Research Article

Statistics in Medicine

Received XXXX

(www.interscience.wiley.com) DOI: 10.1002/sim.0000

A sequential classification rule based on multiple quantitative tests in the absence of a gold standard

Jingyang Zhang[1]*, Ying Zhang[2,3], Kathryn Chaloner[4,5] and Jack T. Stapleton[6]

In many medical applications, combining information from multiple biomarkers could yield a better diagnosis than any single one on its own. When there is a lack of a gold standard, an algorithm of classifying subjects into the case and non-case status is necessary for combining multiple markers. The aim of this paper is to develop a method to construct a composite test from multiple applicable tests and derive an optimal classification rule under the absence of a gold standard. Rather than combining the tests, we treat the tests as a sequence. This sequential composite test is based on a mixture of two multivariate normal latent models for the distribution of the test results in case and non-case groups and the optimal classification rule is derived returning the greatest sensitivity at a given specificity. This method is applied to a real data example and simulation studies have been carried out to assess the statistical properties and predictive accuracy of the proposed composite test. This method is also attainable to implement nonparametrically. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: diagnostic test; EM algorithm; mixture model; multivariate normal

1. Introduction

This work is motivated by some methodological research in HIV/AIDS studies. Some studies have observed that co-infection of a human RNA virus, GB virus C (GBV-C), can prolong the survival for HIV patients [1–8]. GBV-C is not currently known to definitely cause any disease although a recent observational study suggested a potential link between GBV-C and non-Hodgkins lymphoma [9]. Approximately 14%-43% of the individuals with HIV infection have the GBV-C viraemia [10, 11]. Some studies also found an association between GBV-C and improved response to HIV therapy [11]. The mechanism is still under investigation [12–14]. GBV-C viraemia is shown to be cleared in a great portion of patients from several months to several years [15, 16], and antibodies that are directed against the viral envelope glycoprotein 2 (E2) develop [17]. Hence the E2 antibodies are a marker of past GBV-C infection [18, 19]. An association between the

[1] *Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, U.S.A. and* [2] *Department of Biostatistics, Indiana University Fairbanks School of Public Health and School of Medicine, Indianapolis, Indiana 46202, U.S.A.* [3] *Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China.* [4] *Department of Biostatistics* [5] *Department of Statistics and Actuarial Sciences,* [6] *Department of Internal Medicine, University of Iowa, Iowa City, IA, 52242, U.S.A.*

** Correspondence to: Jingyang Zhang, Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, M2-C200, 1100 Fairview Avenue North, Seattle, WA, 98109, U.S.A. Email: jzhang2@fhcrc.org*

Prepared using simauth.cls *[Version: 2010/03/10 v3.00]*

E2 antibodies and the prolonged survival from HIV in subjects without GBV-C viraemia has also been observed [5]. To detect the presence of E2 antibodies in human serum samples, one commonly used method is through the Enzyme Linked Immunosorbent Assay (ELISA); however, there is no commercial and validated test available for the E2 antibodies. In the motivating example from Dr. Jack Stapleton's lab at the University of Iowa, a total of 100 independent blood specimens obtained from HIV infected subjects were tested by two ELISAs, which are not perfect and return a quantitative result with respect to the concentration of the E2 antibodies. The primary goal of the paper is to establish and evaluate a composite diagnostic test based on the two ELISAs in the absence of the true antibody status and any other reference test.

In diagnostic testing, a gold standard is defined as a reference test or a benchmark that is assumed 100% accurate in discriminating case from non-case. When a gold standard is available, the accuracy of a single diagnostic test has been well studied. The accuracy of a binary test is evaluated by the true positive fraction (TPF, or sensitivity) and false positive fraction (FPF, or 1-specificity). For a continuous-scale test, different binary tests can be induced by selecting different threshold values. At each threshold, a pair of TPF and FPF is obtained, and the curve that connects all pairs of TPF and FPF over all possible thresholds, which is the receiver operating characteristic (ROC) curve, is a commonly used tool to evaluate a continuous marker. These are detailed in Zhou *et. al.* [20] and Pepe [21].

When there are multiple imperfect tests available, combining them into one composite test may yield a better diagnostic test than any single test. For continuous tests, a simple case is to repeat testing on a single test. Tolley *et. al.* [22] and Murtaugh [23] consider the scenario of repeated applications of the same continuous test. At each test application, the threshold remains the same. For a set of different tests, various composite tests exist for a given overall specificity. The most straightforward way is to form a linear combination. Suppose $X \sim N(\mu_x, \Sigma_x)$ represents the test results in the case population and $Y \sim N(\mu_y, \Sigma_y)$ represents the test results in the non-case population. The linear composite rule is then based on $U = a^T X$ and $V = a^T Y$. Su and Liu [24] justifies that the linear discriminant function is the optimal linear combination that produces the maximum AUC in this case, i.e., the coefficient for the best linear combination is

$$a_0 \propto (\Sigma_x + \Sigma_y)^{-1}(\mu_y - \mu_x).$$

The linear combination is easy to implement and straightforward to interpret, however, the optimality is only guaranteed when the results are normal and homoscedastic.

Rather combining the multiple tests in parallel, we could also treat them as a sequence. Thompson [25] considers the combination of a sequence of tests. The sequence of tests can be the repeated applications of the same test on the same subject, or different tests simultaneously. The development of the sequential rule does not limit to the linear combination of multiple tests, and the application of the sequential rule on a new population does not require the practice of all tests on each subject. Two main concepts are usually used to define the sequential rule [26, 27]. The first one is "believe negative" (BN), where individuals who have negative diagnosis from any particular test will not receive subsequent tests. The other one is "believe positive" (BP), where individuals who have positive diagnosis from any particular test will not receive sub sequential tests. In this work, we will focus on the sequential rule defined by the BP approach. Thompson [25] provides the evaluation of accuracy of a sequence of tests. For two continuous tests X_1 and X_2, based on the BP rule, an individual is defined as positive if $X_1 > c_{1,p_1}$ or $X_2 > c_{2,p_2}$, where c_{i,p_i} is the p_ith percentile of the distribution of X_i in non-case population for $i = 1, 2$. The ROC curve of the sequence test as a function of an overall false positive fraction s can be expressed as (1) [25].

$$ROC_{X_1 \vee X_2}(s \mid p_1) = 1 - F_{2.1D}\left(F^{-1}_{2.1\bar{D}}\left(\frac{1-s}{p_1}\right)\right) F_{1D}(c_{1p_1}), \tag{1}$$

where $F_{2.1D}$ and $F_{2.1\bar{D}}$ are the conditional distribution functions for X_2 given $X_1 < c_{1p_1}$ in the case and non-case populations. The accuracy of the test sequence could be assessed by the MaxROC curve expressed as

$$MaxROC(s) = \max_{p_1} ROC_{X_1 \vee X_2}(s \mid p_1),$$

so that the sensitivity for any given specificity will be at least as high as that for either of the individual tests, applied on its own with the same threshold. The choice of threshold in implementation, however, is not addressed in this paper.

Both the aforementioned combinations are based on the knowledge of a gold standard, or the true case status. In practice, such information is not always available, because it may be difficult or even impossible to determine the true case status, and even the available reference test against which new tests are compared is subject to error. Kraemer [28] argues the opinion that the true case status is almost never ascertained. For ordinal or continuous-scale tests, the sensitivity and specificity are computed based on a certain classification rule with a specific threshold value, hence are dependent on the choice of the classification rule. When the true case status is unknown and there is no gold standard or even an imperfect binary reference test, like the E2 antibodies data example, a decision rule established from multiple imperfect test needs to be studied [21]. The statistical issues in diagnostic testing without a gold standard are addressed by Hui and Walter [29] mainly focusing on binary tests and summarized by Hui and Zhou [30] with many available methods for quantitative tests. Using the finite mixture model for continuous data , one could acquire the pointwise estimates of the sensitivity and specificity for a continuous-scale test over all possible threshold values by the maximum likelihood method [31]. The estimated ROC curve composed by all estimated sensitivities and specificities, however, may not retain the monotonicity, as in Figures. 2, 3, 4, 5 of [31]. Henkelman et al. [32] propose an estimation of the ROC curve of an ordinal-scale test via a mixture of multivariate normal latent model and Choi et al. [33] provide a parametric Bayesian method for a continuous-scale test under the same distributional assumption. Both methods guarantee that the estimated ROC curve is monotone. The ROC curve can also be estimated nonparametrically instead of assuming the multivariate normal distributions as proposed by Hall and Zhou [34] in which the monotonicity of the estimated ROC curve is assured without any parametric assumptions on the distributions of the test results.

The methods above primarily focus on the evaluation of diagnostic tests when there is no definitive diagnosis or a gold standard, rather on the formulation of a decision rule by combining several available continuous markers. In fact, under some assumptions, those methods could be extended to develop a decision rule from multiple continuous-scale tests. For example, Su-Liu's linear discriminant method is still applicable with the parameters in the normal distributions estimated through the maximum likelihood method using the EM algorithm [35]. Our aim in this paper is to derive an optimal composite test in a sequential way. The optimal sequential composite test is described in Section 2, and applied to the motivating ELISA data in Section 3. The statistical properties are explored through simulation studies in Section 4. We conclude this paper by discussion in Section 5.

2. Optimal sequential composite test without a gold standard

For simplicity in illustration, suppose that there are two quantitative diagnostic tests on each subject and for each test, a greater value of the result indicates a larger chance of case. Denote X_i as the random variable representing the result from test i for $i = 1, 2$ and D as the random variable indicating the case presence, with $D = 1$ meaning case present and $D = 0$ meaning case absent. Moreover, F_1 and F_0 are the joint distribution functions of $\mathbf{X} = (X_1, X_2)$ for the case and non-case populations, respectively, and f_1 and f_0 are the corresponding probability density functions.

2.1. Model setup

Suppose Test 1 is superior to Test 2 judged by a greater value of AUC. The decision rule driven by the sequential composite test is determined by a pair of cut-off values (C_1, C_2) such that:

1. if $X_1 > C_1$, then this subject is classified as positive for the study event; else,
2. if $X_2 > C_2$, then classified as positive;
3. otherwise, classified as negative.

It is theoretically equivalent to the "believe the positive" (BP) rule defined in Marshall [26] and Politser [27] given the threshold values C_1 and C_2.

Given the cut-off (C_1, C_2), the sensitivity and specificity for evaluating this composite test can be expressed as follows:

$$\begin{aligned}\text{Sensitivity} &= \Pr(\text{Positive classification}|\text{case}) \\ &= \Pr(X_1 > C_1|D=1) + \Pr(X_1 \le C_1,\ X_2 > C_2|D=1) \\ &= 1 - F_1(C_1, C_2). \qquad (2)\\ \text{Specificity} &= \Pr(\text{Negative classification}|\text{control}) \\ &= \Pr(X_1 \le C_1,\ X_2 \le C_2|D=0) \\ &= F_0(C_1, C_2). \qquad (3)\end{aligned}$$

Equations (2) and (3) are actually equivalent to (1) assuming that neither p_1 nor p_2 is pre-fixed.

We are searching for the optimal sequential composite test in the sense that it achieves the maximum sensitivity among all the sequential composite tests whose specificity is fixed at p_0. Based on (2) and (3), this task can be converted to a constrained non-linear optimization problem:

$$\min_{F_0(C_1,C_2)=p_0} F_1(C_1, C_2). \qquad (4)$$

An efficient algorithm for finding the optimal (C_1, C_2) in (4) is essential in the development of this sequential method.

2.2. Estimation and statistical inferences

2.2.1. MLE of multivariate normal model Suppose we have a sample of results from two quantitative diagnostic tests $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ that are assumed to be independent and identically distributed copies of $\mathbf{X}$ with distribution F. The implementation of all the foregoing methods requires estimation of F_1 and F_0 from observed data in the first place. Here we follow the set-up of Su and Liu's method [24] for the distribution of the tests results $\mathbf{X}$, i.e. $\mathbf{X}|D=1 \sim F_1 \equiv N(\mu_1, V_1)$ and $\mathbf{X}|D=0 \sim F_0 \equiv N(\mu_0, V_0)$. A mixture distribution of F_1 and F_0 is adopted to model the observed data, that is

$$F_\theta(\cdot) = \pi F_{1,\theta_1}(\cdot) + (1-\pi)F_{0,\theta_0}(\cdot), \qquad (5)$$

where π is an unknown parameter indicating the mixture proportion, or equivalently, the case prevalence, and $\boldsymbol{\theta} = (\pi, \theta_1, \theta_0) = (\pi, (\mu_1, V_1), (\mu_0, V_0))$ denotes the model parameters. The log-likelihood of the observed data can be expressed as:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{k=1}^{n} l_k(\boldsymbol{\theta}) = \sum_{k=1}^{n} \log f_{\boldsymbol{\theta}}(X_{1k}, X_{2k}) \\ &= \sum_{k=1}^{n} \log\left[\pi f_{1,\theta_1}(X_{1k}, X_{2k}) + (1-\pi) f_{0,\theta_0}(X_{1k}, X_{2k})\right]. \end{aligned}$$

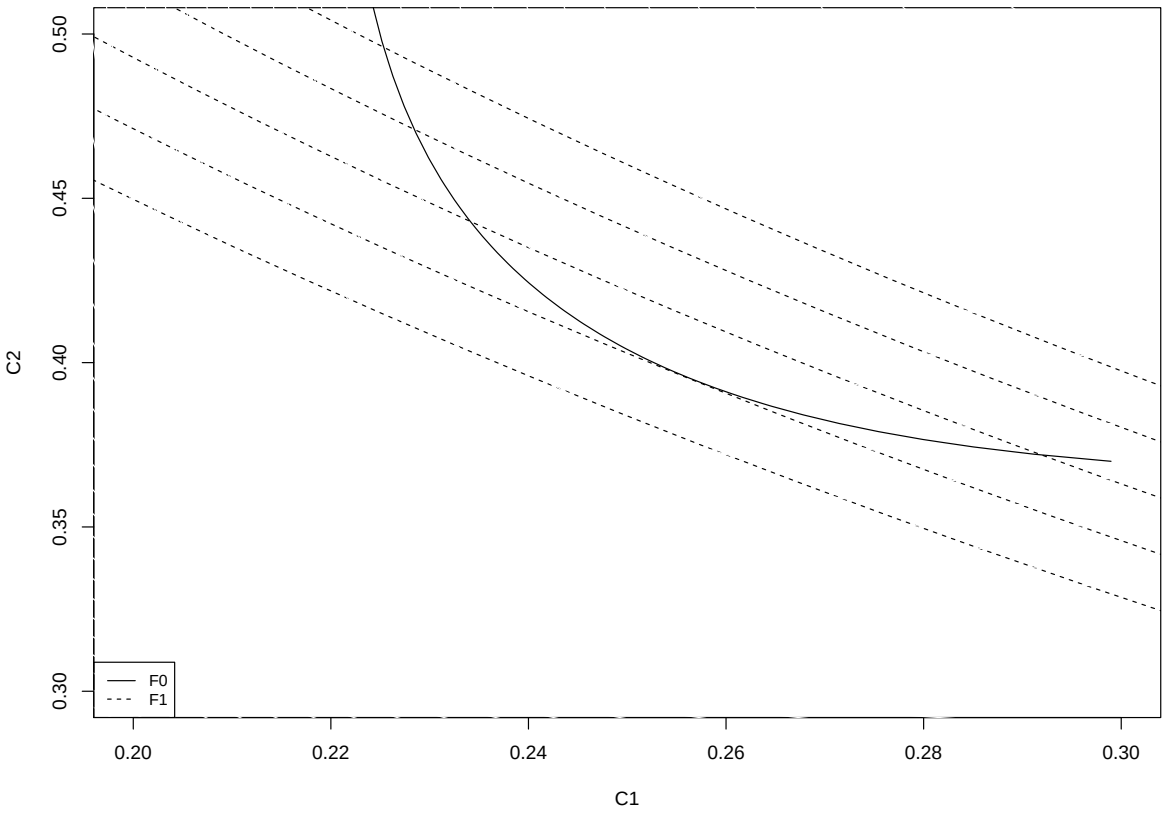

Figure 1. Illustration of the search for the optimal (C_1, C_2) at a given specificity p_0. The solid line represents the contour curve given by $F_0(C_1, C_2) = p_0$, and the dashed lines represent the contour curves given by $F_1(C_1, C_2) = t$ at various values of t.

We note that if the gold standard does exist so that the exact memberships $\mathcal{D} = (D_1, \ldots, D_n)$ are known, the log likelihood for the augmented data $\{(\mathbf{X}_1, D_1), \cdots, (\mathbf{X}_n, D_n)\}$ is given by

$$l_a(\boldsymbol{\theta}) = \sum_{k=1}^{n} D_k \log \pi f_{1,\theta_1}(X_{1k}, X_{2k}) + (1 - D_k) \log(1 - \pi) f_{0,\theta_0}(X_{1k}, X_{2k}) \tag{6}$$

and

$$Pr(D_k = 1 | (\mathbf{X}_1, \cdots, \mathbf{X}_n); \boldsymbol{\theta}) = \frac{\pi f_{1,\theta_1}(X_{1k}, X_{2k})}{\pi f_{1,\theta_1}(X_{1k}, X_{2k}) + (1 - \pi) f_{0,\theta_0}(X_{1k}, X_{2k})}.$$

Hence the MLE of the model parameters $\hat{\boldsymbol{\theta}}_n$ is easily computed using the EM algorithm [35] due to its numerical stability and algorithmic convenience for this problem. The details of the EM algorithm are provided in Appendix A.

2.2.2. Computation of the optimal sequential composite test Under the normality assumption, the feasible set of (C_1, C_2) defined by a given specificity $F_0(C_1, C_2) = p_0$ constitutes a convex contour curve [36]. When the diagnostic markers are more variant for the case subjects, it is expected that the contour given by $F_1(C_1, C_2) = t$ is also convex but with less curvature and moves towards the origin of (C_1, C_2) domain as t decreases. The optimization problem (4) can be illustrated geometrically in Figure 1.

As seen in Figure 1, the constrained optimal value t corresponds to the value given by the contour that touches the

contour of $F_0(C_1, C_2) = p_0$. The threshold vector (C_1, C_2) for the decision rule is simply the tangent point of the two contour lines and can be uniquely determined. Therefore, the original optimization problem (4) is converted to solving the system of bivariate nonlinear equations (7) for the tangent point of the contour lines of F_1 and F_0.

$$G(\mathbf{C}, \boldsymbol{\theta}) = \begin{cases} F_{0,\theta_0}(C_1, C_2) = p_0 \\ \dfrac{\partial F_{1,\theta_1}}{\partial C_1}(C_1, C_2)\dfrac{\partial F_{0,\theta_0}}{\partial C_2}(C_1, C_2) - \dfrac{\partial F_{1,\theta_1}}{\partial C_2}(C_1, C_2)\dfrac{\partial F_{0,\theta_0}}{\partial C_1}(C_1, C_2) = 0. \end{cases} \tag{7}$$

The first equation represents the constraint given by the fixed specificity and the second equation reflects that the two contour lines have the same gradient at the tangent point. The Newton-Raphson method with the step-halving line search procedure is utilized to solve the system.

Let $\hat{\mathbf{C}}_n = (\hat{C}_{1n}, \hat{C}_{2n})$ denote the solution of (7) with the MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \hat{\theta}_{0n})$, then the sensitivity is estimated by $\widehat{\text{sen}}_C = 1 - F_{1,\hat{\theta}_{1n}}(\hat{C}_{1n}, \hat{C}_{2n})$.

2.2.3. Asymptotic properties Suppose $\boldsymbol{\theta}_0$ is the true vector of the model parameters under the mixture of bivariate normal distribution. Assuming that the regularity conditions for MLE hold, it is known that as $n \to \infty$, $\hat{\boldsymbol{\theta}}_n \to_P \boldsymbol{\theta}_0$, and

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \to_d N(0, \mathcal{I}^{-1}),$$

where $\mathcal{I}$ is the Fisher information matrix given by $-E\left[\frac{\partial^2}{\partial \boldsymbol{\theta}^2} l_1(\boldsymbol{\theta}) \middle| \boldsymbol{\theta}_0\right]$ [37].

For the optimal sequential composite test, let $\mathbf{C}_0 = (C_{10}, C_{20})$ denote the solution of the system (7) under $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, then the true sensitivity is $\text{sen}_C = 1 - F_{1,\theta_{10}}(C_{10}, C_{20})$. The estimated sensitivity $\widehat{\text{sen}}_C$ is consistent and asymptotically normal under the mild condition (8) given in Theorem 2.1. The proof of the theorem is also deferred to Appendix B

Theorem 2.1 *If F_0 and F_1 are continuously differentiable with respect to $\mathbf{C} = (C_1, C_2)$ and $\boldsymbol{\theta}$ and satisfy the following inequality (8) at $\mathbf{C}_0$ and $\boldsymbol{\theta}_0$,*

$$\begin{aligned} &\left[\frac{\partial^2 F_1}{\partial C_1 \partial C_2}\frac{\partial F_0}{\partial C_2} + \frac{\partial F_1}{\partial C_1}\frac{\partial^2 F_0}{\partial {C_2}^2} - \frac{\partial^2 F_1}{\partial {C_2}^2}\frac{\partial F_0}{\partial C_1} - \frac{\partial F_1}{\partial C_2}\frac{\partial^2 F_0}{\partial C_1 \partial C_2}\right]\frac{\partial F_0}{\partial C_1} \\ &- \left[\frac{\partial^2 F_1}{\partial {C_1}^2}\frac{\partial F_0}{\partial C_2} + \frac{\partial F_1}{\partial C_1}\frac{\partial^2 F_0}{\partial C_1 \partial C_2} - \frac{\partial^2 F_1}{\partial C_1 \partial C_2}\frac{\partial F_0}{\partial C_1} - \frac{\partial F_1}{\partial C_2}\frac{\partial^2 F_0}{\partial {C_1}^2}\right]\frac{\partial F_0}{\partial C_2} \neq 0 \end{aligned} \tag{8}$$

then as sample size $n \to \infty$, $\sqrt{n}\,(\widehat{\text{sen}}_C - \text{sen}_C)$ converges to a normal distribution with mean 0 and variance given by (B.1).

Remark 2.1 *Condition (8) can be justified algebraically for bivariate normal random variables when F_1 and F_0 have a different covariance matrix. In fact, the left side is the determinant of the Jacobian matrix of (7).*

Remark 2.2 *Although the asymptotic normality holds for the estimator under fairly mild conditions, the asymptotic variance of the sensitivities is hard to estimate directly. Therefore for the inference, the standard error is estimated using the nonparametric bootstrap method [38]. Specifically, 200 samples with the same size are drawn from the original data with replacement. Each sample yields an estimated sensitivity at the given specificity from the estimated optimal sequential composite test, and the standard error is then estimated by the standard deviation of the 200 estimated sensitivities.*

3. An analysis of the ELISA data

The optimal sequential composite test based on the two ELISAs is applied to classify the 100 independent blood samples for the presence of E2 antibody in the motivating data example. Figure 2 presents a scatter plot of the results from the two ELISAs. We fit the data by a mixture of two bivariate normal distributions: $N(\mu_1, V_1)$ and $N(\mu_0, V_0)$, and obtain the MLE of the parameters as $\hat{\mu}_1 = (1.01, 0.84)^T$, $\hat{\mu}_0 = (0.16, 0.24)^T$, and

$$\hat{V}_1 = \begin{pmatrix} 0.54 & 0.22 \\ 0.22 & 0.40 \end{pmatrix}, \hat{V}_0 = \begin{pmatrix} 0.004 & 0.001 \\ 0.001 & 0.017 \end{pmatrix}.$$

Assuming the mixture of two bivariate normal models is true for this data example, the model-based estimated ROC curves are depicted in Figure 3 for the two individual ELISAs. It is apparent in the figure that Test 1 is preferred to Test 2 as it has a higher sensitivity at any prefixed specificity from their ROC curves, and hence Test 1 is utilized as the initial test for the proposed sequential composite test.

The optimal linear composite test for comparison is applicable here by extending the Su-Liu's method [24] under the assumption of normality. The technical details of the extended optimal linear composite test are referred to Zhang [39]. At the specificity of 90%, the ith sample is diagnosed as positive if $X_{1i} > 0.24$ using test 1 alone, $X_{2i} > 0.41$ using test 2 alone, $1.2X_{1i} + 0.8X_{2i} > 0.57$ under the optimal linear composite test and $X_{1i} > 0.27$ or $X_{2i} > 0.43$ under the optimal sequential composite test, where X_{1i} and X_{2i} are the results from test 1 and test 2 on the ith sample. If the future data set is the same as the data used to derive the classification rules, the classifications based on both composite decision rules are represented in Figure 2. The diagnoses from the two composite test do not disagree too much except that the linear composite test tends to attribute more samples into the E2 antibody negative group. We also plot the ROC curves for the two composite tests shown in Figure 3. Combining the two tests into a composite test does improve the discriminant capability compared to any individual test. This improvement is possible because we allow the cut-off value for each of the tests when used as a composite is different from the cut-off value when used individually. The improvement from the linear composite test is not as substantial as the sequential composite test. It appears that the sequential test is superior to the linear composite test at all values of specificity for this case. Moreover, the optimal sequential composite test only needs 53% of the blood samples for the second test at average over 1000 bootstrap samples. This implies that under the optimal sequential composite test, the probability that a patient needs to be tested by T2 is only about 50%. Hence it has a profound significance in practice when the tests are expensive or present some strong side effects.

4. Simulation studies

4.1. Simulations on the model-based estimate of sensitivity for the proposed optimal sequential composite test

In this section, we conduct simulation studies to assess the statistical properties of the model-based sensitivity for the proposed optimal sequential composite test. We generate two diagnostic markers for the case group from a bivariate normal distribution $N(\mu_1, V_1)$ of

$$\mu_1 = (3.77, 1.51)^T \text{ and } V_1 = \begin{pmatrix} 3.97 & 0.69 \\ 0.69 & 1.42 \end{pmatrix},$$

and the markers for the non-case group from a bivariate normal distribution $N(\mu_0, V_0)$ of

$$\mu_0 = (2, 0.81)^T \text{ and } V_0 = \begin{pmatrix} 0.68 & 0.03 \\ 0.03 & 0.18 \end{pmatrix}.$$

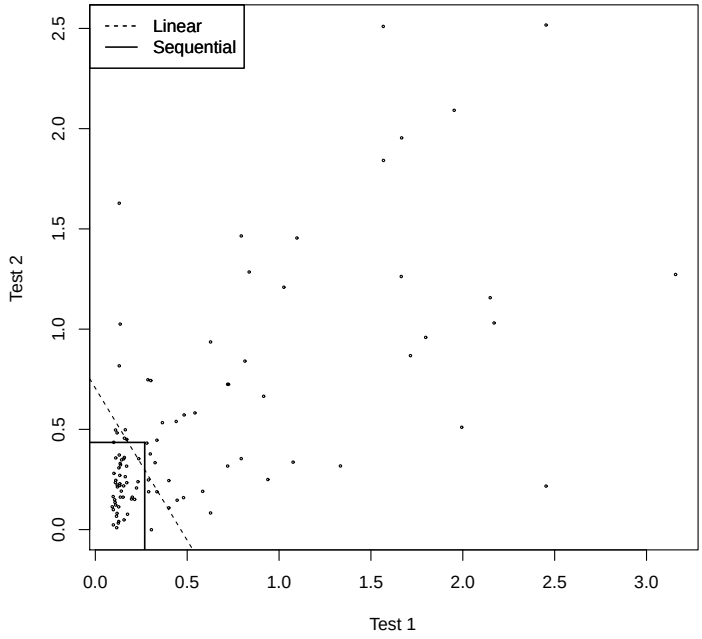

Figure 2. Results from the two tests in 100 blood samples along with the optimal linear composite test and the optimal sequential composite test at specificity $= 0.90$.

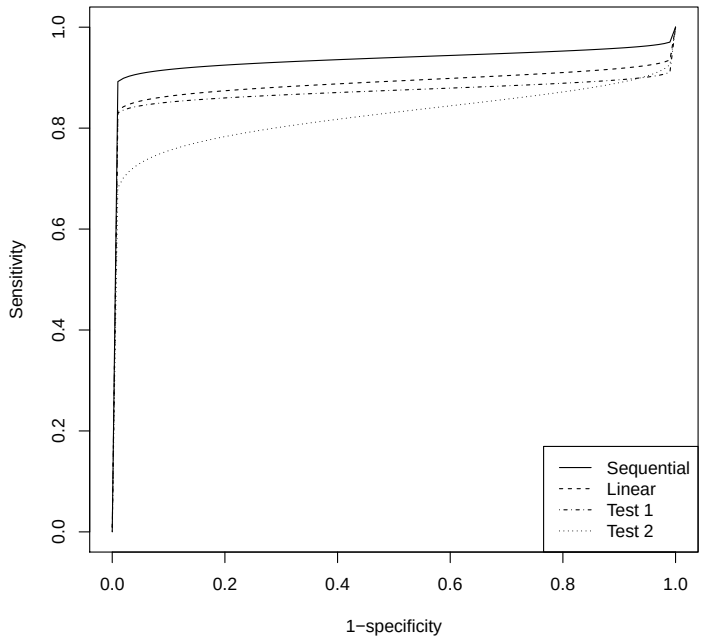

Figure 3. ROC curves for Test 2, Test 1, the optimal linear and sequential composite tests (from bottom to top).

The values of the parameters in the model are selected to mimic our motivating ELISA data example. A sample of 100 simulated data is shown in Figure 4. With the parameter values given above, the sensitivities at specificities 80% and 90% are 71% and 64%, respectively, for Test 1, and 61% and 55% for Test 2. Test 1 is superior to Test 2.

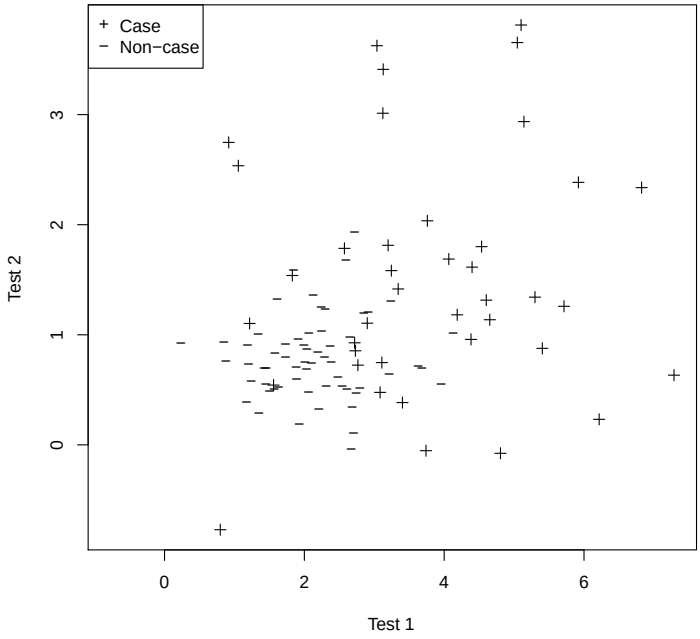

Figure 4. Scatter plot of a simulated data set of 100 subjects from the mixture model of two bivariate normal distributions with case prevalence as 0.5.

Three total sample sizes (100, 200 and 400) and two different case prevalence values (0.25 and 0.5) are examined, respectively. At each combination of sample size and case prevalence, sample data are generated from the underlying mixture of two bivariate normal models. The exact sensitivity of the optimal sequential composite test at a given specificity is computed by solving the nonlinear system (7) with the bivariate normal distribution functions F_0 and F_1, and similarly, the exact sensitivity of the optimal linear composite test is calculated using the true parameters in the bivariate normal distributions as depicted by Su and Liu [24]. The model-based sensitivities are estimated with F_1 and F_0 replaced by their MLE, $\hat{F}_0$ and $\hat{F}_1$. The standard error of the estimated sensitivity is obtained via the nonparametric bootstrap method aforementioned and its 95% Wald confidence interval is constructed using the bootstrap standard error. Subsequently, the bias, root mean square error (RMSE) and coverage probability of the 95% Wald confidence interval (CP) are calculated. In addition, the empirical sensitivity (Esen) and specificity (Espe) are assessed since the true case status is known in simulations. We repeat the Monte-Carlo simulation for 1000 times for each combination of the sample size and case prevalence, and the results are summarized in Table 1.

Indicated by Table 1, the composite tests perform generally better than an individual test and the optimal sequential composite test is superior to the optimal linear composite test in view of the sensitivities for a given specificity. Under the correct normal mixture model, as the sample size increases, the bias of the model-based sensitivity tends to be negligible and the coverage probability tends to arrive at the nominal value 95%, asserting the asymptotic properties declared by Theorem 2.1. Our simulation study indicates for a study with small sample size of 100 as in the ELISA study in Section 3, the estimated sensitivity of the derived optimal sequential composite test is fairly accurate with a negligible bias as illustrated by Table 1. The corresponding classification results are also reliable as illustrated by Table 1, because both the empirical sensitivity and specificity closely agree to their designed values. However making the model-based inference about the sensitivity needs a caution as the coverage probability is systematically lower than its target value (95%). It is also inferred by the RMSE that the estimated sensitivity may be more precise with a higher case prevalence.

Table 1. Summary of the simulation study at the given specificities based on 1000 Monte-Carlo samples in Section 4.1 with different total sample size N and different case prevalences π.

$\pi = 0.25$										
Test 1 alone										
	Specificity=80%, Sensitivity=0.705					Specificity=90%, Sensitivity=0.640				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.038	0.145	0.701	0.797	0.836	0.044	0.138	0.634	0.898	0.852
200	0.030	0.113	0.710	0.799	0.874	0.032	0.099	0.641	0.900	0.900
400	0.010	0.077	0.705	0.799	0.945	0.011	0.070	0.640	0.899	0.948
Optimal Linear Composite Test										
	Specificity=80%, Sensitivity=0.738					Specificity=90%, Sensitivity=0.682				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.058	0.149	0.742	0.793	0.741	0.064	0.141	0.685	0.892	0.771
200	0.039	0.113	0.743	0.797	0.844	0.042	0.100	0.686	0.897	0.867
400	0.014	0.074	0.739	0.798	0.942	0.014	0.066	0.684	0.898	0.946
Optimal Sequential Composite Test										
	Specificity=80%, Sensitivity=0.802					Specificity=90%, Sensitivity=0.750				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.037	0.119	0.794	0.794	0.777	0.044	0.116	0.741	0.895	0.822
200	0.026	0.092	0.800	0.798	0.856	0.029	0.084	0.747	0.899	0.872
400	0.009	0.063	0.800	0.798	0.941	0.010	0.059	0.749	0.899	0.947
$\pi = 0.5$										
Test 1 alone										
	Specificity=80%, Sensitivity=0.705					Specificity=90%, Sensitivity=0.640				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.031	0.120	0.698	0.800	0.882	0.034	0.111	0.630	0.896	0.903
200	0.018	0.084	0.708	0.799	0.929	0.019	0.072	0.642	0.899	0.939
400	0.006	0.065	0.704	0.800	0.957	0.007	0.060	0.637	0.900	0.961
Optimal Linear Composite Test										
	Specificity=80%, Sensitivity=0.738					Specificity=90%, Sensitivity=0.682				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.044	0.122	0.732	0.792	0.818	0.048	0.111	0.674	0.891	0.839
200	0.023	0.081	0.741	0.795	0.917	0.024	0.067	0.684	0.896	0.932
400	0.009	0.061	0.737	0.799	0.959	0.010	0.054	0.681	0.898	0.960
Optimal Sequential Composite Test										
	Specificity=80%, Sensitivity=0.802					Specificity=90%, Sensitivity=0.750				
N	Bias	RMSE	ESen	ESpe	CP	Bias	RMSE	ESen	ESpe	CP
100	0.029	0.098	0.788	0.795	0.848	0.034	0.091	0.730	0.893	0.894
200	0.016	0.066	0.800	0.796	0.927	0.018	0.056	0.747	0.896	0.951
400	0.007	0.053	0.800	0.799	0.947	0.007	0.048	0.747	0.898	0.961

RMSE: Root mean square error. ESen/ESpe: Empirical sensitivity/specificity. CP: 95% confidence interval coverage probability.

4.2. Simulations on the classification accuracy of the proposed sequential composite test

In our motivating example, the true GBV-C status is unknown. A series of simulations are carried out to study the accuracy and robustness of the prediction on a new dataset by the proposed composite classification rule. The simulation study is designed in the following steps:

1. Simulate a training dataset by the two settings, respectively:
   (a) The two markers follow the same mixture of two bivariate normal distributions as in Section 4.1.
   (b) The two markers follow a mixture of two Gaussian copulas with student-t marignals (4 degrees of freedom; parameters are scaled to retain the values of means and variances in the bivariate normal distributions above).

   The case prevalence is 0.5 in both settings.

2. For each simulated training set:
    - Fit the data by the bivariate normal mixture model (5).
    - Use the MLE to derive the classification rules for one single marker, the optimal linear composite test and the optimal sequential composite test under two specificities, 0.8 and 0.9.
    - Simulate one testing dataset with the same sample size and the same simulating distribution as the training data, for each of the three values of the case prevalence, 0.25, 0.5 and 0.75.
    - Apply the classification rules to each testing set and calculate the true positive fraction (TPF) and true negative fraction (TNF).
3. Repeat steps above for 1000 times, for each of the three sample sizes (100, 200 and 400).

Table 2 listed the TPFs and TNFs of the three classification rules averaged over the 1000 experiments. The test based on one single marker alone has a poor discriminating power. Both composite tests improve the accuracy substantially in terms of a higher TPF under the pre-specified specificities in the training set. The optimal sequential composite test outperforms the other two tests with the highest TPF and the accordant TPF, even when the multivariate normal assumption is violated (Setting b). Note that when normal assumption holds (Setting a), the empirical sensitivities of both composite tests in the testing set approach to the exact value as the sample size goes up.

Under both settings, different case prevalences in the new data do not affect the classification accuracy, but the optimal sequential test would be more efficient when applied to a data with a greater case prevalence since more subjects can be identified as case by one test at the first step of the test.

5. Discussion

In this paper, we develop a classification method from an alternative perspective based on multiple quantitative tests without a gold standard. The constitution of the optimal sequential composite test is statistically equivalent to the implementation of a sequence of tests discussed by Thompson [25]. Illustrated by the real data application and simulation studies, for the data of the pattern shown in Figures 2 and 4, the optimal sequential composite test demonstrates a considerable improvement in the discriminating power between case and non-case in view of the area under the ROC curve (AUC). Moreover, it has an additional advantage of engaging fewer tests. This is especially desired when the tests are costly or not applicable to all study subjects under some circumstances. The optimality of the composite test in this article is purely based on the classification accuracy without considering risk or cost associated with the tests. Some modifications of the optimizing system for the decision rule is needed if the risk or cost ought to be considered for determining an optimal decision rule in some applications.

The sequential composite test in this work uses the "believe the positive" rule based on the biological mechanism in our motivating example. It can be constructed by the "believe negative" rule accordingly in other applications. Also there has been some works in the framework of group sequential design to evaluate diagnostic tests with a gold standard [40, 41]. Further topics could be to generalize this method to the design of clinical trials.

The sequential classification method is illustrated with two tests throughout the paper but it can be similarly designed for the situation with more than two tests. It is, however, a mathematically challenging problem because finding the optimal cut-off values may not be equivalently converted to the problem of solving a nonlinear system as it does for the two-test case. The grid search is a straightforward option but it can be very numerically inefficient, especially for high dimensional data. There is still a space for improving the numerical algorithm in order to accommodate an arbitrary number of tests.

The proposed method has a fundamental assumption of multivariate normal distribution for the test results in both case and non-case groups. This assumption is likely violated in applications. In our second simulation study, when the distributional assumption is violated, the predication based on the mis-specified model is quite accurate. When the data are

Table 2. Summary of TPFs and TNFs in the simulation study in Section 4.2. The case prevalence $\pi = 0.5$ in the training set and three different values in the testing set, 0.25, 0.5, and 0.75. Two specificities (Spe), 0.8 and 0.9, are used to derive the classification rules in the training set.

Setting a

TPFs

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100	0.188	0.729	0.793	0.188	0.727	0.794	0.188	0.727	0.794
	200	0.197	0.735	0.802	0.196	0.736	0.802	0.196	0.736	0.802
	400	0.196	0.738	0.802	0.197	0.737	0.802	0.198	0.737	0.802
0.9	100	0.097	0.674	0.741	0.099	0.672	0.740	0.099	0.671	0.741
	200	0.100	0.677	0.749	0.101	0.679	0.750	0.100	0.679	0.750
	400	0.100	0.681	0.750	0.100	0.681	0.751	0.100	0.681	0.751

TNFs

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100	1.000	0.780	0.778	1.000	0.780	0.776	1.000	0.777	0.776
	200	1.000	0.788	0.787	1.000	0.787	0.786	1.000	0.786	0.785
	400	1.000	0.795	0.794	1.000	0.795	0.794	1.000	0.796	0.795
0.9	100	1.000	0.879	0.876	1.000	0.879	0.875	1.000	0.877	0.873
	200	1.000	0.888	0.886	1.000	0.888	0.887	1.000	0.887	0.886
	400	1.000	0.895	0.894	1.000	0.895	0.894	1.000	0.896	0.894

Fraction of subjects classified by one test only in the optimal sequential composite test (%)

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100			28.7			41.2			53.8
	200			27.7			40.6			53.8
	400			27.1			40.3			53.5
0.9	100			21.5			34.3			47.2
	200			20.6			33.7			47.0
	400			20.2			33.5			47.0

Setting b

TPFs

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100	0.164	0.752	0.822	0.164	0.752	0.823	0.164	0.751	0.823
	200	0.163	0.756	0.827	0.163	0.757	0.827	0.163	0.758	0.828
	400	0.159	0.771	0.838	0.161	0.770	0.838	0.161	0.771	0.839
0.9	100	0.085	0.697	0.770	0.085	0.696	0.770	0.085	0.695	0.770
	200	0.081	0.701	0.774	0.082	0.701	0.775	0.081	0.702	0.775
	400	0.077	0.718	0.788	0.077	0.717	0.788	0.077	0.717	0.788

TNFs

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100	0.998	0.814	0.793	0.998	0.814	0.793	0.997	0.813	0.792
	200	0.997	0.821	0.801	0.998	0.821	0.801	0.998	0.821	0.801
	400	0.998	0.830	0.808	0.998	0.831	0.808	0.998	0.831	0.809
0.9	100	0.999	0.888	0.863	0.999	0.887	0.864	0.999	0.885	0.862
	200	0.999	0.895	0.874	0.999	0.895	0.874	0.999	0.894	0.873
	400	0.999	0.904	0.881	0.999	0.904	0.881	0.999	0.904	0.881

Fraction of subjects classified by one test only in the optimal sequential composite test (%)

		$\pi = 0.25$			$\pi = 0.5$			$\pi = 0.75$		
Spe	N	1 marker	Linear	Sequential	1 marker	Linear	Sequential	1 marker	Linear	Sequential
0.8	100			27.7			41.6			55.6
	200			26.8			41.1			55.7
	400			26.4			41.3			56.2
0.9	100			22.2			35.8			49.5
	200			21.4			35.4			49.6
	400			21.2			35.6			50.2

not normal, we could also estimate the mixing probability distributions nonparametrically using the method developed by Hall and Zhou [34]. But Hall-Zhous's estimation method is very complicated to implement and is restrictive. The tensor spline-based sieve maximum likelihood estimation [42] of the multivariate distribution function is a compromise to the Hall-Zhou's nonparametric estimation of mixture distribution. Although the optimal sequential composite tests can still be computed with the tensor spline-based sieve estimation in principle, the numerical implementation of the test is much more demanding and challenging than the multivariate normal model. Moreover, the spline-based model would add more complexity to studying the statistical properties of the test. Study and implementation of the spline-based model for the optimal sequential composite test in this context are currently under our investigation.

Acknowledgement

We are grateful to the associate editor and the anonymous reviewer for their insightful comments.

References

1. Heringlake S, Ockenga J, Tillmann HL, Trautwein C, Meissner D, Stoll M, Hunt J, Jou C, Solomon N, Schmidt RE, *et al.*. GB Virus C/Hepatitis G Virus Infection: A Favorable Prognostic Factor in Human Immunodeficiency Virus-Infected Patients. *The Journal of Infectious Diseases* 1998; **177**(6):1723–1726, doi:10.1086/517431.
2. Lefrère JJ, Roudot-Thoraval F, Morand-Joubert L, Petit JC, Lerable J, Thauvin M, Mariotti M. Carriage of GB virus C/hepatitis G virus RNA is associated with a slower immunologic, virologic, and clinical progression of human immunodeficiency virus disease in coinfected persons. *Journal of Infectious Diseases* 1999; **179**(4):783–789, doi:10.1086/314671.
3. Tillmann HL, Heiken H, Knapik-Botor A, Heringlake S, Ockenga J, Wilber JC, Goergen B, Detmer J, McMorrow M, Stoll M, *et al.*. Infection with GB Virus C and reduced mortality among HIV-infected patients. *New England Journal of Medicine* 2001; **345**(10):715–724, doi:10.1056/NEJMoa010398.
4. Xiang J, Wünschmann S, Diekema DJ, Klinzman D, Patrick KD, George SL, Stapleton JT. Effect of coinfection with GB virus C on survival among patients with HIV infection. *New England Journal of Medicine* 2001; **345**(10):707–714, doi:10.1056/NEJMoa003364.
5. Williams CF, Klinzman D, Yamashita TE, Xiang J, Polgreen PM, Rinaldo C, Liu C, Phair J, Margolick JB, Zdunek D, *et al.*. Persistent GB virus C infection and survival in HIV-infected men. *New England Journal of Medicine* 2004; **350**(10):981–990, doi:10.1056/NEJMoa030107.
6. Xiang J, George SL, Wünschmann S, Chang Q, Klinzman D, Stapleton JT. Inhibition of HIV-1 replication by GB virus C infection through increases in RANTES, MIP-1alpha, MIP-1beta, and SDF-1. *Lancet* 2004; **363**(9426):2040–6, doi:10.1016/S0140-6736(04)16453-2.
7. Van der Bij AK, Kloosterboer N, Prins M, Boeser-Nunnink B, Geskus RB, Lange JMA, Coutinho RA, Schuitemaker H. GB virus C coinfection and HIV-1 disease progression: The Amsterdam Cohort Study. *Journal of Infectious Diseases* 2005; **191**(5):678–685, doi:10.1086/427559.
8. Zhang W, Chaloner K, Tillmann HL, Williams CF, Stapleton JT. Effect of early and late GB virus C viraemia on survival of HIV-infected individuals: a meta-analysis. *HIV Medicine* 2006; **7**(3):173–180, doi:10.1111/j.1468-1293.2006.00366.x.
9. Krajden M, Yu A, Braybrook H, Lai AS, Mak A, Chow R, Cook D, Tellier R, Petric M, Gascoyne RD, *et al.*. GBV-C/hepatitis G virus infection and non-Hodgkin lymphoma: a case control study. *International Journal of Cancer* 2010; **126**:2759–2761, doi:10.1002/ijc.25035.
10. Stapleton JT. GB virus type C/Hepatitis G virusseminars in liver disease. *Seminars in Liver Disease* 2003; **23**(2):137–148, doi:10.1055/s-2003-39943.
11. Stapleton JT, Chaloner K, Zhang J, Klinzman D, Souza IE, Xiang J, Landay A, Fahey J, Pollard R, Mitsuyasu R. GBV-C viremia is associated with reduced CD4 expansion in HIV-infected people receiving HAART and interleukin-2 therapy. *AIDS* 2009; **23**(5):605–610, doi:10.1097/QAD.0b013e32831f1b00.
12. Mohr EL, Stapleton JT. GB virus type C interactions with HIV: the role of envelope glycoproteins. *Journal of Viral Hepatitis* 2009; **16**(11):757–768, doi:10.1111/j.1365-2893.2009.01194.x.
13. Mohr EL, Xiang J, McLinden JH, Kaufman TM, Chang Q, Montefiori DC, Klinzman D, Stapleton JT. GB virus type C envelope protein E2 elicits antibodies that react with a cellular antigen on HIV-1 particles and neutralize diverse HIV-1 isolates. *Journal of Immunology* 2010; **185**(7):4496–4505, doi:10.4049/jimmunol.1001980.
14. Xiang J, McLinden JH, Rydze RA, Chang Q, Kaufman TM, Klinzman D, Stapleton JT. Viruses within the Flaviviridae decrease CD4 expression and inhibit HIV replication in human CD4+ cells. *Journal of Immunology* 2009; **183**(12):7860–7869, doi:10.4049/jimmunol.0902276.
15. Tillmann HL, Heringlake S, Trautwein C, Meissner D, Nashan B, Schlitt HJ, Kratochvil J, Hunt J, Qiu X, Lou SC, *et al.*. Antibodies against the GB virus C envelope 2 protein before liver transplantation protect against GB virus C *de novo* infection. *Hepatology* 1998; **28**(2):379–384, doi:10.1002/hep.510280213.

16. Thomas DL, Vlahov D, Alter HJ, Hunt JC, Marshall R, Astemborski J, Nelson KE. Association of antibody to GB virus C (hepatitis G virus) with viral clearance and protection from reinfection. *Journal of Infectious Diseases* 1998; **177**(3):539–542, doi:10.1086/514245.
17. Pilot-Matias TJ, Carrick RJ, Coleman PF, Leary TP, Surowy TK, Simons JN, Muerhoff AS, Buijk SL, Chalmers ML, Dawson GJ, *et al.*. Expression of the GB virus C E2 glycoprotein using the Semliki Forest virus vector system and its utility as a serology marker. *Virology* 1996; **225**(2):282–292, doi:10.1006/viro.1996.0602.
18. Dille BJ, Surowy TK, Gutierrez RA, Coleman PF, Knigge MF, Carrick RJ, Aach RD, Hollinger FB, Stevens CE, Barbosa LH, *et al.*. An ELISA for detection of antibodies to the E2 protein of GB virus C. *Journal of Infectious Diseases* 1997; **175**(2):458–461, doi:10.1093/infdis/175.2.458.
19. Tacke M, Kiyosawa K, Stark K, Schlueter V, Ofenloch-Haehnle B, Hess G, Engel AM. Detection of antibodies to a putative hepatitis G virus envelope protein. *Lancet* 1997; **349**(9048):318–320, doi:10.1016/S0140-6736(96)06461-6.
20. Zhou XH, McClish DK, Obuchowski NA. *Statistical Methods in Diagnostic Medicine*. Wiley-Interscience: New York, 2002.
21. Pepe MS. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press: Oxford, 2003.
22. Tolley EA, Somes GW, Willey ES. Determining efficacy of monitoring devices: evaluating new technologies. *Statistics in Medicine* 1991; **10**(3):351–360, doi:10.1002/sim.4780100307.
23. Murtaugh RJ. ROC curves with multiple marker measurements. *Biometrics* 1991; **51**:1514–1522.
24. Su JQ, Liu JS. Linear combination of multiple diagnostic markers. *Journal of the American Statistical Association* 1993; **88**(424):1350–1355, doi:10.1080/01621459.1993.10476417.
25. Thompson ML. Assessing the diagnostic accuracy of a sequence of tests. *Biostatistics* 2003; **4**(3):341–351, doi:10.1093/biostatistics/4.3.341.
26. Marshall RJ. The predictive value of simple rules for combining two diagnostic tests. *Biometrics* 1989; **45**:1213–22.
27. Politser P. Reliability, decision rules, and the value of repeated tests. *Medical Decision Making* 1982; **2**(1):47–69, doi:10.1177/0272989X8200200108.
28. Kraemer HC. *Evaluating medical tests: objective and quantitative guidelines*. Sage Publications: Newbury Park, 1992.
29. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. *Biometrics* 1980; **36**(1):167–171.
30. Hui SL, Zhou XH. Evaluation of diagnostic tests without gold standards. *Statistical Methods in Medical Research* 1998; **7**(4):354–370, doi:10.1177/096228029800700404.
31. Nielsen SS, Gronbaek C, Agger JF, Houe H. Maximum-likelihood estimation of sensitivity and specificity of ELISAs and faecal culture for diagnosis of paratuberculosis. *Preventive Veterinary Medicine* 2002; **53**(3):191–204, doi:10.1016/S0167-5877(01)00280-X.
32. Henkelman RM, Kay I, Bronskill MJ. Receiver operating characteristics (ROC) analysis without truth. *Medical Decision Making* 1990; **10**:24–29.
33. Choi YW, Collins M, Gardner I. Bayesian inferences for receiver opterating characteristic curves in the absence of a gold standard. *Journal of Agricultural Biological and Environmental Statistics* 2006; **11**(2):210–229, doi:10.1198/108571106X110883.
34. Hall P, Zhou XH. Nonparametric estimation of component distributions in a multivariate mixture. *The Annals of Statistics* 2003; **31**(1):201–224, doi:10.1214/aos/1046294462.
35. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 1977; **39**(1):1–38.
36. Tihansky DP. Properties of the bivariate normal cumulative distribution. *Journal of the American Statistical Association* 1972; **67**(340):903–905.
37. van der Vaart AW. *Asymptotic statistics*. Cambridge series on statistical and probabilistic mathematics, Cambridge University Press: New York, 2000.
38. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. Chapman & Hall/CRC: Boca Raton, 1994.
39. Zhang J. Making diagnoses with multiple tests under no gold standard. PhD Thesis, University of Iowa 2012.
40. Mazumdar M, Liu A. Group sequential design for comparative diagnostic accuracy studies. *Statistics in Medicine* 2003; **22**(5):727–739, doi:10.1002/sim.1386. URL http://dx.doi.org/10.1002/sim.1386.
41. Tang L, Tan M, Zhou XH. A sequential conditional probability ratio test procedure for comparing diagnostic tests. *Journal of Applied Statistics* 2011; **38**(8):1623–1632, doi:10.1080/02664763.2010.515678.
42. Wu Y, Zhang Y. Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *The Annals of Statistics* 2012; **40**(3):1609–1636, doi:10.1214/12-AOS1016SUPP.
43. Kudryavtsev LD. Implicit function. *Encyclopaedia of Mathematics*, Hazewinkel M (ed.). Springer: http://eom.springer.de/i/i050310.htm, 2001.

Appendices

A. Details of the EM algorithm in Section 2

The EM algorithm treats the exact case membership $\mathcal{D}$ as missing. Therefore, the complete data consist of $\{(\mathbf{X}_1, D_1), \cdots, (\mathbf{X}_n, D_n)\}$, and the complete-data log-likelihood is given by (6).

Let $\theta^{(i)}$ denote the estimate of θ after the ith iteration of the EM algorithm.

- **E step:** The E step computes the conditional expectation of $l_a(\theta)$ given the observed data $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ and the current estimates of θ, $\theta^{(i)}$,

$$E\left(l_a(\theta)|\theta^{(i)}\right) = \sum_{k=1}^{n} \Big\{ \Pr(D_k = 1|\theta^{(i)}) \log \pi f_1(X_{1k}, X_{2k}|\theta_1) + \Pr(D_k = 0|\theta^{(i)}) \log(1-\pi) f_0(X_{1k}, X_{2k}|\theta_0) \Big\}.$$

If we write

$$\begin{aligned} \tilde{\pi}_k^{(i)} &= \Pr(D_k = 1|\theta^{(i)}), \\ f_1^{(i)}(X_{1k}, X_{2k}) &= f_1(X_{1k}, X_{2k}|\theta_1^{(i)}), \\ f_0^{(i)}(X_{1k}, X_{2k}) &= f_0(X_{1k}, X_{2k}|\theta_0^{(i)}), \end{aligned}$$

it is easy to show that

$$\tilde{\pi}_k^{(i)} = \frac{\pi^{(i)} f_1^{(i)}(X_{1k}, X_{2k})}{\pi^{(i)} f_1^{(i)}(X_{1k}, X_{2k}) + (1-\pi^{(i)}) f_0^{(i)}(X_{1k}, X_{2k})}, \tag{A.1}$$

and

$$E\left(l_a(\theta)|\theta^{(i)}\right) = \sum_{k=1}^{n} \tilde{\pi}_k^{(i)} \log \pi f_1(X_{1k}, X_{2k}|\theta_1) + (1-\tilde{\pi}_k^{(i)}) \log(1-\pi) f_0(X_{1k}, X_{2k}|\theta_0). \tag{A.2}$$

- **M step:** The M step updates the estimate $\theta^{(i+1)}$ for θ by maximizing $E\left(l_a(\theta)|\theta^{(i)}\right)$ in (A.2) with respect to θ. We can show that $\theta^{(i+1)}$ has the following explicit expression:

$$\pi^{(i+1)} = \frac{1}{n} \sum_{k=1}^{n} \tilde{\pi}_k^{(i)}, \tag{A.3}$$

$$\mu_1^{(i+1)} = \frac{1}{n\pi^{(i+1)}} \sum_{k=1}^{n} \tilde{\pi}_k^{(i)} \mathbf{X}_k, \tag{A.4}$$

$$V_1^{(i+1)} = \frac{1}{n\pi^{(i+1)}} \sum_{k=1}^{n} \tilde{\pi}_k^{(i)} (\mathbf{X}_k - \mu_1^{(i+1)})(\mathbf{X}_k - \mu_1^{(i+1)})^T, \tag{A.5}$$

$$\mu_0^{(i+1)} = \frac{1}{n(1-\pi^{(i+1)})} \sum_{k=1}^{n} (1-\tilde{\pi}_k^{(i)}) \mathbf{X}_k, \tag{A.6}$$

$$V_0^{(i+1)} = \frac{1}{n(1-\pi^{(i+1)})} \sum_{k=1}^{n} (1-\tilde{\pi}_k^{(i)})(\mathbf{X}_k - \mu_0^{(i+1)})(\mathbf{X}_k - \mu_0^{(i+1)})^T, \tag{A.7}$$

where $\mathbf{X}_k = (X_{1k}, X_{2k})^T$.

B. Proof of Theorem 2.1

Since F_1 and F_0 are the cumulative distribution function of bivariate normal distributions, the function $G(\mathbf{C}, \boldsymbol{\theta})$ is continuously differentiable with respect to $\mathbf{C}$ and $\boldsymbol{\theta}$. Condition (8) is equivalent to the statement that the Jacobian matrix $\nabla_{\mathbf{C}} G(\mathbf{C}_0, \boldsymbol{\theta}_0)$ is invertible by deriving the determinant of $\nabla_{\mathbf{C}} G(\mathbf{C}_0, \boldsymbol{\theta}_0)$ and setting it not equal to zero. Hence, according

to the implicit function theorem [43], there exists an open set U containing $\boldsymbol{\theta}_0$, an open set V containing $\mathbf{C}_0$, and a unique continuous differentiable function $g : U \to V$ such that $\mathbf{C} = g(\boldsymbol{\theta})$ and $G(g(\boldsymbol{\theta}), \boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in U$.

Based on the MLE properties, it is known that $\hat{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_d N(0, \mathcal{I}^{-1})$. So for any $\epsilon > 0$ and $\delta > 0$, there exists an N, such that $n > N$, $\Pr(|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| > \delta) < \epsilon$. This implies that for any $n > N$, $\hat{\boldsymbol{\theta}}_n \in U$ in probability, and hence the proposed method for finding the cut-off $\hat{\mathbf{C}}_n = (\hat{C}_{n,1}, \hat{C}_{n,2})$ through solving for $G(\hat{\mathbf{C}}_n, \hat{\boldsymbol{\theta}}_n) = 0$ results in $\hat{\mathbf{C}}_n = g(\hat{\boldsymbol{\theta}}_n)$ in probability.

Further note that $F_1(\mathbf{C}, \boldsymbol{\theta}) = F_1(g(\boldsymbol{\theta}), \boldsymbol{\theta})$ is a continuously differentiable function of $\boldsymbol{\theta}$, and consequently, by the continuous mapping theorem and the delta method, we have

$$\begin{aligned}
\sqrt{n}\,(\widehat{\text{sen}}_C - \text{sen}_C) &= \sqrt{n}\left(F_1(\hat{\mathbf{C}}_n, \hat{\boldsymbol{\theta}}_n) - F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)\right) \\
&= \sqrt{n}\left(F_1(\hat{\mathbf{C}}_n, \hat{\boldsymbol{\theta}}_n) - F_1(\mathbf{C}_0, \hat{\boldsymbol{\theta}}_n) + F_1(\mathbf{C}_0, \hat{\boldsymbol{\theta}}_n) - F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)\right) \\
&= \sqrt{n}\left(\nabla_{\mathbf{C}} F_1(\mathbf{C}_0, \hat{\boldsymbol{\theta}}_n)(\hat{\mathbf{C}}_n - \mathbf{C}_0) + \nabla_{\boldsymbol{\theta}} F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right) + o_p(1) \\
&= \sqrt{n}\left(\nabla_{\mathbf{C}} F_1(\mathbf{C}_0, \hat{\boldsymbol{\theta}}_n)\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}} F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\right) + o_p(1) \\
&= \left(\nabla_{\mathbf{C}} F_1(\mathbf{C}_0, \hat{\boldsymbol{\theta}}_n)\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}} F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)\right)\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\
&\to_d N(0, B\mathcal{I}^{-1}B^T),
\end{aligned}$$

where

$$B = \nabla_{\mathbf{C}} F_1(\mathbf{C}_0, \boldsymbol{\theta}_0)\nabla_{\boldsymbol{\theta}} g(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}} F_1(\mathbf{C}_0, \boldsymbol{\theta}_0). \tag{B.1}$$