Published in final edited form as:

Stat Med. 2014 April 15; 33(8): 1307–1320. doi:10.1002/sim.6027.

Combining markers with and without the limit of detection

Ting Dong^a, Catherine Chunling Liu^b, Emanuel F. Petricoin^c, and Liansheng Larry Tang^{a,*†}

^aDepartment of Statistics, George Mason University, Fairfax, VA 22030, U.S.A

^bDepartment of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

^cCenter for Applied Proteomics and Molecular Medicine, School of Systems Biology, George Mason University, Fairfax, VA 22030, U.S.A

Abstract

In this paper, we consider the combination of markers with and without the limit of detection (LOD). LOD is often encountered when measuring proteomic markers. Because of the limited detecting ability of an equipment or instrument, it is difficult to measure markers at a relatively low level. Suppose that after some monotonic transformation, the marker values approximately follow multivariate normal distributions. We propose to estimate distribution parameters while taking the LOD into account, and then combine markers using the results from the linear discriminant analysis. Our simulation results show that the ROC curve parameter estimates generated from the proposed method are much closer to the truth than simply using the linear discriminant analysis to combine markers without considering the LOD. In addition, we propose a procedure to select and combine a subset of markers when many candidate markers are available. The procedure based on the correlation among markers is different from a common understanding that a subset of the most accurate markers should be selected for the combination. The simulation studies show that the accuracy of a combined marker can be largely impacted by the correlation of marker measurements. Our methods are applied to a protein pathway dataset to combine proteomic biomarkers to distinguish cancer patients from non-cancer patients.

Keywords

ROC curve; diagnostic accuracy; limit of detection; linear discriminant analysis

1. Introduction

A diagnostic marker is used to screen or diagnose diseases. To evaluate the diagnostic accuracy of a marker, the ROC curve and the area under the curve (AUC) are commonly used. When measurements are subject to the limit of detection (LOD), which is due to the limited detecting ability of the equipment or instrument, it is difficult to measure markers at a relatively low level. If the marker value of a subject is beyond the range of the detection,

the subject is assigned the measurement of NA. The simplest approach to estimate the ROC curve is based on complete case analysis, which discards marker values subject to the LOD and only uses the measured values. However, this will lead to biased ROC estimators particularly when sample size is small or discarding proportion is relatively high. An alternative approach is to substitute NA with the LOD value, but this approach may also result in biases in the estimation. A more appropriate approach is to use the maximum likelihood method based on the normal distributions for the case and control groups while appropriately accounting for the LOD. The maximum likelihood method may lead to an improved AUC estimate compared with other methods [1].

Methods are available to evaluate the diagnostic accuracy of a marker with or without LOD. However, because it is common that different markers are sensitive to different aspects of a disease, one marker may not be accurate enough to distinguish a case from a control. It is thus desired to borrow information from two or more markers to improve the accuracy. Su and Liu [2] consider a linear combination of diagnostic markers to maximize the AUC of the combined markers under the multivariate normal assumption. Liu *et al.* [3] extend Su and Liu's work to maximize the partial AUC. They derive a linear combination method that achieves higher sensitivities in a range of specificities. A logistic regression method is proposed by Pepe and Thompson [4]. Yuan and Ghosh [5] suggest combining logistic regression models by putting a weight on each model. In their paper, several weighted schemes are considered, and the AIC is used to evaluate the performance of combined logistic regression models. However, how to combine multiple markers with the presence of the LOD has been rarely discussed in the literature. In this paper, we propose a method to combine multiple markers with the LOD. In addition, we propose to use the correlation among marker observations to select and combine markers from a large number of candidate markers without the LOD. This method seems to be contradictory with the common belief of combining top markers with the best accuracy. We conduct large-scale simulation studies to evaluate the performance of the proposed methods, and the results are presented in Section 3. In Section 4, we apply the proposed method in an example to illustrate the performance of our proposed methods. Finally, we discuss the implications and limits of our methods in Section 5.

2. Methods

We propose a method to combine markers with the presence of LOD in Section 2.1 and a method to use correlation between markers to select and combine markers from a large number of candidate markers in Section 2.2.

2.1. Combine markers with the limit of detection

Suppose we have m cases and n controls. Marker values, $X_{\ell i}$ and $\tilde{Y}_{\ell j}$, of marker ℓ are measured on case i and control j , respectively, for $\ell = 1, \dots, k$, $i = 1, \dots, m$, and $j = 1, \dots, n$. With the limitation of the measuring devices, marker values below the LOD are not observable, and the actual observations are

$$X_{li} = \begin{cases} \tilde{X}_{li}, & \tilde{X}_{li} \geq d_l \\ NA, & \tilde{X}_{li} < d_l \end{cases}, \quad \text{and} \quad Y_{lj} = \begin{cases} \tilde{Y}_{lj}, & \tilde{Y}_{lj} \geq d_l \\ NA, & \tilde{Y}_{lj} < d_l \end{cases},$$

where d_ℓ is the known LOD value for marker ℓ . One may assume that X_{li} and Y_{lj} both follow normal distributions and empirically estimate means and covariance matrices for two groups using only non-NA observations. Another method is to replace NA with some constant $c = d$ or $d/2$ and then estimate parameters using all the data [6, 7]. But both methods may lead to biased parameter estimators [8].

Existing parametric methods to handle the LOD issue are mainly based on the normal distribution assumption. That is, after some monotonic transformation, $\tilde{X}_i = (\tilde{X}_{1i}, \dots, \tilde{X}_{ki})' \sim N\{\boldsymbol{\mu}_d = (\mu_{1d}, \dots, \mu_{kd})', \boldsymbol{\Sigma}_d\}$ and $\tilde{Y}_j = (\tilde{Y}_{1j}, \dots, \tilde{Y}_{kj})' \sim N\{\boldsymbol{\mu}_d = (\mu_{1d}, \dots, \mu_{kd})', \boldsymbol{\Sigma}_d\}$, where $\boldsymbol{\Sigma}_d$ and $\boldsymbol{\Sigma}_d$ are variance-covariance matrices for cases and controls, respectively. Vexler *et al.* [1] present a maximum likelihood method to estimate parameters in the distributions of two markers with the LOD. For more markers with the LOD, Heather and Robert [7] propose a maximum likelihood method to estimate the parameters of multivariate normal distributions.

There are three types of observations for cases: (i) all of $\tilde{X}_{1i}, \dots, \tilde{X}_{ki}$ are completely observed; (ii) none of $\tilde{X}_{1i}, \dots, \tilde{X}_{ki}$ are observed; and (iii) some of markers, $\tilde{X}_{1i}, \dots, \tilde{X}_{ui}$, are observed and others, $\tilde{X}_{u+1,i}, \dots, \tilde{X}_{ki}$, are below the LOD. The same situation applies to controls. Then the likelihood function for cases is

$$L_X(\boldsymbol{\mu}_d, \sum_d) = \prod_{i=1}^{m_1} t_{i1}^x \prod_{i=m_1+1}^{m_2} t_{i2}^x \prod_{i=m_2+1}^{m_3} t_{i3}^x,$$

with

$$t_{i1}^x = \left((2\pi)^{k/2} |\sum_d|^{1/2} \right)^{-1} \exp \left[-0.5 (\tilde{X}_i - \boldsymbol{\mu}_d)' \sum_d^{-1} (\tilde{X}_i - \boldsymbol{\mu}_d) \right], \quad t_{i2}^x = F_d(d_1, \dots, d_k),$$

and

$$t_{i3}^x = F_d(d_{u+1}, \dots, d_k | \tilde{X}_{1i}, \dots, \tilde{X}_{ui}) f_d(\tilde{X}_{1i}, \dots, \tilde{X}_{ui}),$$

where $F_d(\cdot|\cdot)$ and $f_d(\cdot)$ are the conditional CDF and the density function of the multivariate normal distribution, respectively, for a case. Similarly, the likelihood function for controls is

$$L_Y(\boldsymbol{\mu}_d, \sum_d) = \prod_{j=1}^{n_1} t_{j1}^y \prod_{j=n_1+1}^{n_2} t_{j2}^y \prod_{j=n_2+1}^{n_3} t_{j3}^y,$$

where t_{j1}^y , t_{j2}^y , and t_{j3}^y are obtained by substituting the marker observations and the parameters specific for the control group in t_{i1}^x , t_{i2}^x , and t_{i3}^x , respectively. m_1 , m_2 , and m_3 are

the sample sizes of three types for the case, where $m_1 + m_2 + m_3 = m$. n_1 , n_2 , and n_3 are the sample sizes of three types for the control, where $n_1 + n_2 + n_3 = n$. These two likelihood functions are maximized to get the estimators for the mean parameters, $\hat{\boldsymbol{\mu}}_d = (\hat{\mu}_{1d}, \dots, \hat{\mu}_{kd})'$ and $\hat{\boldsymbol{\mu}}_{\bar{d}} = (\hat{\mu}_{1\bar{d}}, \dots, \hat{\mu}_{k\bar{d}})'$, and the estimators for the covariance matrices, $\hat{\boldsymbol{\Sigma}}_d$ and $\hat{\boldsymbol{\Sigma}}_{\bar{d}}$, which are obtained by substituting the variance and correlation parameter with their estimators in $\boldsymbol{\Sigma}_d$ and $\boldsymbol{\Sigma}_{\bar{d}}$.

After the parameters in the multivariate normal distributions are estimated, we apply the results derived by Su and Liu [2] and Liu *et al.* [3] to obtain the optimal linear combination of markers to maximize the AUC of the combined marker. Let $\boldsymbol{\beta}_c = (\beta_1, \dots, \beta_k)'$. The combined marker has the value of $\boldsymbol{\beta}_c' \tilde{X}_i$ for case i and $\boldsymbol{\beta}_c' \tilde{Y}_j$ for control j . Using the extreme value theorem, Su and Liu [2] and Liu *et al.* [3] show that the coefficient vector of the optimal linear combination is given by

$$\boldsymbol{\beta}^* = c(\boldsymbol{\Sigma}_d + \boldsymbol{\Sigma}_{\bar{d}})^{-1}(\boldsymbol{\mu}_d - \boldsymbol{\mu}_{\bar{d}}),$$

for any positive constant c , and the AUC of the optimally combined marker takes the form of

$$AUC = \Phi \left(\sqrt{(\boldsymbol{\mu}_d - \boldsymbol{\mu}_{\bar{d}})' (\boldsymbol{\Sigma}_d + \boldsymbol{\Sigma}_{\bar{d}})^{-1} (\boldsymbol{\mu}_d - \boldsymbol{\mu}_{\bar{d}})} \right). \quad (1)$$

The coefficient vector can be estimated by substituting the parameters with their consistent estimators:

$$\hat{\boldsymbol{\beta}}^* = c \left(\hat{\boldsymbol{\Sigma}}_d + \hat{\boldsymbol{\Sigma}}_{\bar{d}} \right)^{-1} (\hat{\boldsymbol{\mu}}_d - \hat{\boldsymbol{\mu}}_{\bar{d}}), \quad (2)$$

and the resulting AUC estimator is obtained by plugging respective parameter estimators in (1).

2.2. Combine multiple markers without limit of detection

Without the presence of the LOD, the mean and variance parameters in the normal distributions can be empirically estimated. These estimators can be plugged in (2) to estimate the coefficient vector of the best linear combination of markers and then in (1) to estimate the AUC of the combined marker. Note that for the sake of simplicity, the combined marker will refer to the marker generated from the optimal combination of markers hereafter. If one is only to select a few out of a large pool of candidate markers, a common understanding is to choose the markers with the highest AUCs. However, this may not be the case as will be indicated through our example. Instead, the correlation between measurements of two markers may play a more important role.

We notice that when fixing the means and variances in (1), the AUC of the combined marker becomes a function of correlation coefficients. Let ρ_d , ℓ_1 , ℓ_2 and $\rho_{\bar{d}}$, ℓ_1 , ℓ_2 be the correlation coefficients of measurements between markers ℓ_1 and ℓ_2 in a case and a control,

respectively. If we further assume that the correlation coefficients from the case and the control are the same, say, $\rho_{d, \ell_1, \ell_2} = \rho_{d, \bar{\ell}_1, \bar{\ell}_2} = \rho_{\ell_1, \ell_2}$, Figure 1 shows that the AUC value of the combined marker in (1) is a convex function of ρ_{ℓ_1, ℓ_2} . The convex property can be showed for bivariate normal distributions by taking the first derivative of the AUC expression in (1) with regard to ρ_{ℓ_1, ℓ_2} , but the tedious algebra is omitted here. We see that the further away $\rho_{1,2}$ is from the minimum $\rho_{1,2}^*$, the larger the AUC is. This provides us the motivation to select markers based on the distance between the estimated correlation coefficient, r_{ℓ_1, ℓ_2} , and the minimum, ρ_{ℓ_1, ℓ_2}^* . However, some complication may arise in practice because it is uncommon to have the same correlation coefficients between two markers in the case and the control. To fully account for the effect of the correlation in both groups, a possible way is to use $r_{max, \ell_1, \ell_2} = \max(r_{d, \ell_1, \ell_2}, r_{d, \bar{\ell}_1, \bar{\ell}_2})$ where r_{d, ℓ_1, ℓ_2} and $r_{d, \bar{\ell}_1, \bar{\ell}_2}$ estimate ρ_{d, ℓ_1, ℓ_2} and $\rho_{d, \bar{\ell}_1, \bar{\ell}_2}$, respectively. The distance can then be calculated between r_{max, ℓ_1, ℓ_2} and ρ_{ℓ_1, ℓ_2}^* using $|r_{max, \ell_1, \ell_2} - \rho_{\ell_1, \ell_2}^*|$. We hereafter call this distance the *correlation distance*. In our procedure for selecting a subset of markers, we propose to start with the marker with the highest AUC. The marker is denoted as marker 1. We then obtain the $r_{d, 1, \ell}$, $r_{d, \bar{1}, \bar{\ell}}$ and $r_{max, 1, \ell} = \max(r_{d, 1, \ell}, r_{d, \bar{1}, \bar{\ell}})$ between marker 1 and marker ℓ , $\ell = 2, \dots, L$. Fixing the sample means and sample variances derived from measurements of markers 1 and ℓ , we write the AUC of the combined marker from markers 1 to ℓ as a function of $\rho_{1, \ell}$ and obtain the minimum $\rho_{1, \ell}^*$. Subsequently, we find the marker with the largest value of *correlation distance*, $|r_{max, 1, \ell} - \rho_{1, \ell}^*|$ for $\ell = 2, \dots, L$, and denote this marker as marker 2. Markers 1 and 2 are then combined in a linear fashion using the empirically estimated coefficient vector given in (2). If the goal of the marker combination is to combine L_0 markers out of a pool of L markers for $L > L_0$, and $L_0 = 2$, we stop selecting more markers; otherwise, we continue with selecting the next marker from the rest of $L - 2$ markers. We calculate the *correlation distance* between each of the $L - 2$ markers and the combined marker in the previous selection, and identify the marker with the largest *correlation distance*. The marker is then combined with previously selected markers to obtain a new combined marker using Su and Liu [2]'s method. These steps are repeated until L_0 markers are selected and combined.

One may also be interesting in selecting and combining a subset of markers to achieve some desired value of the AUC. Suppose that a single marker does not have the desired AUC and more than one markers need to be combined to improve the accuracy. The final steps in the aforementioned procedure can be modified accordingly. We stop the selection if the combination of two markers achieves the desired AUC; otherwise, we continue with selecting and combining the next marker from the rest of markers based on their *correlation distance* from the combined marker in the previous selection until the desired accuracy is reached.

A traditional method to choose a subset of L_0 markers is to consider all subsets of L_0 markers and choose the subset that gives the highest AUC of the combined marker. The proposed method does not show much computational advantage over this traditional method in the setting of selecting a fixed number of markers. However, the advantage of the proposed method is more clear when the selection criteria is to obtain a desired AUC value.

Using the traditional method, we may need to consider all possible subsets of 1 marker, 2 markers, 3 markers, ..., until the desired accuracy is reached, while the proposed procedure considers a much less number of combinations. Suppose that both the traditional method and the proposed method choose l_0 markers at the end. The former method requires the

computation of $L + \binom{L}{2} + \dots + \binom{L}{l_0}$ marker combinations, while the latter requires $L + (L-1) + \dots + (L-l_0+1)$ marker combinations. The savings on the computation could become substantial with a large number of candidate markers.

3. Simulation studies

In this section, we demonstrate the performance of our methods for combining two or three markers with the LOD in Sections 3.1 and 3.2. The comparison of the proposed method with existing methods is shown, respectively. Section 3.3 demonstrates the selection of combined markers without the LOD. It illustrates how the AUC of the combined marker changes with the correlation between marker measurements.

3.1. Simulation for the combination of two markers with limit of detection

The ROC curve under the binormal distribution (normal distributions for two groups) takes on the form of $ROC(p) = \Phi(\alpha + \beta\Phi^{-1}(p))$, where $\alpha = (\mu_d - \mu_{\bar{d}})/\sigma_d$ and $\beta = \sigma_{\bar{d}}/\sigma_d$. The parameters, α and β , determine the shape of the ROC curve. We simulate observations from bivariate normal distributions with the forms of $(X_1, X_2) \sim N\{(\mu_1, \mu_2), \Sigma\}$, and $(Y_1, Y_2) \sim N\{(0, 0), \Sigma\}$, where Σ has diagonal elements 1 and off-diagonal elements ρ . We choose $m = n = 200$ and ρ from (0.1, 0.25) in our simulation. μ_1 and μ_2 are computed according to the specified AUC values. Three pairs of AUCs, (0.70, 0.75), (0.75, 0.80) and (0.70, 0.80), are specified. The three corresponding pairs of (μ_1, μ_2) are (0.7416, 0.9539), (0.9539, 1.1902), and (0.7416; 1.1902).

In our simulation, we use the same LOD value, d , to truncate measurements for the case and control groups. Four different LOD values ($-5, -2, -1, 0$) are used to vary the number of measurements subjected to the LOD. We compare the proposed method with the procedure, which ignores the LOD and the procedure, which replaces the truncated measurements with the LOD value. The bias and mean square error (MSE) for the estimators used by these methods are computed. The results are presented in Tables I and II. The bias and MSE of α are displayed in Table I. It is shown that the bias and MSE tend to increase when the correlation coefficient increases. By comparing three methods, our method has the smallest bias and MSE than the other two methods. The ignoring LOD method has the worst performance in terms of the bias and MSE. The bias and MSE of β are displayed in Table II. When the correlation coefficient increases, the bias and MSE tend to increase. Among three methods, our method has the smallest bias and MSE. The method, which replaces NA with the LOD, performs differently from the one ignoring the LOD. The former has smaller bias and MSE when $d = -2$ and $d = -1$, while the latter has smaller bias and MSE when $d = 0$. In general, our method has the smallest biases and MSEs in all the cases.

Figure 2 shows four ROC curves in the case of $d = 0$. The red curve is the true ROC curve of the combined marker. The blue curve is the ROC curve of the combined marker using our

method. The black curve is the ROC curve of the combined marker using the method ignoring LOD. The green curve is the ROC curve of the combined marker using replacing value 0. The ROC curve of the combined marker generated by our method is closer to the true ROC curve. This can be seen from Figure 2.

3.2. Simulation for the combination of three markers with limit of detection

We simulate observations from multivariate normal distributions with $(X_1, X_2, X_3) \sim N\{(0.7416, 0.9539, 1.1902), \Sigma\}$ and $(Y_1, Y_2, Y_3) \sim N\{(0, 0, 0), \Sigma\}$, where Σ has diagonal elements 1 and off-diagonal elements ρ . We choose $m = n = 1000$ and $\rho = 0.1$ in our simulation. The LOD d takes values of -5, -3, -2, and -1.5. We compare our procedure with the procedures that ignore the truncated measurements (ignoring LOD) or replace the truncated measurements with the LOD (replacing LOD).

Table III presents the biases and MSEs for α and β . It is shown that in the case of no truncated measurements, $d = -5$, three methods perform equally well in terms of the bias and MSE. When setting $d = -3$ truncates less than 0.5% of the observations, all methods generate relatively accurate estimates for α , but the proposed method and the replacing LOD method perform better in the estimation of β . The advantage of the proposed method can be seen clearly when $d = -1.5$ because its biases and MSEs are the smallest among all three methods. The ignoring LOD method tends to yield least accurate estimates for α and β when d gets smaller.

3.3. Simulation for the selection of combined markers without the limit of detection

In this section, we illustrate how the correlation between markers affects the AUC of the combined markers. Two hundred observations are simulated for the cases and for the controls from multivariate normal distributions. The bivariate normal model has the forms of $(X_1, X_2) \sim N\{(\mu_1, \mu_2), \Sigma_2\}$ for the case and $(Y_1, Y_2) \sim N\{(0, 0), \Sigma_2\}$ for the control, where Σ_2 has diagonal elements 1 and off-diagonal elements ρ . We use the normal model, $(X_1, X_2, X_3) \sim N\{(\mu_1, \mu_2, \mu_3), \Sigma_3\}$ and $(Y_1, Y_2, Y_3) \sim N\{(0, 0, 0), \Sigma_3\}$, where Σ_3 has diagonal elements 1 and off-diagonal elements ρ , to generate observations of three markers. In our simulations, the mean values are computed according to the AUC values of 0.55, 0.60, 0.70, and 0.75, respectively. The corresponding mean values were 0.1777, 0.3583, 0.7416, and 0.9539. Moreover, we choose the correlation coefficient between markers to be from -0.8, -0.5, 0, 0.1, 0.5, 0.8 to evaluate how much improvement can be achieved through the combination of markers with different AUCs.

In Table IV, we consider combining two markers with AUCs of 0.55, 0.60, two more accurate markers with AUCs of 0.75, 0.70 as well as markers with AUCs of 0.75, 0.55, respectively. From these results, we can see that all the CIs of the combined markers are greater than zero, that is, the combinations statistically improve the diagnostic accuracy compared with each individual marker. With the same correlation coefficient between markers, to combine more accurate markers would have better diagnostic accuracy than combining less accurate markers. These results are also shown in Figure 1 in the previous section.

In Table V, three markers with different AUCs of 0.75, 0.70, and 0.55 are combined with various correlation coefficients between these markers. The correlation coefficient between markers 1 and 2 is ρ_1 . The correlation coefficient between markers 1 and 3 is ρ_2 . The correlation coefficient between markers 2 and 3 is ρ_3 . In scenario 1, we assume the same correlation coefficient among three markers. In scenario 2, markers 1 and 2 have a moderate correlation ($\rho_1 = 0.5$). We set equal correlations for ρ_2 and ρ_3 . In scenario 3, we keep the correlation between markers 1 and 2 as $\rho_1 = 0.5$ and let markers 1 and 3 be independent ($\rho_2 = 0$). Various ρ_3 , the correlation coefficient between markers 2 and 3, are considered. Results listed in the Table V indicate that the correlation coefficient between markers plays an important role on the improvement of the diagnostic accuracy for a combined marker.

4. Example

4.1. Combination of lung cancer markers without the limit of detection

To illustrate the application of our proposed marker selection method, we use a lung cancer dataset with protein markers collected by Center for Applied Proteomics and Molecular Medicine at George Mason University. The objective of this study is to use the protein pathway activation to distinguish molecular subtypes of non-small-cell lung carcinoma tumors. There are 27 adenocarcinomas cells and 20 squamous cells in this dataset. Laser capture microdissection was used to take tumor cells from 47 frozen non-small-cell lung carcinoma tumors from lung cancer patients. The activation of signaling proteins was evaluated by reverse phase protein microarrays [9]. First, we want to combine some of these markers to increase the diagnostic accuracy. Four markers, Bad S136, EGFR Y1289, Foxo 01/03, GSK3 alpha-beta S21/9 in our dataset are selected because they do not have missing values, and the log transformation is suggested by the Box-Cox method for all of them. Using the optimal linear combination method in [2], four markers are normalized first, and then the coefficient vector for the linear combination is estimated to be $(1.1032, 0.6568, 0.4412, -0.5844)'$. The AUC of the combined marker is estimated to be 0.872, which is larger than the estimated AUC of each individual marker. This can also be seen in Figure 3 regarding their empirically estimated ROC curves.

Second, we illustrate the proposed marker selection procedure based on the idea of *correlation distance* using the following example. Suppose there are six markers without any missing values: 4E-BP1T70, Ack1Y284, AuroraA(T288)/B(T232)/C(T198), Bad S136, c-AbIT735, and ErbB3/HER3(1B2), with estimated AUCs of 0.7426, 0.7056, 0.7056, 0.8037, 0.7130, and 0.8185, respectively. After the normalization of their measurements, we either combine the best two markers in terms of their AUCs or follow the proposed algorithm to combine markers. When we combine the best two markers, Bad S136 and ErbB3/HER3(1B2), we obtain an AUC of 0.9074. However, on the basis of our procedure, we want to combine marker ErbB3/HER3(1B2) with marker AuroraA (T288)/B(T232)/C(T198), and combining these two markers has the AUC of 0.9148, which offers better accuracy than combining two best markers. The empirical ROC curves of combining Bad S136 and ErbB3/HER3(1B2), and combining ErbB3/HER3(1B2) and AuroraA(T288)/B(T232)/C(T198), are shown in Figure 4. The estimated correlation coefficients are shown in Table VI. Marker 1, ErbB3/HER3(1B2), has the largest AUC, 0.8185. Then, we calculate

the correlation coefficients between marker 1 and other five markers. We compute the correlation $r_{\max, 1, \ell}$ between marker 1 and marker ℓ , for $\ell = 2, \dots, 6$. For marker ℓ , we obtain the AUC from combining marker ℓ and marker 1 as a function of $\rho_{1, \ell}$. The minimum $\rho_{1, \ell}^*$ is then calculated as the value of $\rho_{1, \ell}$ where the minimum of the AUC occurs. The *correlation distance* between $r_{\max, 1, \ell}$ and $\rho_{1, \ell}^*$ was calculated for marker ℓ . The marker AuroraA(T288)/B(T232)/C(T198) has the largest *correlation distance* 0.32 in our example. Thus, we combine ErbB3/HER3(1B2) with AuroraA(T288)/B(T232)/C(T198).

Suppose that in order to obtain a higher diagnostic accuracy, we want to combine three markers instead of two markers. A possible way is to combine the best three markers, the ErbB3/HER3(1B2), Bad S136, and 4E-BP1T70, which gives the AUC of 0.9111. However, by using the proposed procedure based on the *correlation distance*, three markers, ErbB3/HER3(1B2), AuroraA(T288)/B(T232)/C(T198), and Ack1Y284, were selected, and the AUC of the combined marker is 0.9370, which is higher than that of combining the three most accurate markers. The empirical ROC curves of these two combined markers are shown in Figure 5. The estimated correlation coefficients are shown in Table VII. Based on the results in Table VI, we combined markers, ErbB3/HER3(1B2) and AuroraA(T288)/B(T232)/C(T198) first, and then computed the *correlation distance* of this combined marker with other individual markers. We see that in Table VII, marker Ack1Y284 has the largest distance of 1.0395 and thus include it in the final set of three markers for the combination.

4.2. Combination of lung cancer markers with the limit of detection

To illustrate the proposed method for combining markers with the LOD, we first create a pseudo dataset with two markers, 4E-BP1T70 and Ack1Y284, from the aforementioned lung cancer dataset. After the log transformation of original marker values, the transformed measurements are then truncated using the LOD value of 8.55. The biomarker, Ack1Y284, has four case measurements and one control measurement below the LOD. The biomarker, 4E-BP1T70, has one case measurement below the LOD. We then apply the proposed method to the dataset and compare with the methods, which ignores the LOD or replaces the LOD. Because parameter estimates for α and β can be obtained from the original complete marker values without truncation, these estimates can be regarded as ‘true’ parameter values so that comparison can be made among three methods regarding the bias and the MSE. The results are shown in Table VIII. It indicates that our method has the smallest bias and MSE among the three methods.

We also create another pseudo dataset with three markers, ErbB3/HER3(1B2), 4E-BP1T70, and Ack1Y284. Original marker measurements are log transformed, and then the values of the LOD, 8.0 and 8.5, are used to truncate ErbB3/HER3(1B2) and 4E-BP1T70, respectively. The biomarker, ErbB3/HER3(1B2), has eight measurements below the LOD. Only one measurement is below the LOD for the biomarker, 4E-BP1T70. The results of α and β in Table IX indicate that the proposed method has the smallest bias and MSE.

5. Discussion

In this paper, we propose algorithms to combine markers with or without the LOD. We conduct extensive simulation studies to evaluate the performance of the proposed procedure. Our simulation results show that the ROC parameter estimators from the proposed method are much closer to the true ones than existing methods.

Similar to existing ROC methods based on the parametric distribution assumption, our procedure may be sensitive to the departure of the normal assumption. We applied the proposed combination method to datasets from a t-distribution, but the Newton-Raphson procedure for estimating the normal distribution parameters could not converge. Even if the procedure converges and generates parameter estimates, we expect that the departure of the normal assumption may lead to biases in parameter estimation and inaccurate coefficient estimates in the marker combination.

The proposed procedure requires the maximization of the likelihood function, which is derived from the multivariate normal distributions. In the simulation studies and the applications to the real example, we adopted an efficient algorithm developed by Genz [10] for the computation of the multivariate normal probabilities in the likelihood function. The same algorithm is also recommended by Heather and Robert for quick computation [7]. In fact, it takes around one minute to estimate parameters for the three-dimensional normal distribution on one simulated dataset in Section 3.2. The R code is available from the first author upon request.

Moreover, although many methods are proposed for combining markers in the literature, existing methods do not provide good rationale for choosing the best markers in the combination. Our simulation studies show that simply combining the set of markers with the best AUCs does not generate the highest AUC. Instead, the correlation played a bigger role in obtaining the best AUC for combined markers. This serves as our motivation for finding the set of markers. We propose a procedure for combining the markers based on the *correlation distance*. The accuracy of the combined marker based on our procedure is showed to be better than the one based on combining the most accurate markers. There are two limitations of our procedure. First, the *correlation distance*, which we rely on for the marker selection, is based on the assumption of the same correlation coefficients for the case and the control. It is worthwhile to explore more complex situations on the combination methods when the assumption does not hold. Second, we use results from the linear discriminant analysis to combine markers, which assumes that the measurements follow multivariate normal distributions. The distribution misspecification may lead to an incorrect AUC of the combined marker.

In general, there are two possible applications of the methods proposed in this paper. First, due to the hierarchical structure of the disease, one marker may not be accurate enough to detect the disease. For example, a new type of proteomic concept, serum proteomic pattern diagnostics, is based on the combination of markers [11]. In a high dimensional mass spectrometry data, how to find the subset of available markers to combine is important. In contrast to existing methods, our marker selection procedure chooses a marker with the

correlation further away from the one with the smallest AUC. Our procedure can be used to find the subset of markers to obtain a desired accuracy. Second, in practice, it is common that equipments or instruments have a detection range. Beyond this range, the LOD occurs. Ignoring these measurements can bias our accuracy estimation. The combination procedure proposed can be used to combine the markers with the LOD.

Acknowledgments

The authors would like to thank the associate editor and a referee for their constructive comments.

References

1. Vexler A, Liu A, Eliseeva E, Schisterman EF. Maximum likelihood ratio tests for comparing the discriminatory ability of biomarkers subject to limit of detection. *Biometrics*. 2008; 64:895–903. [PubMed: 18047527]
2. Su JQ, Liu JS. Linear combinations of multiple diagnostic markers. *Journal of the American Statistical Association*. 1993; 88:1350–1355.
3. Liu A, Schisterman EF, Zhu Y. On linear combinations of biomarkers to improve diagnostic accuracy. *Statistics in Medicine*. 2005; 24:37–47. [PubMed: 15515132]
4. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics*. 2000; 1:123–140. [PubMed: 12933515]
5. Yuan Z, Ghosh D. Combining multiple biomarkers models in logistic regression. *Biometrics*. 2008; 64:431–439. [PubMed: 18325068]
6. Hornung RW, Reed LD. Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*. 1990; 5:46–51.
7. Hoffman HJ, Johnson RE. Estimation of multiple trace metal water contaminants in the presence of left-censored and missing data. *Journal of Environmental Statistics*. 2011; 2 <http://www.jenvstat.org/v02/i02/paper>.
8. Perkins NJ, Schisterman EF, Vexler A. Receiver operating characteristic curve inference from a sample with a limit of detection. *American Journal of Epidemiology*. 2007; 165:325–333. [PubMed: 17110640]
9. Zupa A, Improta G, Deng J, Aieta M, Musto P, Liotta LA, Belluco C, Mammano E, Wulfkühle JD, Petricoin EF. Use of protein pathway activation mapping of NSCLC to identify distinct molecular subtypes and a prognostic signature for aggressive node-negative tumors. *Journal of Clinical Oncology*. 2010; 28:15s. http://meeting.ascopubs.org/cgi/content/abstract/28/15_suppl/10594.
10. Genz A. Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*. 1992; 1:141–149.
11. Petricoin EF, Ornstein DK, Liotta LA. Clinical proteomics: applications for prostate cancer biomarker discovery and detection. *Urologic Oncology: Seminars and Original Investigations*. 2004; 22:322–328.

Figure 1.

The figure shows the simulation results of combining two markers. The black curve represents the AUC of the combined marker from markers with AUCs of 0.55 and 0.60. The blue curve represents the AUC of the combined marker of two markers with AUCs of 0.55 and 0.75. The red curve represents the combination of two markers with AUCs of 0.75 and 0.70. The dash lines show the ρ_{min} for each combination of two markers.

Figure 2.

ROC curves of the combined marker using different methods. The red solid curve is the ROC curve of the combined marker without the limit of detection (LOD). The blue dashed curve is the ROC curve of the combined marker using our method. The black dotted curve is the ROC curve of the combined marker using the method ignoring LOD. The green dash-dot curve is the ROC curve of the combined marker using the method replacing LOD. We use $d = 0$.

Figure 3.
The empirical ROC curves of individual markers and the combined marker.

Figure 4. Combination of three markers. The blue curve is the ROC curve of the combined marker 1 using our method. The red curve is the ROC curve of the combined marker 2 by combining the best two markers.

Figure 5. Combination of three markers. The blue curve is the ROC curve of the combined marker 1 using our method. The red curve is the ROC curve of the combined marker 2 by combining best three markers.

Table I

Simulation summary ($m = n = 200$) for α .

ρ	d	$\Omega_1^A \backslash \Omega_2^A$	Bias						MSE					
			Proposed			Ignoring LOD			Replacing LOD			Proposed		
			0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75
0.1	-5	0.75	-0.051	-0.055	-0.051	-0.054	-0.051	-0.054	-0.051	-0.054	0.121	0.122	0.121	0.121
		0.8	-0.046	-0.068	-0.048	-0.067	-0.048	-0.067	-0.048	-0.067	0.119	0.130	0.120	0.132
	-2	0.75	-0.058	-0.061	-0.107	-0.114	-0.107	-0.114	-0.060	-0.063	0.124	0.122	0.151	0.125
		0.8	-0.052	-0.073	-0.106	-0.129	-0.106	-0.129	-0.057	-0.077	0.122	0.133	0.151	0.136
0.25	-5	0.75	-0.089	-0.088	-0.276	-0.305	-0.276	-0.305	-0.120	-0.131	0.140	0.137	0.292	0.164
		0.8	-0.086	-0.102	-0.296	-0.333	-0.296	-0.333	-0.124	-0.149	0.140	0.150	0.312	0.180
	-2	0.75	-0.090	-0.100	-0.506	-0.575	-0.506	-0.575	-0.306	-0.351	0.139	0.147	0.516	0.359
		0.8	-0.106	-0.117	-0.578	-0.636	-0.578	-0.636	-0.349	-0.386	0.155	0.160	0.587	0.353
0.5	-5	0.75	-0.114	-0.143	-0.111	-0.135	-0.111	-0.135	-0.111	-0.135	0.155	0.177	0.153	0.173
		0.8	-0.109	-0.150	-0.108	-0.146	-0.108	-0.146	-0.108	-0.146	0.158	0.185	0.158	0.183
	-2	0.75	-0.119	-0.147	-0.164	-0.193	-0.164	-0.193	-0.120	-0.145	0.158	0.181	0.193	0.180
		0.8	-0.113	-0.153	-0.163	-0.204	-0.163	-0.204	-0.117	-0.155	0.161	0.189	0.197	0.190
0.75	-5	0.75	-0.144	-0.166	-0.331	-0.376	-0.331	-0.376	-0.180	-0.210	0.177	0.197	0.344	0.232
		0.8	-0.138	-0.176	-0.352	-0.405	-0.352	-0.405	-0.184	-0.226	0.180	0.208	0.364	0.248
	-2	0.75	-0.150	-0.172	-0.569	-0.643	-0.569	-0.643	-0.378	-0.428	0.185	0.204	0.576	0.435
		0.8	-0.153	-0.190	-0.620	-0.700	-0.620	-0.700	-0.406	-0.465	0.189	0.219	0.627	0.471

The results are based on 1000 simulated datasets. The true areas under the curve of two markers are Ω_1^A and Ω_2^A . The correlation coefficient between two markers is ρ .

MSE, mean square error; LOD, limit of detection. 'Proposed', our proposed method; 'Ignoring LOD', replacing the measurements subjected to the LOD with 'NA'; 'Replacing LOD', replacing the measurements subjected to the LOD with d , where d is the LOD point.

Table II

Simulation summary ($m = n = 200$) for β .

ρ	d	$\Omega_1^A \backslash \Omega_2^A$	Bias						MSE					
			Proposed			Ignoring LOD			Replacing LOD			Proposed		
			0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75	0.7	0.75
0.1	-5	0.75	-0.045	-0.046	-0.045	-0.045	-0.045	-0.045	-0.045	-0.045	0.064	0.059	0.066	0.064
		0.8	-0.042	-0.049	-0.042	-0.049	-0.042	-0.049	-0.042	-0.049	0.066	0.066	0.067	0.069
	-2	0.75	-0.023	-0.030	-0.091	-0.092	-0.091	-0.092	-0.062	-0.062	0.076	0.048	0.101	0.077
		0.8	-0.014	-0.026	-0.091	-0.097	-0.091	-0.097	-0.060	-0.066	0.080	0.052	0.103	0.078
0.25	-5	0.75	-0.018	-0.026	-0.178	-0.185	-0.178	-0.185	-0.145	-0.150	0.155	0.045	0.182	0.151
		0.8	-0.001	-0.014	-0.191	-0.196	-0.191	-0.196	-0.150	-0.156	0.161	0.045	0.196	0.201
	0	0.75	-0.090	-0.081	-0.245	-0.253	-0.245	-0.253	-0.319	-0.327	0.329	0.095	0.252	0.258
		0.8	-0.084	-0.083	-0.272	-0.277	-0.272	-0.277	-0.345	-0.347	0.349	0.104	0.278	0.281
	-2	0.75	-0.108	-0.115	-0.108	-0.114	-0.108	-0.114	-0.108	-0.114	0.173	0.117	0.121	0.118
		0.8	-0.088	-0.111	-0.088	-0.111	-0.088	-0.111	-0.088	-0.111	0.119	0.102	0.118	0.102
	-1	0.75	-0.088	-0.101	-0.149	-0.156	-0.149	-0.156	-0.123	-0.129	0.136	0.100	0.108	0.155
		0.8	-0.060	-0.090	-0.132	-0.153	-0.132	-0.153	-0.103	-0.126	0.133	0.081	0.100	0.141
	0	0.75	-0.082	-0.098	-0.227	-0.235	-0.227	-0.235	-0.198	-0.204	0.208	0.093	0.104	0.230
		0.8	-0.044	-0.078	-0.229	-0.243	-0.229	-0.243	-0.189	-0.207	0.210	0.070	0.089	0.233
		0.75	-0.131	-0.136	-0.277	-0.289	-0.277	-0.289	-0.348	-0.351	0.353	0.141	0.144	0.283
		0.8	-0.112	-0.133	-0.302	-0.310	-0.302	-0.310	-0.384	-0.374	0.376	0.128	0.143	0.310

The results are based on 1000 simulated datasets. The true areas under the curve of two markers are Ω_1^A and Ω_2^A . The correlation coefficient between two markers is ρ .

MSE, mean square error; LOD, limit of detection; 'Proposed', our proposed method; 'Ignoring LOD', replacing the measurements subjected to the LOD with 'NA'; 'Replacing LOD', replacing the measurements subjected to the LOD with d , where d is the LOD point.

Table III

Simulation summary ($m = n = 1000$) for α and β .

ρ	d	Bias		MSE	
		Proposed	Ignoring LOD	Replacing LOD	Replacing LOD
0.1	-5	0.0054	0.0034	0.0033	0.0039
	-3	0.0055	-0.0021	0.0035	0.0040
	-2	0.0103	-0.0620	-0.0064	0.0040
	-1.5	-0.0129	-0.1455	-0.0291	0.0043
Estimation of α					
0.1	-5	0.0005	0.0008	0.0008	0.0011
	-3	0.0006	-0.0064	-0.0003	0.0011
	-2	0.0058	-0.0580	-0.0183	0.0010
	-1.5	0.0042	-0.1118	-0.0530	0.0011
Estimation of β					
0.1	-5	0.0005	0.0008	0.0008	0.0011
	-3	0.0006	-0.0064	-0.0003	0.0011
	-2	0.0058	-0.0580	-0.0183	0.0010
	-1.5	0.0042	-0.1118	-0.0530	0.0011

The results are based on 1000 simulated datasets. The correlation coefficient between two markers is ρ . MSE, mean square error; LOD, limit of detection; 'Proposed', our proposed method; 'Ignoring LOD', replacing the measurements subjected to the LOD with 'NA'; 'Replacing LOD', replacing the measurements subjected to the LOD with d , where d is the LOD point.

Table IV

The simulation results of combining two markers ($m = n = 100$).

ρ	AUC_1 CI (New – marker)	AUC_2 CI (New – marker)	AUC_c
–0.8	0.55 (0.17, 0.18)	0.60 (0.12, 0.14)	0.73
–0.5	0.55 (0.095, 0.107)	0.60 (0.05, 0.06)	0.65
0	0.55 (0.06, 0.07)	0.60 (0.016, 0.02)	0.62
0.1	0.55 (0.06, 0.07)	0.60 (0.01, 0.02)	0.61
0.5	0.55 (0.05, 0.06)	0.60 (0.007, 0.01)	0.61
0.8	0.55 (0.06, 0.07)	0.60 (0.017, 0.023)	0.62
–0.8	0.55 (0.35, 0.36)	0.75 (0.15, 0.16)	0.91
–0.5	0.55 (0.25, 0.26)	0.75 (0.05, 0.06)	0.81
0	0.55 (0.19, 0.21)	0.75 (0.006, 0.008)	0.76
0.1	0.55 (0.19, 0.20)	0.75 (0.003, 0.005)	0.75
0.5	0.55 (0.20, 0.22)	0.75 (0.01, 0.02)	0.76
0.8	0.55 (0.27, 0.29)	0.75 (0.08, 0.09)	0.83
–0.8	0.7 (0.27, 0.28)	0.75 (0.22, 0.23)	0.97
–0.5	0.7 (0.18, 0.19)	0.75 (0.13, 0.14)	0.89
0	0.7 (0.10, 0.11)	0.75 (0.05, 0.06)	0.81
0.1	0.7 (0.09, 0.10)	0.75 (0.04, 0.05)	0.79
0.5	0.7 (0.06, 0.07)	0.75 (0.011, 0.014)	0.76
–0.8	0.7 (0.05, 0.06)	0.75 (0.002, 0.004)	0.76

The correlation coefficient between two markers is ρ . The area under the curve (AUC) of marker 1 is AUC_1 . The AUC of marker 2 is AUC_2 . The AUC of the combined marker is AUC_c . ‘CI (New-marker)’ represents the confidence interval of the difference between the combined marker and the corresponding individual marker.

Table V

The simulation results of combining three markers ($m = n = 100$).

$\rho_1 = \rho_2 = \rho_3$	AUC_1 CI (New – marker)	AUC_2 CI (New – marker)	AUC_3	AUC_c
-0.5	0.75 (0.24, 0.25)	0.70 (0.30, 0.31)	0.55 (0.44, 0.45)	1
0	0.75 (0.057, 0.06)	0.70 (0.106, 0.11)	0.55 (0.25, 0.26)	0.81
0.1	0.75 (0.043, 0.049)	0.70 (0.094, 0.10)	0.55 (0.24, 0.25)	0.80
0.5	0.75 (0.036, 0.041)	0.70 (0.08, 0.09)	0.55 (0.23, 0.24)	0.79
0.8	0.75 (0.09, 0.10)	0.70 (0.14, 0.15)	0.55 (0.29, 0.30)	0.85
$\rho_1 = 0.5$	AUC_1	AUC_2	AUC_3	New
$\rho_2 = \rho_3 = -0.8$	0.75 (0.23, 0.24)	0.70 (0.28, 0.29)	0.55 (0.42, 0.43)	0.98
$\rho_2 = \rho_3 = -0.5$	0.75 (0.08, 0.09)	0.70 (0.13, 0.14)	0.55 (0.27, 0.28)	0.83
$\rho_2 = \rho_3 = 0$	0.75 (0.017, 0.021)	0.70 (0.06, 0.07)	0.55 (0.21, 0.22)	0.77
$\rho_2 = \rho_3 = 0.1$	0.75 (0.015, 0.019)	0.70 (0.06, 0.07)	0.55 (0.21, 0.22)	0.77
$\rho_2 = \rho_3 = 0.8$	0.75 (0.18, 0.19)	0.70 (0.23, 0.24)	0.55 (0.38, 0.39)	0.94
$\rho_1 = 0.5, \rho_2 = 0$	AUC_1	AUC_2	AUC_3	New
$\rho_3 = -0.8$	0.75 (0.15, 0.16)	0.70 (0.35, 0.36)	0.55 (0.30, 0.31)	0.90
$\rho_3 = -0.5$	0.75 (0.12, 0.13)	0.70 (0.31, 0.33)	0.55 (0.27, 0.28)	0.88
$\rho_3 = 0$	0.75 (0.09, 0.10)	0.70 (0.28, 0.30)	0.55 (0.24, 0.25)	0.85
$\rho_3 = 0.1$	0.75 (0.09, 0.10)	0.70 (0.28, 0.29)	0.55 (0.24, 0.25)	0.85
$\rho_3 = 0.5$	0.75 (0.078, 0.081)	0.70 (0.27, 0.28)	0.55 (0.22, 0.24)	0.83
$\rho_3 = 0.8$	0.75 (0.067, 0.074)	0.70 (0.26, 0.27)	0.55 (0.22, 0.23)	0.82

The correlation coefficient between two markers is ρ . The area under the curve (AUC) of marker 1 is AUC_1 . The AUC of marker 2 is AUC_2 . The AUC of marker 3 is AUC_3 , and AUC_c is the AUC of the combined marker. 'CI (New-marker)' represents the confidence interval of the difference between the combined marker and the individual marker.

Table VI

Illustration of our procedure to combine two markers.

<i>AUC</i>	0.7426	0.7056	0.7056	0.8037	0.7130	0.8185
<i>Marker</i>	<i>4E – BP1T70</i>	<i>Ack1Y284</i>	<i>AuroraA (T288)</i>	<i>BadS136</i>	<i>c-AbIT735</i>	<i>ErbbB3/HER3(1B2)</i>
$r_{d,1,\ell_2}$	0.53	0.358	0.68	0.42	0.443	
$r_{d,1,\ell_2}$	0.022	0.220	0.41	0.046	0.03	
$r_{max,1,\ell}$	0.53	0.358	0.68	0.42	0.443	
$\rho_{1,\ell}^*$	0.236	0.382	0.187	0.236	0.606	
$ r_{max,1,\ell} - \rho_{1,\ell}^* $	0.294	0.024	0.493	0.183	0.162	

AUC, area under the curve.

Table VII

Illustration of our procedure to combine three markers.

<i>AUC</i>	0.7426	0.7056	0.8037	0.7130	0.9148
<i>Marker</i>	<i>4E – BP1T 70</i>	<i>Ack1Y 284</i>	<i>BadS136</i>	<i>c – AbIT 735</i>	<i>Combined</i>
$r_{d,1,\ell_2}$	0.0929	-0.0559	0.1519	0.1337	
$r_{d,1,\ell_2}^*$	0.2065	-0.0445	0.1906	0.0667	
$r_{max,1,\ell_2}$	0.2065	-0.0445	0.1906	0.1337	
$\rho_{1,\ell}^*$	0.9860	0.9788	0.9986	0.9788	
$ r_{max,1,\ell_2} - \rho_{1,\ell}^* $	0.7795	1.0233	0.8080	0.8451	

AUC, area under the curve.

Table VIII

Bias and mean square error of α and β for combining two markers.

	Bias		MSE	
	Proposed	Ignoring LOD	Replacing LOD	Replacing LOD
α	-0.120	0.247	0.170	0.061
β	-0.048	0.267	0.192	0.071
				0.029
				0.037

MSE, mean square error; LOD, limit of detection; 'Proposed', our proposed method; 'Ignoring LOD', replacing the measurements subjected to the LOD with 'NA'; 'Replacing LOD', replacing the measurements subjected to the LOD with d , where d is the LOD point.

Table IX

Bias and MSE of α and β for combining three markers.

	Bias			MSE		
	Proposed	Ignoring LOD	Replacing LOD	Proposed	Ignoring LOD	Replacing LOD
α	0.243	0.861	0.980	0.059	0.741	0.961
β	-0.094	0.933	0.583	0.009	0.871	0.339

MSE, mean square error; LOD, limit of detection; 'Proposed', our proposed method; 'Ignoring LOD', replacing the measurements subjected to the LOD with 'NA'; 'Replacing LOD', replacing the measurements subjected to the LOD with d , where d is the LOD point.