7-11-2003

Adjusting for Non-Ignorable Verification Bias in Clinical Studies for Alzheimer's Disease

Xiao-Hua Zhou
University of Washington, azhou@u.washington.edu

Pete Castelluccio
Purdue University, pcastell@purdue.edu

1. INTRODUCTION

Two-stage designs have commonly been used in studies of Alzheimer's disease (AD) because of the relatively low prevalence of the disease and high cost of ascertaining a diagnosis. In the first stage, a large sample of subjects are screened for AD with inexpensive screening tests. In the second stage, a subset of the screened subjects are selected for the clinical diagnosis by a more extensive but more expensive clinical assessment. The probability of selection for the clinical assessment depends on the screening result from the first stage, as well as other factors, such as subject's age. Furthermore, not all selected subjects have the clinical diagnosis because some subjects may refuse and others are unable to be clinically assessed.

For example, in the motivational example, as reported in Zhou and Higgs (2000), Hendrie and his colleagues employed a two-stage design to study the role of environmental risk factors for development of dementia in two black populations in Indianapolis, USA, and Ibadan, Nigeria. The study used a two-stage design. The first stage consisted of screening a large sample of subjects for dementia with the screening instruments, and the second stage consisted of selecting a small number of the screened subjects for the clinical diagnosis of dementia by a clinical assessment, which consisted of a neurological examination, a neuropsychological test battery, laboratory tests, CT scans, and a detailed interview with a relative of the subject. Selection for the clinical assessment was based on the score of the screening instruments and the age of a subject. All subjects who were screened into the 'poor performance' category were selected for the clinical assessment; a subject who was screened into the 'intermediate performance' category had the 50% chance of being selected for the clinical assessment; and a subject who was screened into the 'good performance' group had the 5% chance of being selected for the clinical assessment. In addition, a stratified random sample was taken in the 'good performance' group, selecting 75% from those age 75 years and older in order to have enough older subjects. The clinical diagnosis, however, was not obtained for all subjects who were selected for disease verification because

3

some subjects were unable to be clinically assessed, while others refused.

We are interested in comparing the accuracy of a new screening test (T_1) with that of a standard screening test (T_2) for subjects age 75 years or older. The standard test is based on results from the cognitive test only; and the new test uses information from the cognitive assessment given to the subject as well as informant data from someone who knows the subject. Table 1 displays the resulting classification data in Indianapolis and Ibadan for subjects who were 75 years or older.

TABLE 1 GOES HERE

It has been shown that inferences on the accuracy of screening tests that use only verified cases may result in verification bias. Under the ignorable verification bias assumption that the probability of selection for verification depends only on the observed data, Zhou and Higgs (2000) proposed a maximum likelihood (ML) approach for comparing the accuracies of two tests while correcting for verification bias.

However, if the verification mechanism depends on unobserved variables that are related to the disease status, the verification mechanism is not ignorable. For example, if selected subjects who refuse to be verified are likely demented, the verification mechanism most likely depends on the unobserved disease status and hence is non-ignorable. If the verification mechanism is non-ignorable, the method proposed in Zhou and Higgs (2000) cannot be applied to assess the relative accuracy of two screening tests. Zhou and Castelluccio (2003) have developed a general approach for comparing the relative accuracy of two screening tests in the presence of non-ignorable verification bias.

In this paper, we apply the method of Zhou and Castelluccio (2003) to comparing the relative accuracy of two screening tests in a two-stage design study in the presence of non-ignorable verification bias. This paper is organized as follows. In Section 2, we introduce notation and parameters of interest. In Section 3, we propose a particular model for the non-ignorable verification mechanism and discuss an estimation method under the assumed non-ignorable verification bias model. In Section 4, we contrast

4

the difference between ML estimates under the ignorable verification bias assumption and ones derived under a non-ignorable verification bias model and study effects of non-ignorability on ML estimates for ROC curve areas. In Section 5, we present results from the analysis of our AD data set using the proposed methods.

2. Notation and Parameters of Interest

We denote ordinal-scale responses of two tests by T_1 and T_2, ranging from 1 to K. We let D and V denote two indicator variables for the true disease status and disease verification of a patient, respectively. Here $D = 1$ for a diseased patient and 0 otherwise, and $V = 1$ for a verified patient and 0 for a non-verified patient. With such notation, we can display the observed data in Table 2.

Table 2 goes there

We let A_1 and A_2 denote the areas under the ROC curves of the screening test 1 and screening test 2, respectively, under the trapezoidal rule. If we let $\phi_{1jk} = P(T_1 = j, T_2 = k)$ and $\phi_{2jk} = P(D = 1 \mid T_1 = j, T_2 = k)$, we can show that A_1 and A_2 are functions of ϕ_{1jk} and ϕ_{2jk}:

$$A_1 = \frac{1}{\gamma(1-\gamma)}\Big[\sum_{j=1}^{K-1}\sum_{k=1}^{K}(1-\phi_{2jk})\phi_{1jk}\sum_{l=j+1}^{K}\sum_{k=1}^{K}\phi_{2lk}\phi_{1lk} + (1/2)\sum_{j=1}^{K}\sum_{k=1}^{K}(1-\phi_{2jk})\phi_{1jk}\sum_{k=1}^{K}\phi_{2jk}\phi_{1jk}\Big], \quad (1)$$

and

$$A_2 = \frac{1}{\gamma(1-\gamma)}\Big[\sum_{k=1}^{K-1}\sum_{j=1}^{K}(1-\phi_{2jk})\phi_{1jk}\sum_{l=k+1}^{K}\sum_{j=1}^{K}\phi_{2jl}\phi_{1jl} + (1/2)\sum_{l=1}^{K}\sum_{j=1}^{K}(1-\phi_{2jl})\phi_{1jl}\sum_{j=1}^{K}\phi_{2jl}\phi_{1jl}\Big], \quad (2)$$

respectively, where $\gamma = P(D = 1) = \sum_{j,l=1}^{K}\phi_{1jl}\phi_{2jl}$.

Our statistical problem is to perform hypothesis testing about the two ROC curve areas and to construct a confidence interval for the difference in two ROC areas, $A_1 - A_2$.

3. Estimation Procedure

Without the ignorable verification bias assumption, we have to model the verification mechanism to draw a valid inference about the ROC curves and their areas of two screening tests. Let $\boldsymbol{W}$ be a (p+1) dimensional vector of covariates, derived from the test results T_1 and T_2. Zhou and Castelluccio (2002) have proposed the following non-ignorable verification bias model:

$$\text{logit} P(V = 1 \mid T_1, T_2, D) = \boldsymbol{W}'\boldsymbol{\beta} + \alpha D, \tag{3}$$

where $\boldsymbol{\beta}$ is a column vector of $p+1$ parameters, and α is a scalar parameter.

Because of possible several local maximum likelihood estimates and boundary solutions, Zhou and Castelluccio (2002) proposed a profile maximum likelihood method for finding the global ML estimates. We next introduce this profile ML approach.

For the problem under consideration, the observed data consist of the two test results and true disease status for a verified patient and only the two test results for an unverified patient. Denote $n = \sum_{j,l=1}^{K} n_{jl}$, $\boldsymbol{\phi}_1 = (\phi_{111}, \ldots, \phi_{1(K-1)K})'$, $\boldsymbol{\phi}_2 = (\phi_{211}, \ldots, \phi_{2KK})'$, and $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2')'$. Let V_i, $T_1^{(i)}$, $T_2^{(i)}, D_i$ be the values of V, T_1, T_2, and D for the ith patient, respectively. Let $\sum_{V_i=1}$ denote a sum over all cases in which $V_i = 1$ and let $\sum_{V_i=0}$ denote a sum over all cases in which $V_i = 0$.

Without the ignorable verification bias assumption, a valid likelihood needs to be based on the data $V, D, \mathbf{W}$ if V=1 and $V, \mathbf{W}$ if V=0, and the resulting log-likelihood function is given by

$$l(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{j,l=1}^{K} [s_{jl} \log(\psi_{1jl}\phi_{2jl}) + r_{jl} \log(\psi_{0jl}(1-\phi_{2jl})) + u_{jl} \log((1-\psi_{1jl})\phi_{2jl} + (1-\psi_{0jl})(1-\phi_{2jl})) + n_{jl} \log \phi_{1jl}], \tag{4}$$

where

$$\psi_{1jl} = P(V = 1 \mid T_1 = j, T_2 = l, D = 1) = \exp(\boldsymbol{W}_{jl}'\boldsymbol{\beta} + \alpha)/(1 + \exp(\boldsymbol{W}_{jl}'\boldsymbol{\beta} + \alpha)),$$

and

$$\psi_{0jl} = P(V = 1 \mid T_1 = j, T_2 = l, D = 0) = \exp(\boldsymbol{W}_{jl}'\boldsymbol{\beta})/(1 + \exp(\boldsymbol{W}_{jl}'\boldsymbol{\beta})).$$

If we denote

$$l_1(\boldsymbol{\phi}_1) = \sum_{j,l=1}^{K} n_{jl} \log \phi_{1jl}$$

and

$$l_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2) = \sum_{j,l=1}^{K} [s_{jl} \log(\psi_{1jl}\phi_{2jl}) + r_{jl} \log(\psi_{0jl}(1-\phi_{2jl})) + u_{jl} \log((1-\psi_{1jl})\phi_{2jl} + (1-\psi_{0jl})(1-\phi_{2jl}))],$$
(5)

we can write $l(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}) = l_1(\boldsymbol{\phi}_1) + l_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2)$. Therefore, maximizing $l_1(\boldsymbol{\phi}_1)$ with respect to $\boldsymbol{\phi}_1$, we obtain the ML estimates for ϕ_{1jl} as $\hat{\phi}_{1jl} = n_{jl}/n$. We can obtain the ML estimates for the remaining parameters, α, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}_2$, by maximizing the log-likelihood $l_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2)$.

By considering (V, D, W) as complete-data, we can use the expectation-maximization (EM) algorithm to find ML estimates. Because the log-likelihood may have the several local maxima and boundary solutions, the direct use of the EM algorithm may fail to find the global ML estimates. If we knew the value of α, it would be easier to estimate the remaining parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\phi}_2$. Therefore, we propose to combine the profile method and the EM algorithm to find the global ML estimates for α, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}_2$.

Specifically, we choose a set of α values. At each of the selected α values, we use the EM algorithm to find the ML estimates $\hat{\boldsymbol{\beta}}(\alpha)$ and $\hat{\boldsymbol{\phi}}_2(\alpha)$, and we compute the profile likelihood for α as $l_P(\alpha) = l(\alpha, \hat{\boldsymbol{\beta}}(\alpha), \hat{\boldsymbol{\phi}}_2(\alpha))$, where l is given by Equation 4. Then, we find the value of α that maximizes the profile likelihood with respect to α and the corresponding values for $\hat{\boldsymbol{\beta}}(\alpha)$ and $\hat{\boldsymbol{\phi}}_2(\alpha)$. See Zhou and Castellucio (2003) for a detailed description on the profile EM method.

4. Likelihood-based inferences

After obtaining the ML estimators $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\phi}}$ for α, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}$, we can derive the ML estimators for the non-parametric ROC curves of the two tests by substituting the unknown parameters in (1) and (2) with their ML estimators.

We may also use the likelihood-based method for testing the null hypothesis that $H_0 : A_1 = A_2$ and for constructing confidence intervals for the difference $A_1 - A_2$. By substituting the ML estimates into (1) and (2), we obtain the ML estimators, $\hat{A}_1$ and $\hat{A}_2$, for the areas under the ROC curves of the screening tests 1 and 2, respectively.

To find the asymptotic covariance matrix of $\hat{A}_1$ and $\hat{A}_2$, we first compute the expected Fisher information matrix for α, $\boldsymbol{\beta}$, $\boldsymbol{\phi}_1$, and $\boldsymbol{\phi}_2$. Observe that we can write $l(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ as the sum of $l_1(\boldsymbol{\phi}_1)$ and $l_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2)$ and that the vectors $\boldsymbol{\phi}_1$ and $(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2)$ are distinct. Therefore, the expected Fisher information matrix for $(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ is $\mathrm{diag}(I_1(\boldsymbol{\phi}_1), I_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2))$, where $I_1(\boldsymbol{\phi}_1)$ and $I_2(\boldsymbol{\phi}_2)$ are the expected Fisher information matrices for $l_1(\boldsymbol{\phi}_1)$ and $l_2(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}_2)$, respectively. Then, using the delta method (Agresti, pp. 56-58, 1990) we find a consistent estimator for the covariance matrix of $\hat{A}_1$ and $\hat{A}_2$, and denote its (k, j)th element by $\hat{\nu}_{kl}$. Using the statistic,

$$ Z = (\hat{A}_1 - \hat{A}_2)/\sqrt{\hat{\nu}_{11} + \hat{\nu}_{22} - 2\hat{\nu}_{12}}, $$

we can test the null hypothesis that $A_1 = A_2$, basing the large-sample normal distribution of ML estimators. Similarly we can derive a $(1 - \alpha)\%$ confidence interval for $A_1 - A_2$ as

$$ [\hat{A}_1 - \hat{A}_2 - z_{1-\alpha/2}\sqrt{\hat{\nu}_{11} + \hat{\nu}_{22} - 2\hat{\nu}_{12}}, \hat{A}_1 - \hat{A}_2 + z_{1-\alpha/2}\sqrt{\hat{\nu}_{11} + \hat{\nu}_{22} - 2\hat{\nu}_{12}}], $$

where $z_{1-\alpha}$ is the $(1 - \alpha/2)100$th percentile of the standard normal distribution.

5. Comparison of ML estimates with and without the ignorable assumption

From Section 3 we see that without the ignorable verification bias assumption, the valid log-likelihood function is given by

$$ l_{NonMAR}(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{j,l=1}^{K} [s_{jl}\log(\psi_{1jl}\phi_{2jl}) + r_{jl}\log(\psi_{0jl}(1-\phi_{2jl})) + u_{jl}\log((1-\psi_{1jl})\phi_{2jl} + (1-\psi_{0jl})(1-\phi_{2jl}))] + $$

$$ \sum_{j,l=1}^{K} n_{jl}\log\phi_{1jl}], \tag{6} $$

8

where $\psi_{1jl} = P(V = 1 \mid T_1 = j, T_2 = l, D = 1) = \psi_{1jl}(\alpha, \beta), and \psi_{0jl} = P(V = 1 \mid T_1 = j, T_2 = l, D = 0) = \psi_{0jl}(\alpha, \beta).$

If we assume the verification bias is ignorable (e.g. $\alpha = 0$), we obtain the following valid log-likelihood function,

$$l_{MAR}(\boldsymbol{\beta}, \boldsymbol{\phi}) = \sum_{j,l=1}^{K} [s_{jl} \log(\phi_{2jl}) + r_{jl} \log(1 - \phi_{2jl})] +$$

$$\sum_{j,l=1}^{K} [(s_{jl} + r_{jl}) \log(\psi_{jl}) + u_{jl} \log(1 - \psi_{jl}) + n_{jl} \log \phi_{1jl}], \tag{7}$$

where $\psi_{jl} = P(V = 1 \mid T_1 = j, T_2 = l).$

From (6) and (7), we see that the ML estimate for ϕ_{1jl} is equal to the same expression, n_{jl}/n, with or without the ignorable verification bias assumption. However, the ML estimates for ϕ_{2jl}'s will be different, depending on whether we assume the ignorable verification bias. Under the ignorable verification bias assumption, the valid log-likelihood can be written as three separate terms, the first one involving only the parameters ϕ_{2jl}'s, the second one involving only the parameters ϕ_{1jl}'s, and the third one involving only the verification mechanism parameters ψ_{jl}'s. Therefore, the ML estimates for ϕ_{2jl}'s under the ignorable verification bias assumption do not depend on ψ_{jl}'s, the parameters for the verification bias mechanism. However, without the ignorable verification bias assumption, the valid log-likelihood cannot be separated into a term that involves only the parameters ϕ_{2jl}' and a different term that involves only the verification bias mechanism parameters. Therefore, the ML estimates for ϕ_{2jl}'s will depend on the estimated values of the parameters α and β in a non-ignorable model. Hence, although the ROC curve area does not depend on model parameters α and $\boldsymbol{\beta}$ for the verification mechanism and depends only on ϕ_{1jl}'s and ϕ_{2jl}'s, without the ignorable verification bias assumption, the parameters α and $\boldsymbol{\beta}$ can still affect the ROC curve area through their effects on ϕ_{2jl}; sometime their effects can be dramatic. We will further discuss this point when we analyze our motivating data set.

6. Results on Screening for Alzheimer's Disease Data

Let A_1 and A_2 be the areas under the ROC curves of the new screening test and the standard screening test in a particular site (Indianapolis or Ibadan), respectively. We wish to know whether the two screening tests are same in a particular site (Indianapolis and Ibadan).

We proposed the following model for the non-ignorable verification mechanism:

$$\text{logit}P(V = 1 \mid W, D) = \beta_0 + \beta_1 I_{[T_1=1]} + \beta_2 I_{[T_1=2]} + \beta_3 I_{[T_2=1]} + \beta_4 I_{[T_2=2]} + \alpha D, \tag{8}$$

This model assumes that no interactions between T_1 and T_2 can affect the verification mechanism.

To apply the profile likelihood approach described in Section 3, we need to choose the set for α. We chose this set to be between -8 and +8, with an increment value of 0.5; that is, $\alpha = -8 + 0.5 * i, i = 1, \ldots, 32$. With this chosen set, we used the profile approach described in Section 3 to find the ML estimates for α, β, and ϕ. To avoid a computational problem due to too many zero cells in the data, we added a small number of 0.3 to the data given in Table 1 before we ran our profile likelihood algorithm, and we summarized the resulting profile ML estimates for α, β, and ϕ in Table 3.

<div align="center">Table 3 goes here</div>

Replacing unknown ϕ by their ML estimates in Equations 1 and 2, we obtained the ML estimates for the ROC curve areas of the two screening tests, separately for Indianapolis and Ibadan sites. We then calculated the associated covariance matrices by Fisher's information matrix. We summarized those results in Table 4, which also included 95% confidence intervals for the differences between two ROC curve areas.

<div align="center">Table 4 goes here</div>

For comparison purposes, we also included the inference results under the ignorable verification bias assumption in Table 4.

To assess the goodness-of-fit of this non-ignorable verification bias model (8), we derived the following Pearson's goodness-of-fit statistic. We assume that our data, given in Table 1, follow a multinomial

<div align="center">10</div>

distribution with the cell probabilities, $\pi_{jl1} = P(T_1 = j, T_2 = l, D = 1, V = 1)$, $\pi_{jl2} = P(T_1 = j, T_2 = l, D = 0, V = 1)$, and $\pi_{jl3} = P(T_1 = j, T_2 = l, V = 0)$. Under the proposed non-ignorable verification mechanism model, we see that the cell probabilities π_{jlk}'s are functions of the parameters, α, $\boldsymbol{\beta}$, $\boldsymbol{\phi}_1$, and $\boldsymbol{\phi}_2$, and we can write them as $\pi_{jlk}(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi})$. The total number of unknown parameters in the model is $p + 2K^2$. Let $\hat{\pi}_{jlk} = \pi_{jlk}(\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\phi}})$ be the ML estimate of $\pi_{jlk}(\alpha, \boldsymbol{\beta}, \boldsymbol{\phi})$, where $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\phi}}$ are the ML estimates of α, $\boldsymbol{\beta}$, and $\boldsymbol{\phi}$, respectively, under the non-ignorable verification bias model. For a formal goodness-of-fit test, we proposed the following Pearson's test statistic:

$$\chi^2_{obs} = \sum_{j,l=1}^{K} \left[\frac{(s_{jl} - n\hat{\psi}_{1jl}\hat{\phi}_{1jl}\hat{\phi}_{2jl})^2}{n\hat{\psi}_{1jl}\hat{\phi}_{1jl}\hat{\phi}_{2jl}} + \frac{(r_{jl} - n\hat{\psi}_{0jl}\hat{\phi}_{1jl}(1 - \hat{\phi}_{2jl}))^2}{n\hat{\psi}_{0jl}\hat{\phi}_{1jl}(1 - \hat{\phi}_{2jl})} \right] +$$

$$\sum_{j,l=1}^{K} \frac{[u_{jl} - n(\hat{\psi}_{1jl}\hat{\phi}_{1jl}\hat{\phi}_{2jl} + \hat{\psi}_{0jl}\hat{\phi}_{1jl}(1 - \hat{\phi}_{2jl}))]^2}{n(\hat{\psi}_{1jl}\hat{\phi}_{1jl}\hat{\phi}_{2jl} + \hat{\psi}_{0jl}\hat{\phi}_{1jl}(1 - \hat{\phi}_{2jl}))}.$$

From Table 3 we see the non-ignorable verification mechanism operates slightly different in the Indianapolis and Ibadan sites. From the values of the goodness-of-fit statistics, we concluded that the proposed non-ignorable verification bias models fit the data well in both Indianapolis and Ibadan sites.

From Table 4 we also see that the ML estimates for the ROC curve areas can be dramatically different, depending on whether the ignorable verification bias is assumed. For example, the estimated ROC curve area for the first screening test in the Indianapolis site is 0.87 under the ignorable verification bias assumption and is reduced to 0.69 under the non-ignorable verification bias model. Similarly, in the Ibadan site, the estimated ROC curve area of the first screening test derived under the ignorable verification bias assumption is much lower than the one derived under the non-ignorable verification bias model. We also note that the variance and covariance estimates under the non-ignorable verification bias model tend to be larger than the ones derived under the ignorable verification bias assumption; this phenomena is what we expect since we have more parameters to estimate under a non-ignorable verification bias model than under the ignorable verification bias assumption.

However, the conclusion on the significant difference between the two ROC curve areas is unchanged,

11

regardless whether we assume the ignorable verification bias or use a non-ignorable verification bias model. To assess the sensitivity of adding a constant to our data, we have tried to add different small constants, and our results did not change.

7. DISCUSSION

In two-stage studies of Alzheimer's disease, the verification bias is a serious problem in estimation of the accuracy of screening tests.

In this paper, we analyzed a particular data set from a two-stage study of dementia in two different sites under both the ignorable verification bias assumption and a non-ignorable verification bias model. The goodness-of-fit statistics suggest that the proposed models fit the data well. From the data analysis results, we concluded that the two screening tests for Alzheimer's disease are not statistically significant in both the Indianapolis and Ibadan sites. The accuracy of the standard screening test is higher in Indianapolis than in Ibadan.

The current paper has one limitation due to the limitation of the available data. There are three reasons why subjects did not have unverified disease status: (1) they were not selected to undergo verification; (2) they were selected but were too frail to undergo verification; and (3) they were selected but refused to undergo verification. Unfortunately, since we do not have information on which category a unverified subject belongs to, we have lumped the three categories into one unverified group. It is likely that the verification mechanism for those subjects who were not selected for verification would be ignorable. If we had had such data, we would have build a better model for the verification mechanism that would have specified a non-ignorable verification mechanism for unverified subjects who were selected but either too frail or refused to undergo verification and would have specified a different ignorable verification mechanism for unverified subjects who were not selected to undergo verification.

It is worth to note that the current paper jointly modelled the non-ignorable parameter α and out-

come data, and the identifiability of α comes from the parametric model assumption for the verification bias mechanism that is unverifiable. An alternative approach to deal with non-ignorable verification bias is to apply the semi-parametric sensitivity analysis models proposed by Scharfstein et al (2002) for non-ignorable drop-outs. It is a future research topic to compare these two approaches.

ACKNOWLEDGMENTS

REFERENCES

Baker, S. G. (1995). Evaluating multiple diagnostic tests with partial verification. *Biometrics* **51**, 330-337.

Scharfstein, D. O., Rotnitzky, A., and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of American Statistical Association* **94**, 1096-1120.

Zhou, X. H. (1993). Maximum likelihood estimators of sensitivity and specificity corrected for verification bias. *Communications in Statistics-Theory and Methods* **22**, 3177-3198.

Zhou, X. H. and Rodenberg, C. A. (1998). Estimating a ROC curve in the presence of verification bias. *Communications in Statistics-Theory and Methods* **27**, 635-657.

Zhou, X. H. and Higgs , R. E. (2000). Assessing the relative accuracies of two screening tests in the presence of verification bias. *Statistics in Medicine* **19**, 1697-1705.

Zhou, X. H. and Castelluccio , P. (2003). Nonparametric analysis for the ROC curves of two diagnostic tests in the presence of nonignorable verification bias. *Journal of Statistical Planning and Inferences*, 193-213.

Table 1: Cross-classification of two screening tests

		$T_1 = 1$			$T_1 = 2$			$T_1 = 3$		
		$T_2 = 1$	$T_2 = 2$	$T_2 = 3$	$T_2 = 1$	$T_2 = 2$	$T_2 = 3$	$T_2 = 1$	$T_2 = 2$	$T_2 = 3$

Indianapolis Site

		$T_2=1$	$T_2=2$	$T_2=3$	$T_2=1$	$T_2=2$	$T_2=3$	$T_2=1$	$T_2=2$	$T_2=3$
V=1	D=0	37	4	10	14	0	9	8	2	25
	D=1	0	0	0	1	0	3	5	0	31
V=0		287	45	53	11	3	12	6	0	22
Total		324	49	63	26	3	24	19	2	78

Ibadan Site

		$T_2=1$	$T_2=2$	$T_2=3$	$T_2=1$	$T_2=2$	$T_2=3$	$T_2=1$	$T_2=2$	$T_2=3$
V=1	D=0	33	17	25	3	5	38	2	3	111
	D=1	0	1	1	0	0	2	0	0	17
V=0		146	85	119	4	3	37	1	1	47
Total		179	103	145	7	8	77	3	4	175

Table 2: Observed Data

		$T_1 = 1$			...	$T_1 = K$		
		$T_2 = 1$	...	$T_2 = K$	...	$T_2 = 1$	...	$T_2 = K$
V=1	D=1	s_{11}	...	s_{1K}	...	s_{K1}	...	s_{KK}
	D=0	r_{11}	...	r_{1K}	...	r_{K1}	...	r_{KK}
V=0		u_{11}	...	u_{1K}	...	u_{K1}	...	u_{KK}
Total		n_{11}	...	n_{1K}	...	n_{K1}	...	n_{KK}

Table 3: The ML estimates for α and β in the non-ignorable model and the goodness-of-fit

α	β_0	β_1	β_2	β_3	β_4	Goodness-of-fit
			Indianapolis Site			
-2.86	3.27	-4.59	-1.68	-0.58	-0.42	0.188
			Ibadan Site			
-2.84	2.14	-3.18	-1.43	-0.34	0.14	0.119

14

Table 4: The ML estimates under the ROC curve areas of the screening tests in Indianapolis and Ibadan sites

Verification model	A_1	$\text{Var}(A_1)$	A_2	$\text{Var}(A_2)$	$\text{Cov}(A_1, A_2)$	95% Normal CI
Indianapolis Site						
The ignorable assumption	0.87	0.0036	0.78	0.0019	0.0011	(-0.022, 0.203)
The non-ignorable model	0.69	0.0890	0.70	0.0414	0.054	(-0.303, 0.283)
Ibadan Site						
The ignorable assumption	0.68	0.0067	0.61	0.0032	0.0024	(-0.070, 0.210)
The non-ignorable model	0.53	0.0295	0.54	0.0112	0.0096	(-0.297, 0.277)