioinformatics Branch, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland, USA

Abstract

New biomarkers continue to be developed for the purpose of diagnosis, and their diagnostic performances are typically compared with an existing reference biomarker used for the same purpose. Considerable amounts of research have focused on receiver operating characteristic curves analysis when the reference biomarker is dichotomous. In the situation where the reference biomarker is measured on a continuous scale and dichotomization is not practically appealing, an index was proposed in the literature to measure the accuracy of a continuous biomarker, which is essentially a linear function of the popular Kendall's tau. We consider the issue of estimating such an accuracy index when the continuous reference biomarker is measured with errors. We first investigate the impact of measurement errors on the accuracy index, and then propose methods to correct for the bias due to measurement errors. Simulation results show the effectiveness of the proposed estimator in reducing biases. The methods are exemplified with hemoglobin A1c measurements obtained from both the central lab and a local lab to evaluate the accuracy of the mean data obtained from the metered blood glucose monitoring against the centrally measured hemoglobin A1c from a behavioral intervention study for families of youth with type 1 diabetes.

Keywords

Bias; diagnostic accuracy; Kendall's tau; measurement error; reliability coefficient

1. Introduction

Diagnostic biomarkers are used to identify subjects with a health condition, such as a disease or high risk of developing the disease. Prostate-specific antigen (PSA) and hemoglobin A1c are two good examples that have been used for the diagnosis of prostate cancer and diabetes, respectively. However, even the most well-known and commonly used biomarkers are often measured with errors that, if not properly accounted for, may incur substantial bias in evaluating the performance of the biomarkers.

The performance of a continuous biomarker against a binary reference biomarker has received enormous attention, resulting in a large body of work in the literature that are well summarized in a few books (Pepe, 2003; Krzanowski and Hand, 2009; Zhou et al., 2011; Zou et al., 2011). In this setting the biomarker's performance is usually assessed with the aid of its receiver operating characteristic (ROC) curves, which is obtained by plotting the biomarker's sensitivity versus 1 minus its specificity as its specificities range from 0 to 1. When comparing two biomarkers, a higher curve in a range of specificity indicates better performance of the biomarker in that range. Often, the area and partial area under the ROC curve are used to compare the performance of two biomarkers.

A reference biomarker used for diagnosis is often measured on a continuous scale, and frequently encountered in practice is that dichotomization may be inappropriate. Furthermore, when lacking scientifically established threshold, dichotomization may dilute the measure of accuracy and can mislead the true relationship between the reference biomarker and diagnostic biomarker (Obuchowski, 2005, 2006). Obuchowski (2006) illustrated some problems when treating a continuous-scale gold standard as if it were binary.

Unlike a binary reference biomarker, however, evaluation of the diagnostic accuracy of a continuous biomarker against a continuous reference biomarker has received less attention in the literature. Obuchowski (2005, 2006) suggested an index measure of accuracy, which is essentially a linear function of the popular Kendall's (1938) tau, and proposed a nonparametric estimator of the index.

In reality, it is quite common that the reference biomarker is measured with error, often due to the use of an imperfect instrument/technique to reduce cost or to ease the procedure. When such measurement errors are not properly accounted for, substantial bias can be introduced in estimation of the accuracy index. To the best of our knowledge, little work has been done in the literature on the impact of such measurement errors and on how to correct the errors to reduce or eliminate the biases in estimating the accuracy index. To fill the gap, we investigate the effect of measurement errors on estimation of the accuracy index and propose methods to account for the measurement errors in estimating the accuracy index. Simulation studies show that the proposed error-corrected estimator substantially reduces the biases due to measurement errors. The methods are exemplified with hemoglobin A1c measurements obtained from both the central lab and a local lab to evaluate the accuracy of the mean data obtained from the metered blood glucose monitoring as compared with the centrally measured hemoglobin A1c from a behavioral intervention study for families of youth with type 1 diabetes.

2. An accuracy measure of a biomarker against a continuous reference biomarker

Let Y be the true value of the continuous reference biomarker which may or may not be observed, and X be the value of a continuous biomarker whose diagnostic accuracy against the reference biomarker is of primary interest. Let (X_1, Y_1) and (X_2, Y_2) be two independent

copies of (X, Y) . Then the accuracy index proposed by Obuchowski (2005, 2006) is given by:

$$\eta = P((X_1 - X_2)(Y_1 - Y_2) > 0).$$

The popular Kendall's (1938) tau for X and Y , as a measure of ordinal association between X and Y , is defined as:

$$\tau = P((X_1 - X_2)(Y_1 - Y_2) > 0) - P((X_1 - X_2)(Y_1 - Y_2) < 0).$$

It follows immediately that:

$$\eta = \frac{1}{2}\tau + \frac{1}{2}.$$

Thus the accuracy index η is linear in Kendall's tau. However, the index η deserves its own merit in diagnostic medicine in that it is the probability that the biomarker and the reference biomarker fall into the same direction (simultaneously larger/smaller).

The Kendall's tau is closely related to the Pearson's correlation coefficient ρ between X and Y , and closed forms are available for certain family of distributions. Indeed, if the paired measurements (X, Y) follow a bivariate normal distribution, then $\tau = 2 \arcsin(\rho)/\pi$. This yields a closed form for η in terms of ρ :

$$\eta = \frac{1}{\pi} \arcsin(\rho) + \frac{1}{2}. \quad (1)$$

As shown in Fang et al. (2002), this functional relationship holds generally if (X, Y) follows an elliptical distribution of which the bivariate normal distribution is a special case.

3. Effects of measurement errors on accuracy

Now suppose that the reference biomarker Y is measured with error. Instead of Y , we observe Y_o , with:

$$Y_o = Y + \varepsilon,$$

where ε (with mean 0) is independent of Y (and X). The diagnostic accuracy of X against Y_o is:

$$\eta_o = P((X_1 - X_2)(Y_{o1} - Y_{o2}) > 0),$$

where (X_1, Y_{o1}) is independent of (X_2, Y_{o2}) , both random copies of (X, Y_o) . Under this simple additive measurement error model, the correlation coefficient r between Y and Y_o is

given by $\sigma_Y/\sigma_{Y_o} = \sigma_Y/\sqrt{\sigma_Y^2 + \sigma_\varepsilon^2}$ where σ stands for the standard deviation of the corresponding variable. The reliability coefficient,

$$r^2 = \frac{\sigma_Y^2}{\sigma_{Y_o}^2} = \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_\varepsilon^2},$$

is commonly used to assess the magnitude of measurement errors with larger values indicating smaller errors (1 if there is no error). The correlation coefficient ρ_o between X and Y_o is easily shown to be $r\rho$.

Note that the diagnostic accuracy η is invariant to monotone transformations of both the test and reference biomarkers. For simplicity we assume that (X, Y, Y_o) follows, possibly after certain proper transformation, a trivariate normal distribution and satisfies an additive error model. Under these assumptions, we have:

$$\eta_o = \frac{1}{\pi} \arcsin(\rho r) + \frac{1}{2}. \quad (2)$$

Clearly using Y_o instead of Y yields a different accuracy index. The resulting index η_o differs from η , referred to as the true index in contrast to η_o is mentioned, by:

$$b(\rho, r) = \frac{1}{\pi} (\arcsin(\rho r) - \arcsin(\rho)). \quad (3)$$

It is worth noting that η_o is always smaller than η unless $r=1$, that is, when there are no measurement errors. This is because under additive measurement models, Y_o has larger variation than Y . As demonstrated later, the difference between these two indexes can be substantially large.

Suppose a sample of size n , (X_i, Y_{oi}) , $i = 1, \dots, n$, from (X, Y_o) is collected. Without correcting for the measurement errors, the true accuracy index can be estimated by:

$$\hat{\eta}_o = \frac{1}{\pi} \arcsin(\hat{\rho}_o) + \frac{1}{2},$$

where $\hat{\rho}_o$ is the sample correlation coefficient. The estimator $\hat{\eta}_o$ is asymptotically unbiased for η_o , and hence biased for η . The bias is asymptotically identical to $b(\rho, r)$ since $E(\hat{\eta}_o) - \eta \approx \eta_o - \eta$. Therefore, $\hat{\eta}_o$ always underestimates η and in some cases can be badly so.

For selected values of $r = 0.55, 0.75, 0.95$, which indicate relatively large, moderate, and small measurement errors, Figure 1 shows the asymptotic bias of $\hat{\eta}_o$ in estimating η as the correlation coefficient ρ ranges from 0.35 to 1, corresponding to relatively poor accuracy ($\eta = 0.61$) to high accuracy. As the figure shows, the magnitude of the bias can be as large as

0.25, and becomes larger as the correlation (or true accuracy) gets larger. The magnitude of the bias also increases as the measurement error gets larger (or the reliability coefficient r gets smaller).

Similar findings apply to any estimators that are unbiased (or asymptotically so) for η_o but used to estimate η .

4. Bias correction in estimating the accuracy index η

The substantial bias incurred by measurement errors needs to be corrected in order to reduce/eliminate biases in estimating the accuracy index η of interest. Failure to do so may produce misleading results that would undermine the importance of the biomarker in comparison with the reference biomarker. To this end, note that from (2) it follows that:

$$\rho = \frac{1}{r} \sin(\pi(\eta_o - 0.5)).$$

Plugging into (1) then yields:

$$\eta = \frac{1}{\pi} \arcsin\left(\frac{1}{r} \sin(\pi(\eta_o - 0.5))\right) + \frac{1}{2}. \quad (4)$$

Assuming that the reliability coefficient r^2 is known or can be estimated accurately from independent source, then a bias-corrected estimator of η can be derived by substituting $\hat{\eta}_o$ for η_o in (4), leading to:

$$\hat{\eta} = \frac{1}{2} + \frac{1}{\pi} \arcsin\left(\min\left\{1, \frac{1}{r} \sin(\pi(\hat{\eta}_o - 0.5))\right\}\right).$$

This estimator of η is expected to substantially reduce the bias, even when the sample sizes are relatively small. To investigate its finite-sample biases and mean square errors, we conducted Monte Carlo studies and compared various estimators. We generated a random sample of size n from a trivariate normal distribution for (X, Y, Y_o) with zero mean vector, and variance-covariance matrix:

$$\text{cov}\begin{pmatrix} X \\ Y \\ Y_o \end{pmatrix} = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.9 & 1 & 1/\sqrt{1.5} \\ 0.9 & 1/\sqrt{1.5} & 1.5 \end{pmatrix}.$$

The parameter settings imply that $\rho = 0.9$, $r^2 = 1/1.5$, and $\rho_o = 0.9 / \sqrt{1.5}$, yielding $\eta \approx 0.86$ and $\eta_o \approx 0.76$. Thus from (3) the true (asymptotic) bias is about 0.0937. For selected values of $n = 20, 30, 50, 100, 150$, and 200 , the average biases of $\hat{\eta}$ from 10,000 simulations were found to be 0.0155, 0.0123, 0.0094, 0.0052, 0.0039, and 0.0027, respectively. In contrast, the simulated corresponding biases of $\hat{\eta}_o$ are $-0.0933, -0.0929, -0.0935, -0.0935, -0.0937$, and -0.0939 , respectively, all close to the true (asymptotic) bias. The corresponding

simulated root mean square errors of $\hat{\eta}_o / \hat{\eta}$ are 0.1073/0.0968, 0.1028/0.0822, 0.0981/0.0659, 0.0964/0.0466, 0.0964/0.0465, 0.0954/0.0368, 0.0951/0.0307, respectively. The simulation results clearly demonstrate that the proposed error-corrected estimator $\hat{\eta}$ substantially reduces both the biases and root mean square errors.

5. Applications: A real example

Children with type 1 diabetes can have serious health complications including problems with the heart, blood vessels, eyes, and kidneys. Risks of these complications can be substantially reduced by keeping the child's blood sugar level close to normal level at most of the time. Maintaining a normal blood sugar level, however, is a difficult task especially for diabetic children.

To face the challenge, a clinic-integrated behavioral intervention study was conducted by the Division of Intramural Population Health Research of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, to improve families' Type 1 diabetes management. Families randomized to the intervention group received the "WE-CAN" manage diabetes intervention at each routine clinic visit focusing on facilitating problem-solving skills, communication skills, and appropriate responsibility sharing. Families in the control group received usual care. Before each visit, research staff contacted families assigned to usual care to facilitate appointment keeping, including providing parking vouchers and acting as a liaison with clinic appointment staff. After completion of the study, families assigned to usual care also received a notebook containing the intervention educational information. The intervention was demonstrated to be effective in improving glycemic control; see Nansel et al. (2012) for details.

Glycated hemoglobin (HbA1c), a biomarker of glycemic control with lower values indicating better control, was the primary outcome of the study. Patients were recruited at study sites by research staff during routine clinic visits. Blood samples were obtained at each visit and shipped to a central laboratory for HbA1c assaying. Simultaneously blood samples were processed in local labs with DCA-2000 at each site. Blood glucose meter data were also obtained at each clinic visit, and the mean number of checks per day for the previous 14 days was calculated.

As secondary analysis of the study, we are interested in the diagnostic usefulness of the metered blood glucose monitoring (BxDof) as compared to the HbA1c measured in the central lab. To illustrate our method with the data, we treat the HbA1c level measured in the central lab as the reference biomarker and the HbA1c from on-site facilities (BxHbA1c) as the reference biomarker measured with errors. For simplicity, we confined our analysis to data halfway during the study from the control group at a selected study site. Since data for all three variables are available, the study is ideal to evaluate the effectiveness of our method in correcting bias.

Data from a total of 44 patients are used for demonstration. The sample means and variances of (HbA1c, BxHbA1c, BxDof) are (9.03, 8.56, 203.49) and (1.97, 2.29, 2225.60), respectively, and the sample correlation matrix is:

$$\begin{pmatrix} 1 & 0.93 & 0.66 \\ 0.93 & 1 & 0.57 \\ 0.67 & 0.57 & 1 \end{pmatrix}.$$

From these summary statistics, the reliability coefficient r is estimated to be 0.86, indicating moderate measurement errors. The accuracy index of BxDOF against HbA1C is estimated to be 0.729, while the accuracy index of BxDOF against BxHbA1C is found to be 0.693, a slight decrease of about 4%. The error-corrected estimator is found to be 0.725, essentially identical to the observed one.

6. Discussion

Using the multivariate normal distributions of continuous reference biomarker and biomarker as a motivating example, we showed that measurement errors can substantially bias the estimation of the accuracy index. We proposed a bias-corrected estimator and demonstrated its ability to reduce the biases. The proposed method requires the reliability coefficient be known a priori, which may not be feasible in practice. In the presence of unknown reliability, one solution is to conduct a validation study on a sample of subjects and estimate the reliability coefficient. Future work is needed to address issues related to the validation such as selection and size of validating sample and accounting for variations brought by estimating the coefficient.

In the present paper we confined our attention to developments for multivariate normal distributions because of the existence of a closed form of the accuracy index in terms of the Pearson correlation coefficient. Such a closed form provides an ideal situation to exactly evaluate the biases due to measurement errors and to develop methods to reduce biases. Beyond multivariate elliptical distributions (as a generalization of multivariate normal distributions), we are not aware of such a closed form for other multivariate distributions that would allow similar developments.

When knowledge is limited about the distributions of the reference biomarker and the biomarker, we can use the nonparametric estimate suggested by Obuchowski (2005, 2006) based on the observations (X_i, Y_{oi}) :

$$\tilde{\eta}_o = \frac{2}{n(n-1)} \sum_{i \neq j}^n \left(\psi_{X^{(i,j)}} \psi_{Y_o^{(ij)}} - (1 - \psi_{X^{(i,j)}})(1 - \psi_{Y_o^{(ij)}}) \right),$$

where ψ is defined as:

$$\psi_{X^{(i,j)}} = \begin{cases} 1, & \text{if } X_i - X_j > 0, \\ 0.5, & \text{if } X_i - X_j = 0, \\ 0, & \text{otherwise;} \end{cases} \quad \psi_{Y_o^{(i,j)}} = \begin{cases} 1, & \text{if } Y_{oi} - Y_{oj} > 0, \\ 0.5, & \text{if } Y_{oi} - Y_{oj} = 0, \\ 0, & \text{otherwise.} \end{cases}$$

The estimator $\tilde{\eta}_o$ is unbiased for η_o and thus biased for η when used as an estimator of η . Future research needs to investigate the magnitude and direction of the biases and how such biases can be corrected. One challenge is that the absence of a functional form of the biases makes it hard to correct for the biases. Efficiency and robustness of such a nonparametric estimator against various distributional assumptions and parametric estimators is also worth exploring.

Acknowledgments

The authors thank Dr. Tonja Nansel for providing the glycemic control data from the intervention study to illustrate the methods.

Funding

Research of Aiyi Liu was fully supported by the Intramural Research Program of the National Institute of Child Health and Human Development, National Institutes of Health. Research of Mixia Wu was partially supported by the General Program of Science and Technology Development Project of Beijing Municipal Education Commission (km201410005011).

References

- Fang HB, Fang KT, Kotz S (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis* 82:1–16.
- Kendall M (1938). A new measure of rank correlation. *Biometrika* 30:81–89.
- Krzanowski WJ, Hand DJ (2009). *ROC Curves for Continuous Data*. Boca Raton, FL: Chapman & Hall/CRC.
- Nansel TR, Iannotti RJ, Liu A (2012). Clinic-integrated behavioral intervention for families of youth with type 1 diabetes: Randomized clinical trial. *Pediatrics* 129:866–873.
- Obuchowski NA (2005). Estimating and comparing diagnostic tests accuracy when the gold standard is not binary. *Academic Radiology* 12:1198–1204. [PubMed: 16099683]
- Obuchowski NA (2006). An ROC-type measure of diagnostic accuracy when the gold standard is continuous-scale. *Statistics in Medicine* 25:481–493. [PubMed: 16287217]
- Pepe MS (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, UK: Oxford University Press.
- Zhou XH, Obuchowski NA, McClish DK (2011). *Statistical Methods in Diagnostic Medicine*, 2nd ed. New York, NY: Wiley.
- Zou KH, Liu A, Bandos AI, Ohno-Machado L, Rockette HE (2011). *Statistical Evaluation of Diagnostic Performance: Topics in ROC Analysis*. New York, NY: Chapman and Hall/CRC.

Figure 1.
Bias incurred by measurement error in estimating the true accuracy index.