

# Relative accuracy of the screening tests when absolute determination of the disease status is not feasible

## Abstract

Comparison of the accuracy of a new screening test to that of the standard one can be implemented by administering both screening tests to a group of asymptomatic subjects for which the disease status can be determined using a gold standard (GS) test. Nevertheless, the GS test may be too costly or too invasive hence unethical to administer to all the study subjects, including those who screen negative on all the screening tests. When this is the case, relative accuracy of the two screening tests can be estimated when a randomized paired screen positive (RPSP) design is used to collect the data. However, this design contains cells with missing data, thus the likelihood function is not available. The objective of this study is to demonstrate a parsimonious way of estimating relative accuracy of the screening tests when the determination of the disease status of the subjects who screened negative on the new and the standard screening tests was not conducted due to ethical concerns.

Markov Chain Monte Carlo simulation technique is used to parsimoniously address the aforementioned shortcoming of the RPSP design when subjective approach to estimation is used. Multiple data imputation using Gibbs sampler is performed. Monte Carlo point and interval estimates of the missing data, measures of accuracy and the relative rates are computed when the tests are treated to be:

I. Conditionally independent

II. Dependent

The accuracy when the tests are dependent is higher than when the tests are conditionally independent. This is apparent from the narrower credible bounds and smaller standard errors of the true and false positive rates under this setting. Secondly, the estimates obtained under the two assumptions are better than when the missing data are ignored. Whether these differences were true or just due to chance warrant statistical investigation under different assumptions.

**Keywords:** conditional independence, dependence, gibbs sampling, markov chain monte carlo, subjective approach, credible bounds

Volume I Issue 3 - 2014

Keter K Alfred,<sup>1</sup> Otieno R Argwings<sup>2</sup>

<sup>1</sup>Academic Model Providing Access to Healthcare (AMPATH), Kenya

<sup>2</sup>Department of Mathematics and Computer Science, University of Eldoret, Kenya

**Correspondence:** K Alfred, Academic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya, P.O Box 4606-30100, Tel +254 722 792 928, Email keteralfred@gmail.com

**Received:** November 05, 2014 | **Published:** December 04, 2014

**Abbreviations:** TPR, true positive rate; FPR, false positive rate; DP, detection probabilities; FP, false referral probabilities; MCMC, markov chain monte carlo simulation; rTPR, relative true positive rate; rFPR, relative false positive rate; MAR, missing at random

## Introduction

Before a new test is approved for public health use its accuracy should be compared to that of the existing standard tests to determine whether the new test is superior to or at least as accurate as the available tests. This is achieved through the use of standard metrics used to compare the true positive rate (TPR) and the false positive rate (FPR) of two tests. This comparison is quantified using the relative performance or the relative accuracy of the tests used for detecting the disease.<sup>1-5</sup> Screen positive designs administer the gold standard test to subjects with positive outcome on at least one of the screening tests. These designs are preferred if the diagnostic test is too costly or too invasive to be applied to all the study subjects.

It is unethical to administer a cancer-confirming biopsy to subjects who screen negative on both the new and the standard screening tests because the results already suggest a low probability of cancer.<sup>2,4</sup>

Therefore the biopsy data are missing with negative screening tests. This is what we end up with in screen positive designs. In this case disease detection probabilities (DP) and false referral probabilities (FP) are the key measures of accuracy that can be derived.<sup>4</sup> These estimates will be biased because some subjects might have been missed by the screening test due to error. A way to mitigate this issue is to impute the missing data. Therefore the aim of this study is to demonstrate a parsimonious way of estimating relative accuracy of the screening tests when the determination of the disease status of the subjects who screened negative on the new and the standard screening tests was not conducted due to ethical concerns. Markov Chain Monte Carlo simulation (MCMC) technique was used to address the aforementioned shortcoming when subjective approach to estimation was adopted.<sup>6,7</sup> Non-informative conjugate prior distributions were utilized.

Inferences about the performance of the screening tests are based on the estimates of relative true positive rate (rTPR) and relative false positive rate (rFPR). These rates are respectively obtained by taking the ratio of the point estimates of TPR and FPR of the new screening test to that of the standard screening test.<sup>1-4</sup> In section 2 we give an outline of the methods that were used. The randomized paired

screen positive design by<sup>4</sup> is discussed in section 3. In this section we compute the measures of accuracy under the assumptions that the tests are

I. Conditionally independent

II. Dependent

The other assumptions were: data were observed with the RPSP design, the missing data were assumed to be missing at random (MAR), the TPR and the FPR of the standard screening test were assumed to be 0.5 and 0.02 respectively, and the rTPR and rFPR were assumed to be 1.5 and 1.25 respectively. In section 4 we provide the results while in section 5 we discuss, conclude and give recommendations.

## Methods

Due to the missing data the likelihood function for the data was not available. Consequently the subjective as well as the classical approaches to estimation were intractable. Thus MCMC technique was adopted. When the new and the standard screening tests were assumed to be conditionally independent, the data followed binomial distribution thus the conjugate prior distribution was Beta distribution. When the new and the standard screening tests were assumed to be dependent the data followed a multinomial distribution thus the conjugate prior distribution was Dirichlet distribution.<sup>8</sup> Non-informative prior distributions were used. To get the posterior distributions data augmentation technique was utilized to circumvent

the complexity resulting due to missing data. The missing data were assumed to be MAR. MCMC approach was applied and Monte Carlo point and interval estimates were reported. Multiple data imputation using Gibbs sampler was employed. The missing data and the measures of accuracy were imputed using the binomial and beta distributions respectively. Assuming we had the observed data resulting from the subjects who screened positive for at least one screening test under RPSP design, the simulation proceeded as follows: Initialize by estimating the parameter of interest from the posterior distribution of interest then draw a missing value from the posterior distribution of the missing data. Next draw a parameter of interest from the posterior distribution of interest again then the missing data from the posterior distribution of the missing data. This procedure was iterated 10000 times. One run of 10000 draws was implemented. Analysis of the chain was performed to establish the convergence.<sup>9,10</sup> The sample size in our study was 45900 and the prevalence of the disease was assumed to be 1%.

## Design and estimation

### Randomized paired screen positive (RPSP) design

In RPSP design  $n$  subjects are randomized to first receive test  $X_1$  or test  $X_2$ . The second screening test and the GS test are administered only if the first screening test was positive.<sup>4</sup> Data from such a design are summarized in (Table 1).

The variables  $n_{jkl}^p$  and  $z_{jkl}^p$ ,  $p = 1, 2$ ,  $j, k, l = 0, 1$ , where

$$j = \begin{cases} 0 & \text{for a negative result on test } X_1 \\ 1 & \text{for a positive result on test } X_1 \end{cases}, \quad k = \begin{cases} 0 & \text{for a non-diseased subject} \\ 1 & \text{for a diseased subject} \end{cases}, \quad \text{and} \quad l = \begin{cases} 0 & \text{for a negative result on test } X_2 \\ 1 & \text{for a positive result on test } X_2 \end{cases}$$

denote the observed and the missing data, respectively, while the cells with “?” denote the unknown marginal totals. The superscripts denote the test that the subjects were first randomized to receive. Test  $X_1$  is the standard test while test  $X_2$  is the new test.

### Measures of accuracy when the tests are assumed to be conditionally independent

The total number of subjects with the disease and without the disease in the design is fixed. Further, the row totals are fixed and the column totals are random. Thus binomial sampling model was appropriate for each test. Assume that the  $n_{\bullet\bullet}$  is equal to the actual sample size<sup>11</sup> determined and  $n_{\bullet\bullet}$  is equal to  $\left(\frac{1-\psi}{\psi}\right)n_{\bullet\bullet}$  where  $\psi$  is the prevalence of the disease in the population. The marginal totals of the tests can be approximated based on the assumed TPRs or FPRs of the tests. (Table 2) summarizes the data from the RPSP design under the assumption that the tests are conditionally independent. (Table 2) is similar to the unpaired screen positive (USP) design.<sup>1-4</sup>

Where

$$\begin{aligned} n_{111} &= n_{111}^1 + n_{111}^2 + n_{110}^1 + z_{110}^2, \\ n_{110} &= n_{011}^2 + z_{011}^1 + z_{010}^1 + z_{010}^2, \\ n_{011} &= n_{111}^1 + n_{111}^2 + n_{011}^2 + z_{011}^1, \\ n_{010} &= n_{110}^1 + z_{110}^2 + z_{010}^1 + z_{010}^2, \end{aligned}$$

$$\begin{aligned} n_{101} &= n_{101}^1 + n_{101}^2 + n_{100}^1 + z_{100}^2, \\ n_{100} &= n_{001}^2 + z_{001}^1 + z_{000}^1 + z_{000}^2, \\ n_{001} &= n_{101}^1 + n_{101}^2 + n_{001}^2 + z_{001}^1, \\ n_{000} &= n_{100}^1 + z_{000}^1 + z_{100}^2 + z_{000}^2. \end{aligned}$$

The  $n_{jkl}^p$  variables denote the observed data, while the  $z_{jkl}^p$  variables denote the missing data from the RPSP design,  $j, k, l = 0, 1$ .

Let  $Y \in \{0, 1\}$  denote the outcome of the screening test where 1 denotes a positive outcome and 0 a negative outcome. Similarly, let  $X \in \{0, 1\}$  denote the chosen screening test such that  $X=1$  if test  $X_1$  is chosen and  $X=0$  if test  $X_2$  is chosen. This study treated the outcome of the screening tests as response and the type of test as the explanatory variable. Hence to describe the data, conditional distributions of the response variable  $Y$  given the explanatory variable  $X$  were computed. This led to computation of measures of accuracy of the screening tests. This setting treated the row totals (type of test) as fixed and forming separate binomial samples.

Under this setting the relative TPR is given by  $\frac{\Pr(X_1=1|D=1)}{\Pr(X_2=1|D=1)}$

while the relative FPR is given by  $\frac{\Pr(X_1=1|D=0)}{\Pr(X_2=1|D=0)}$ .

Under this assumption the missing data under the new and the standard screening tests were 35 and 23 respectively.

### Measures of accuracy when the tests are assumed to be dependent

The second setting treated the screening tests to be dependent of each other. In (Table 3) the marginal totals for the levels of each test were random rather than fixed. Hence the joint distributions of the two tests follow a multinomial distribution. The total number with the disease and without the disease can be obtained as was done above when the tests were assumed to be conditionally independent. (Table 3) is similar to the paired screen positive (PSP) design.<sup>1-4</sup>

Where

$$n_{111} = (n_{111}^1 + n_{111}^2), \quad n_{110} = (n_{110}^1 + n_{110}^2),$$

$$n_{011} = (n_{011}^2 + z_{011}^1), \quad n_{010} = (z_{010}^1 + z_{010}^2),$$

$$n_{101} = (n_{101}^1 + n_{101}^2), \quad n_{100} = (n_{100}^1 + z_{100}^2),$$

$$n_{001} = (n_{001}^2 + z_{001}^1), \quad n_{000} = (z_{000}^1 + z_{000}^2).$$

When the tests were treated to be dependent of each other the relative TPR is defined as

$$\frac{\Pr(X_1=1, X_2=1|D=1) + \Pr(X_1=1, X_2=0|D=1)}{\Pr(X_1=1, X_2=1|D=1) + \Pr(X_1=0, X_2=1|D=1)} \quad \text{while the relative}$$

FPR was defined

$$\text{as } \frac{\Pr(X_1=1, X_2=1|D=0) + \Pr(X_1=1, X_2=0|D=0)}{\Pr(X_1=1, X_2=1|D=0) + \Pr(X_1=0, X_2=1|D=0)}.$$

**Table 1** Data from a Randomized Paired Screen Positive Design

		Disease Status					
		D = 1			D = 0		
		Test 2			Test 2		
		$X_2 = 1$	$X_2 = 0$	Total	$X_2 = 1$	$X_2 = 0$	Total
Randomized to receive test 1 first							
Test 1	$X_1 = 1$	$n_{111}^1$	$n_{110}^1$	$n_{11.}^1$	$n_{101}^1$	$n_{100}^1$	$n_{100.}^1$
	$X_1 = 0$	$z_{011}^1$	$z_{010}^1$	$z_{01.}^1$	$z_{001}^1$	$z_{000}^1$	$z_{00.}^1$
	Total	$?_{.1}$	$?_{.0}$	$?_{.1}$	$?_{.1}$	$?_{.0}$	$?_{.1}$
Randomized to receive test 2 first							
Test 1	$X_1 = 1$	$n_{111}^2$	$z_{110}^2$	$?_{.1}$	$n_{101}^2$	$z_{100}^2$	$?_{.1}$
	$X_1 = 0$	$n_{011}^2$	$z_{010}^2$	$?_{.0}$	$n_{001}^2$	$z_{000}^2$	$?_{.0}$
	Total	$n_{.11}^2$	$z_{.10}^2$	$?_{.1}$	$n_{.01}^2$	$z_{.00}^2$	$?_{.0}$

**Table 2** Summarized data from RPSP design when the tests were assumed to be conditionally independent

		D=1			D=0		
		Test Outcome, Y			Test Outcome, Y		
		Positive	Negative	Total	Positive	Negative	Total
Test	Test $X_1$	$n_{111}$	$n_{110}$	$n_{11.}$	$n_{101}$	$n_{100}$	$n_{10.}$
	Test $X_2$	$n_{011}$	$n_{010}$	$n_{01.}$	$n_{001}$	$n_{000}$	$n_{00.}$
	Total	$?$	$?$	$n_{.1.}$	$?$	$?$	$n_{.0.}$

Using the data from (Table 3), the two ratios are similar to the ratios from the PSP design. The eight cell counts were the sample values from a multinomial distribution having eight categories. Thus the likelihood function<sup>12</sup> is of the form:

$$l(\pi) = \frac{n!}{\prod_{i=1}^7 n_i! \left(n - \sum_{i=1}^7 n_i\right)!} \prod_{i=1}^7 \pi_i^{n_i} \left(1 - \sum_{i=1}^7 \pi_i\right)^{n - \sum_{i=1}^7 n_i}; \forall i=1, 2, \dots, 8$$

where

$$\begin{aligned} \underline{\pi} &= (\pi_1, \pi_2, \dots, \pi_7)', \quad n = \sum_{i=1}^8 n_i, \quad n_1 = n_{111}^1 + n_{111}^2, \\ n_2 &= n_{110}^2 + z_{110}^1, \quad n_3 = n_{011}^2 + z_{011}^1, \quad n_4 = n_{101}^1 + n_{101}^2, \\ n_5 &= n_{100}^1 + z_{100}^2, \quad n_6 = n_{001}^2 + z_{001}^1, \quad n_7 = z_{010}^1 + z_{010}^2, \\ n_8 &= z_{000}^1 + z_{000}^2. \\ \pi_1 &= \Pr(X_1 = 1, X_2 = 1 | D = 1), \quad \pi_2 = \Pr(X_1 = 1, X_2 = 0 | D = 1), \end{aligned}$$

$$\begin{aligned} \pi_3 &= \Pr(X_1 = 0, X_2 = 1 | D = 1), \quad \pi_4 = \Pr(X_1 = 0, X_2 = 1 | D = 0), \\ \pi_5 &= \Pr(X_1 = 1, X_2 = 0 | D = 0), \quad \pi_6 = \Pr(X_1 = 0, X_2 = 1 | D = 0), \\ \pi_7 &= \Pr(X_1 = 0, X_2 = 0 | D = 1), \quad \pi_8 = \Pr(X_1 = 0, X_2 = 0 | D = 0) \end{aligned}$$

The  $n_i$ 's are not complete. That is, they include missing data. Thus to avoid ambiguity in the model let  $n_i = n_{jkl}^1 + z_{jkl}^2$ , for  $i = 2, 3, 5$  and  $6$ ,  $j, k, l = 0, 1$  be written in the form:  $n_i = n_{ii} + z_i$  for

$i = 2, 3, 5$  and  $6$  and  $n_7 = z_7$  and  $n_8 = z_8$  where the  $n_{ii}$  denote the observed data while the  $z_i$  denote the missing data. Under this assumption the missing data due to the new screening test were assumed to be  $z_2=69$ ,  $z_3=34$ , and  $z_7=41$  while due to the standard screening test they were assumed to be  $z_5=227$ ,  $z_6=182$ , and  $z_8$  was the difference between the observed and the sum of the imputed missing data i.e  $z_2+z_3+z_5+z_6+z_7$ .

**Table 3** Summarized data from the RPSP design when the tests were assumed to be dependent

		Diseased, D=1			Non-Diseased, D=0		
		Test $X_2$			Test $X_2$		
		Positive	Negative	Total	Positive	Negative	Total
Test $X_1$	Positive	$n_{111}$	$n_{110}$	?	$n_{101}$	$n_{100}$	?
	Negative	$n_{011}$	$n_{010}$	?	$n_{001}$	$n_{000}$	?
	Total	?	?	?	?	?	?

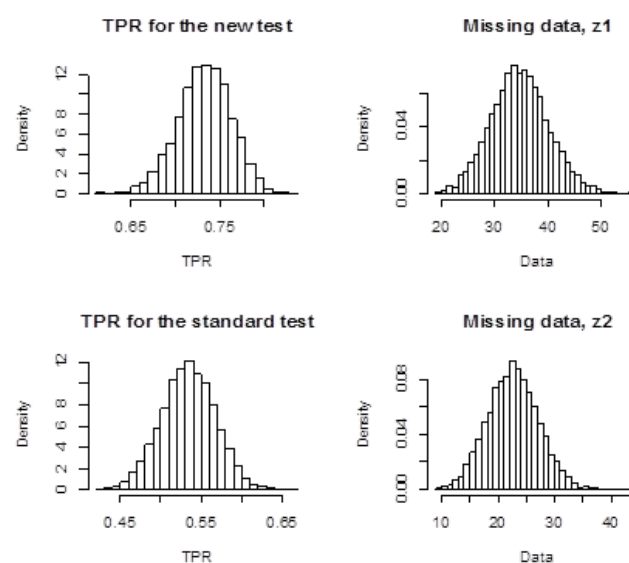
## Results

The Geweke convergence diagnostic<sup>9</sup> and the correlograms for the first order autoregressive (AR(1)) processes<sup>10</sup> confirmed that the parameter posterior means and the missing data posterior means converged to stationarity after 4000 iterations. Estimation of the point estimates of the TPR and FPR of the tests and the missing data together with their corresponding credible bounds and variances was then conducted.

(Figures 1 & 2) show the distributions of the TPRs, FPRs and the missing data under the assumption that the tests were conditionally independent while (Figure 3) shows the distributions of the relative TPR and relative FPR. The point estimates and the 95% credible bounds for the TPRs, FPRs, the missing data and the rTPR and rFPR were computed. The point and credible bounds estimates under conditional independence of the tests were as shown in (Table 4).

(Figure 4) shows the distribution of the missing data when the tests were assumed to be dependent. In (Table 4) we have denoted  $z_{110}^1 = z_2$ ,  $z_{011}^1 = z_3$ ,  $z_{100}^2 = z_5$ , and  $z_{001}^2 = z_6$  for convenience. In (Figure 4)  $z_8$  represents the data when the outcome of both screening tests is negative and the diagnostic test confirms that the subjects have the disease while  $z_8$  represents the data when the outcome of both screening tests is negative and the diagnostic test confirms that the subjects do not have the disease.

The relative TPR and relative FPR of the two screening tests were computed and their posterior distributions were as shown in (Figure 5).

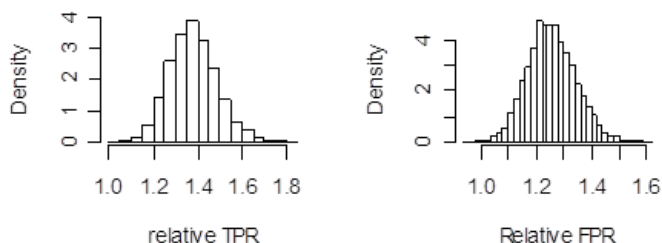


**Figure 1** Posterior distributions of the TPRs and the missing data under conditional independence of the tests.



**Figure 2** Posterior distributions of the FPRs and the missing data under conditional independence of the tests.

The point estimates and the 95% credible bounds of the parameters of the Dirichlet posterior distribution, the measures of accuracy and the missing data were also computed and the results were as shown in (Table 5). The relative TPR was computed as  $((\pi_1 + \pi_2)/(\pi_1 + \pi_3))$  while the relative FPR was computed as  $((\pi_4 + \pi_5)/(\pi_4 + \pi_6))$  which are the probabilities of the appropriate cells in (Table 3).

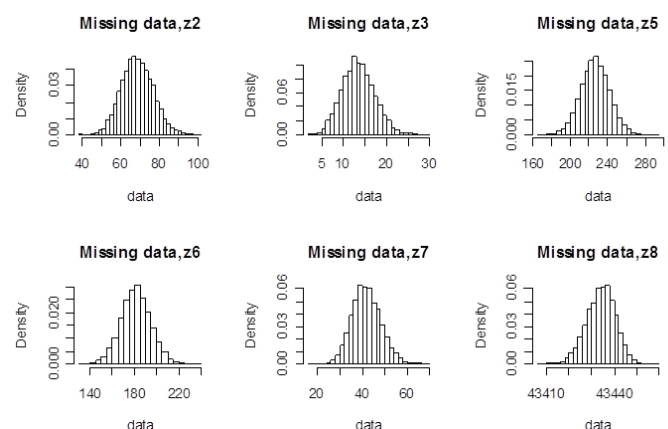


**Figure 3** Posterior distributions of the relative TPR and relative FPR of the tests under conditional independence of the tests.

## Discussions, conclusion and recommendations

When the tests were conditionally independent the TPR of the new screening test was estimated to be 0.73 while that of the standard test was estimated to be 0.53. Thus the relative TPR of the new screening test to that of the standard test was 1.38 (95% credible bounds: 1.19, 1.59). When the missing data was ignored the disease detection probability (DP) of the new screening test was 0.58 while the DP of the standard screening test was 0.44. This gives a relative TPR of 1.34 (95% credible bounds: 1.12, 1.62). The relative TPR obtained when the missing data was ignored was different from that obtained when the missing data was ignored indicating downward bias.

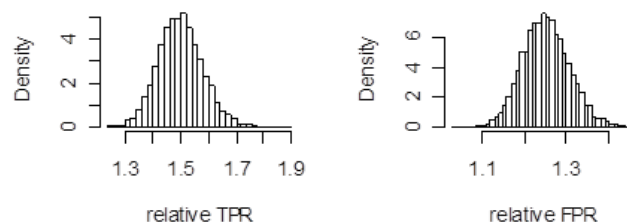
The FPR of the standard screening test was estimated to be 0.02 while that of new screening test was estimated to be 0.025. Thus the relative FPR of the new screening test to that of the standard screening test was 1.25 (95% credible bounds: 1.09, 1.43). The false referral probability (FP) of the new screening test when the missing data was ignored was 0.016 while that of the standard screening test was 0.02 giving a relative FPR of 1.25 (95% credible bounds: 1.09, 1.43). There seems to be small effect on the relative FPR when the missing data was ignored since the estimate accounting for missing data is similar to that ignoring the missing data. When the tests were treated to be dependent of each other the relative TPR and the relative FPR of the new to the standard screening tests were 1.50 (95% credible bounds: 1.35, 1.67) and 1.25 (95% credible bounds: 1.15, 1.36) respectively. The relative TPR and the relative FPR when the missing data were ignored were 1.37 (95% credible bounds: 1.27, 1.47) and 1.25 (95% credible bounds: 1.14, 1.36) respectively. Apparently when the missing data were ignored, the relative TPR was downward biased while the relative FPR was not biased. Estimation ignoring the missing data leads to the estimation of the DP and FP because the number of subjects who (would have) screened negative due to one screening test and positive due to the other is not taken into account to estimate the TPRs. This means that the estimates obtained are biased. Since the assumption was that the TPR of the new screening test is 0.75 and that of the standard screening test is 0.5 the relative TPR is 1.50 which is higher than that obtained when the tests were considered to be conditionally independent from each other. This value is, however, similar to that obtained when the tests were considered to be dependent of each other which were precisely found to be 1.496. Similarly, the FPR of the standard screening test was assumed to be 0.02 and that of new screening test was estimated to be 0.025. Thus the relative FPR of the new screening test to the standard screening test was 1.25. We find that the relative FPRs in all the settings (the assumed values, under conditional independence and under dependence of the tests) are similar.



**Figure 4** Distributions of the missing data.

In this study the missing data have been simulated. This means that the TPRs and the FPRs are estimable under the assumptions stated. Therefore, we can no longer estimate the DP and the FP as demonstrated in this study. This is an achievement from this study. As to whether the observed differences in the estimates of relative TPR are true or just due to chance a statistical assessment needs to

be conducted. It is recommended that if the tests are known a priori to be dependent or conditionally independent of each other it is good practice to estimate the measures of accuracy in that setting. This is because there is clear evidence from the study that the estimates are affected by the association of the tests. The relative TPR and the relative FPR resulting after the missing data were included in (Tables 2 & 3) can be interpreted as the ratio of two posterior averages. Although these are the “best” posterior estimates of the measures of accuracy, the ratio of these posterior estimators are not necessarily the best posterior estimates. Thus, Bayesian alternatives of obtaining the relative accuracy are warranted.



**Figure 5** Distributions of the relative TPR and relative FPR of the tests under dependence of the test.

**Table 4** Estimates of the measures of accuracy for the two screening test under the assumption of conditional independence

Estimator	Estimate	variance	95% Credible bounds	
			Lower Limit	Upper Limit
$TPR_{X_1}$	0.73	0.0009	0.67	0.79
$TPR_{X_2}$	0.53	0.0012	0.47	0.6
$FPR_{X_1}$	0.025	1.29E-06	0.023	0.027
$FPR_{X_2}$	0.02	1.01E-06	0.018	0.022
Z2	35	31.01	25	46
Z3	23	20.56	15	32
Z5	114	119.74	93	136
Z6	91	93.378	73	111
$rTPR_{X_1:X_2}$	1.38	0.0109	1.19	1.59
$rFPR_{X_1:X_2}$	1.25	0.0075	1.09	1.43

**Table 5** Estimates of measures of accuracy and the missing data for the two screening tests

Estimator	Point Estimate	Variance	95% Credible Bounds	
			Lower bound	Upper Bound
Pi1	0.00379	8.32e-08	0.00325	0.00436
Pi2	0.00377	8.46e-08	0.00322	0.00435
Pi3	0.00126	2.76e-08	0.00097	0.00161
Pi4	0.00052	1.18e-08	0.00034	0.00076
Pi5	0.02428	5.07e-07	0.02290	0.02572
Pi6	0.01931	4.08e-07	0.01805	0.02056
Pi7	0.00091	1.93e-08	0.00066	0.00121
Pi8	0.94616	1.15e-06	0.94407	0.94819
Z2	69	70.47	53	86



Table Continued

Estimator	Point Estimate	Variance	95% Credible Bounds	
			Lower bound	Upper Bound
<b>Z3</b>	14	33.47	7	22
<b>Z5</b>	227	217.24	197	257
<b>Z6</b>	182	183.21	156	209
<b>Z7</b>	42	40.64	31	56
<b>Z8</b>	43435	40.64	43421	43446
$rTPR_{x_1:x_2}$	1.50	0.0065	1.35	1.67
$rFPR_{x_1:x_2}$	1.25	0.0029	1.15	1.36

## Acknowledgement

None.

## Conflict of interests

Authors declare that there is no conflict of interest.

## References

- Alonzo TA, Pepe MS, Moskowitz CS. Sample size calculation for comparative studies of medical tests for detecting presence of disease. *Stat Med*. 2002;21(6):835–852.
- Alonzo TA, Braun M, Moskowitz CS. Small sample estimation of relative accuracy for binary screening tests. *Stat Med*. 2004;23(1):21–34.
- Alonzo TA, Pepe MS. Assessing accuracy of a continuous screening test in presence of verification bias. *Royal Statistical Society series*. 2005;54(1):173–190.
- Alonzo TA, Kittelson JM. A novel design for estimating relative accuracy of screening tests when complete disease verification is not feasible. *Biometrics*. 2006;62(2):605–612.
- Cheng H, Macaluso M. Comparison of the accuracy of two screening tests with a confirmatory procedure limited to positive results. *Epidemiology*. 1997;8(1):104–106.
- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993;88(422):669–679.
- Leonard T. Bayesian simultaneous estimation for several multinomial distributions. *Communications in Statistics in Statistics A*. 1997;6(7):619–630.
- Laurence J, Gyorkos TW. Inferences for likelihood ratios in the absence of a gold standard. *Med Decis Making*. 1996;16(4): 412–417.
- Walsh B. Markov Chain Monte Carlo and Gibbs sampling. Lecture Notes for EEB 581 version 26. 2004.
- Johnstone J, Dinardo J. *Econometrics Methods* (4<sup>th</sup> edn), McGraw–Hill International Editions, Singapore, 1997;pp. 496.
- Keter KA, Otieno RA, Kinyanjui KJ. Sample size for a randomized paired screen positive design. *JP Journal of biostatistics*. 2011;6(2):97–108.
- Parimal M. *Theory of Probability*. New central Book Agency Ltd., Kolkata, India. 2012.