Nonparametric Estimation of ROC Curves in the Absence of a Gold Standard (Biometrics 2005)

Xiao-Hua Zhou, Pete Castelluccio and Chuan Zhou

As interpreted by: JooYoon Han

Department of Biostatistics
University of Washington

May 27, 2014

Problem

The procedure that establishes the patient's true disease status is referred to as a gold standard. But..

- A perfect gold standard may exist but unavailable
- A perfect gold standard may not exist
- A perfect gold standard may be impossible to perform

$\Rightarrow$ What if we want to evaluate the accuracy of the diagnostic test by estimating ROC curves when a gold standard does exist but is unavailable?

Problem

There are not many published papers have dealt with the estimation of ROC curves in the absence of a gold standard (especially with continuous or ordinal scale tests)

- Henkelman, Kay, and Bronskill (1990) - MLE method for the ROC curve of a 5-point rating scale using a multivariate normal mixture latent model
  - Limitation: Multivariate normal distribution assumption
- Hall and Zhou (2003) - Nonparametric method for continuous-scale tests under conditional independence assumption when the number of tests is more than two

$\Rightarrow$ Apply ideas of Hall and Zhou (2003) for ordinal-scale tests when the number of tests is more than two

Setup

- N patients, K diagnostic tests with scale from 1 to J (ordinal)
- Disease status D is unknown for all N patients
- $T_1, ..., T_K$: responses from K tests for a particular patient
- $y_{ikj} = \begin{cases} 1 & \text{if } x = \textit{response of kth test is j for the ith patient} \\ 0 & \text{if } \textit{otherwise} \end{cases}$
- $\phi_{0kj} = P(T_k = j | D = 0)$
- $\phi_{1kj} = P(T_k = j | D = 1)$
- $p_0 = P(D = 0)$ and $p_1 = P(D = 1)$

Setup

$$
\begin{aligned}
g_d(\mathbf{y_i}) &= P(\mathbf{y_i}|D_i = d) \\
&= \prod_{k=1}^{K}\prod_{j=1}^{J} P(T_k = j|D_i = d)^{y_{ikj}} (\textit{conditional indep of the K tests}) \\
&= \prod_{k=1}^{K}\prod_{j=1}^{J} [\phi_{dkj}]^{y_i kj}
\end{aligned}
$$

Setup

$\mathrm{FPR}_k(j) = \sum_{l=j}^{J} \phi_{0kl}$ and $\mathrm{TPR}_k(j) = \sum_{l=j}^{J} \phi_{1kl}$

The area under the ROC curve for the k_{th} test can be written as follows:
$A_k = \sum_{j=1}^{J-1}[\phi_{0kj} \sum_{l=j+1}^{J} \phi_{1kl}] + \frac{1}{2} \sum_{j=1}^{J} \phi_{0kj}\phi_{1kj}$

EM Algorithm

- Observed data: $(\mathbf{y})$
- Unobserved data: $(\mathbf{D})$
- Complete data: $(\mathbf{y}, \mathbf{D})$
- Parameter: $\theta = (p_1, \phi_0, \phi_1)$
- Estimate of θ after the t^{th} iteration: $\theta^{(t)}$

EM algorithm

E step: $\mathrm{E}(l_c(\theta)|\mathbf{y}, \theta = \theta^{(t)})$

$$= \sum_{i=1}^{N} \sum_{d=0}^{1} P(D_i = d|\mathbf{y_i}, \theta^{(t)}) logp_d g_d(\mathbf{y_i})$$

$$= \sum_{i=1}^{N} \sum_{d=0}^{1} q_{id}^{(t)} logp_d g_d(\mathbf{y_i})$$

M step: $p_1^{(t+1)} = \frac{1}{N} \sum_{i=1}^{N} q_{i1}^{(t)}$

$$\phi_{dkj}^{(t+1)} = \frac{\sum_{i=1}^{N} q_{id}^{(t)} y_{ikj}}{\sum_{i=1}^{N} q_{id}^{(t)}}$$

Initial Values

- Impute the missing true disease status by the majority rule
- Get initial values for $p_1, \phi_{0kj}, \phi_{1kj}$
- EM algorithm with these initial values for simulation

Simulation-Set up

- N=118
- J=5
- K=7
- True prevalence p_1=0.5, 0.7, and 0.9
- Calculate Bias and MSE of the estimators (p_1 and AUC)

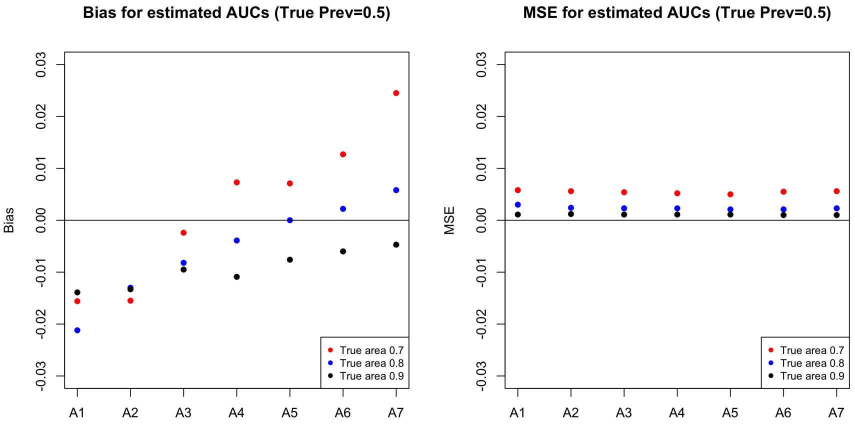
Simulation-Set up

500 simulations for

- Equal AUCs for 7 diagnostic tests (0.7, 0.8, and 0.9)
- Unequal AUCs for 7 diagnostic tests
- Compare nonparametric approach to a parametric approach

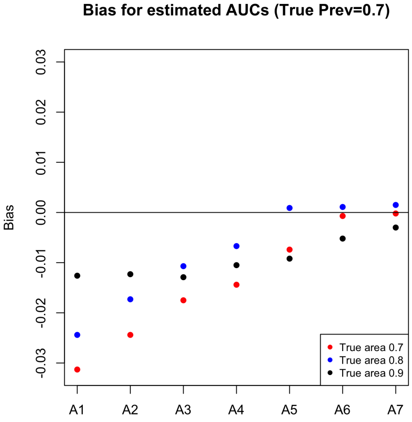
Simulation-Result

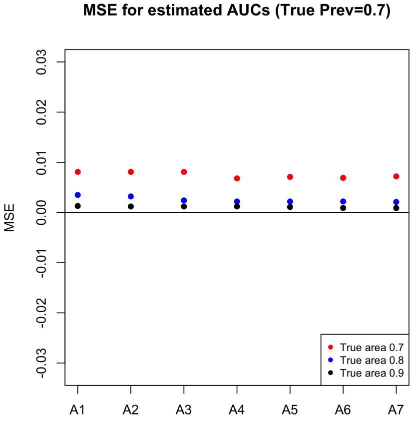
Result from 500 simulations
for equal AUCs for 7 diagnostic tests when the true prevalence is 0.5.



Simulation-Result

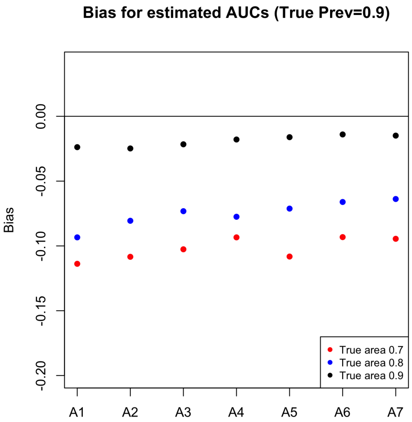
Result from 500 simulations
for equal AUCs for 7 diagnostic tests when the true prevalence is 0.7.

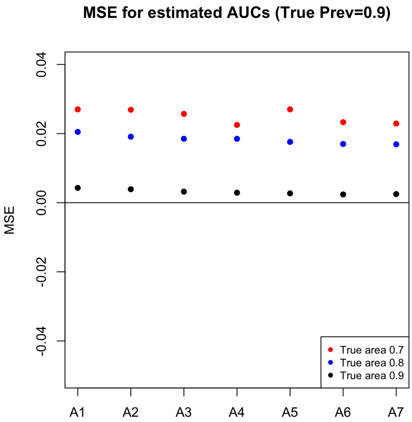




Simulation-Result

Result from 500 simulations
for equal AUCs for 7 diagnostic tests when the true prevalence is 0.9.



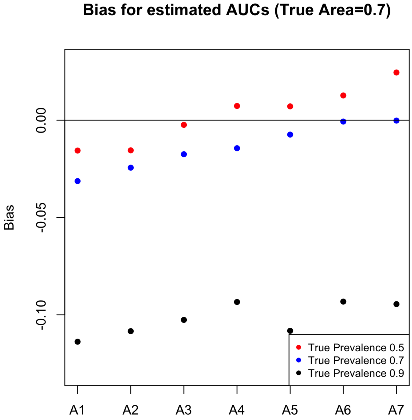


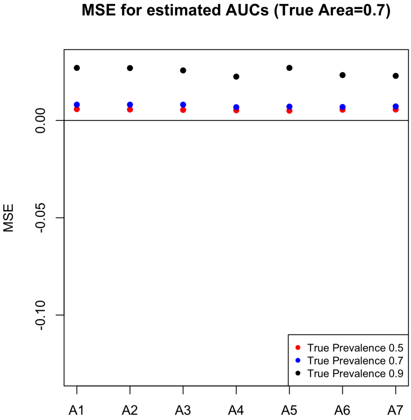
Simulation-Result

- In general, smaller bias and MSE for the higher AUCs
- The estimators perform better when the tests distinguish the disease status better

Simulation-Result

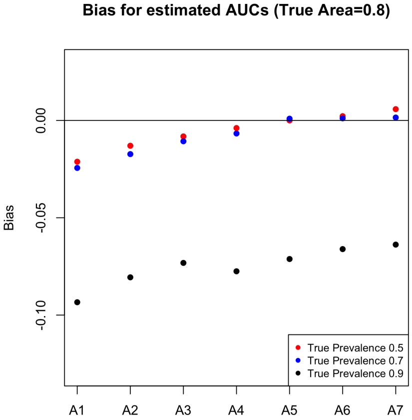
Result from 500 simulations when
AUCs are 0.7 for 7 diagnostic tests for different true prevalence rates.

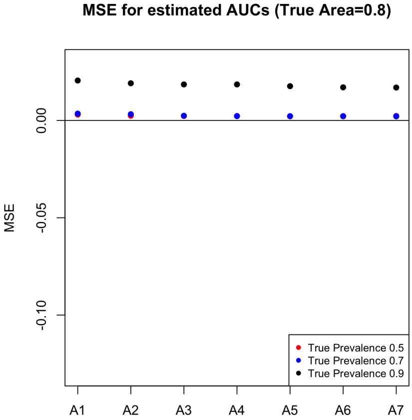




Simulation-Result

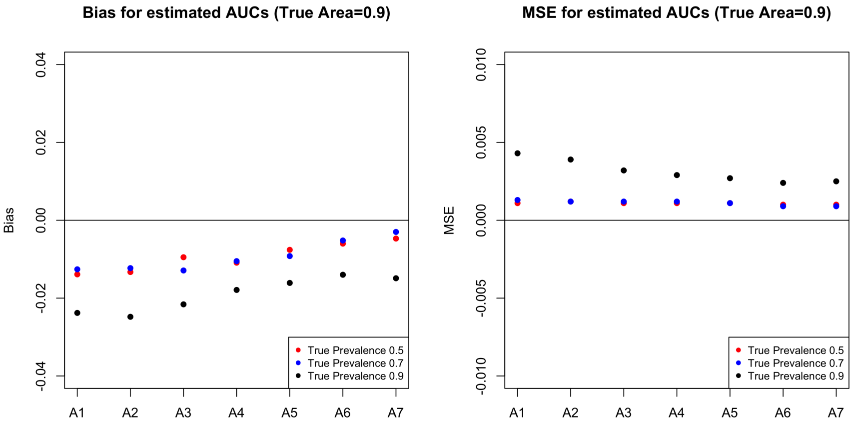
Result from 500 simulations when
AUCs are 0.8 for 7 diagnostic tests for different true prevalence rates.





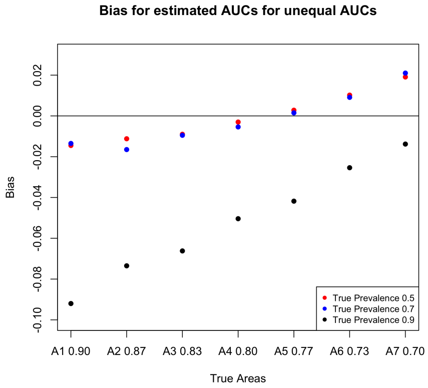
Simulation-Result

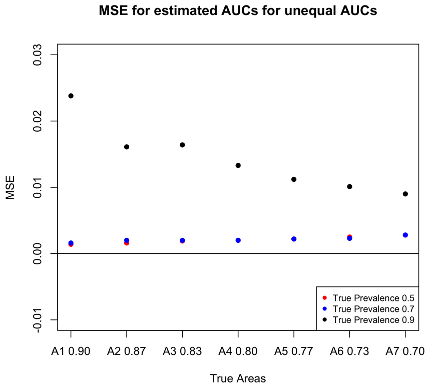
Result from 500 simulations when
AUCs are 0.9 for 7 diagnostic tests for different true prevalence rates.



Simulation-Result

Bias and MSE for estimated AUCs from 500 simulations for unequal AUCs for 7 diagnostic tests.



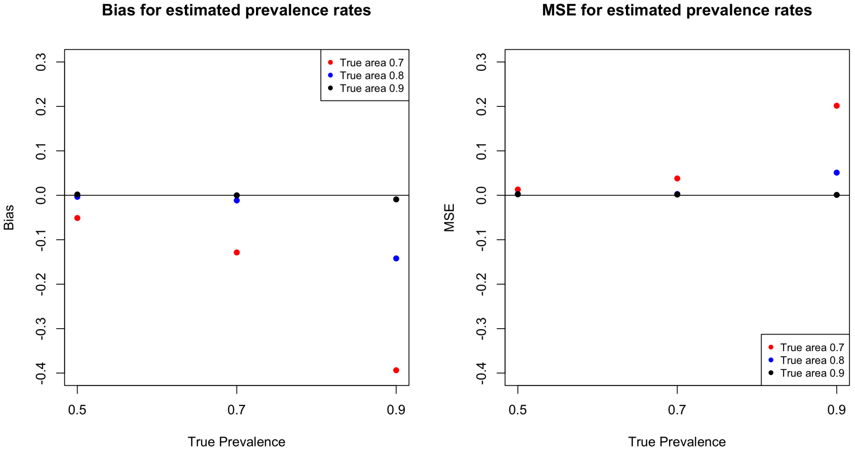


Simulation-Result

- In general, smaller bias and MSE for the smaller true prevalence rate
- The estimators perform better when the true prevalence rate is 0.5

Simulation-Result

Bias and MSE for estimated prevalence rates from 500 simulations for equal AUCs for 7 diagnostic tests.

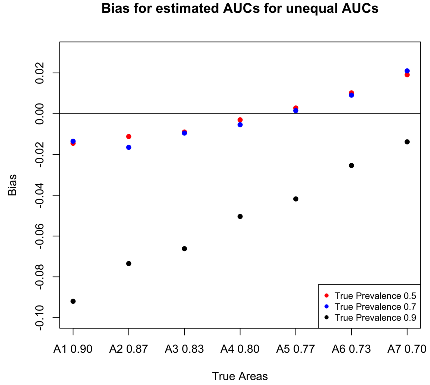


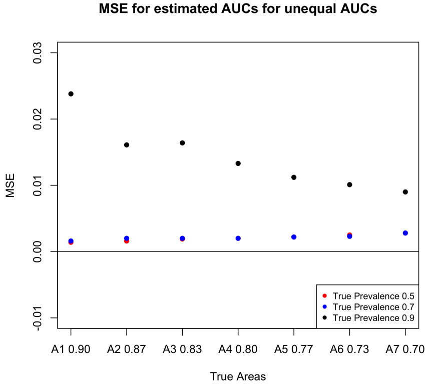
Simulation-Result

- In general, smaller bias and MSE for the higher AUCs
- The estimators perform better when the tests distinguish the disease status better

Simulation-Result

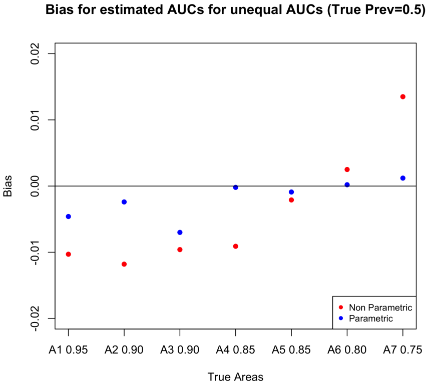
Bias and MSE for estimated AUCs from 500 simulations with unequal AUCs for 7 diagnostic tests for different true prevalence rates.

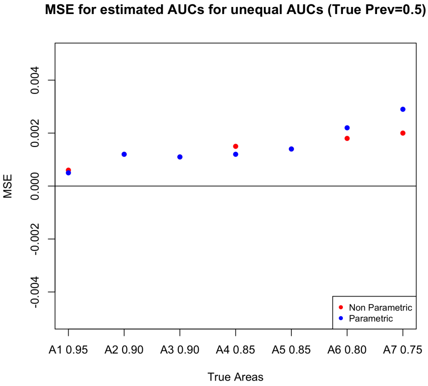




Simulation-Result

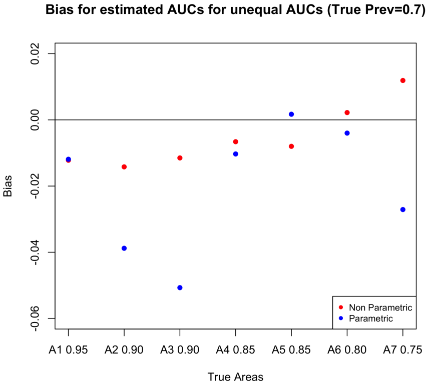
Bias and MSE for estimated AUCs from 500 simulations with unequal AUCs for 7 diagnostic tests for non parametric approach and parametric approach when the true prevalence is 0.5.

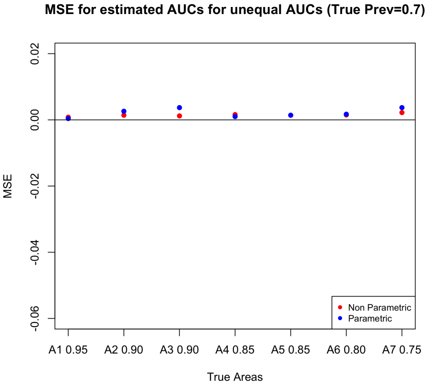




Simulation-Result

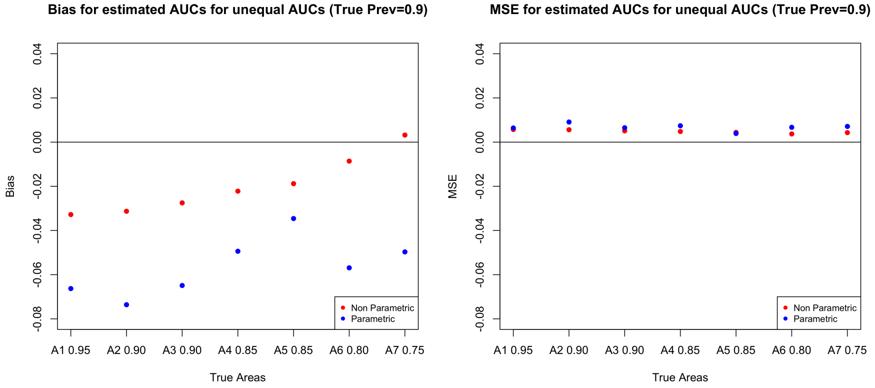
Bias and MSE for estimated AUCs from 500 simulations with unequal AUCs for 7 diagnostic tests for non parametric approach and parametric approach when the true prevalence is 0.7.





Simulation-Result

Bias and MSE for estimated AUCs from 500 simulations with unequal AUCs for 7 diagnostic tests for non parametric approach and parametric approach when the true prevalence is 0.9.



Simulation-Result

- In general, small bias and MSE for both approach
- In general, non parametric approach seems more stable
- Non parametric approach does not have distributional assumptions

Conclusion

- This method can evaluate performances of tests in the absence of a gold standard when we have ordinal-scale tests
- Two assumptions: conditional independence of the K tests and the number of tests is more than two
- Simulation studies show that this method works well in terms of bias and MSE
- Simulation studies show that this method is more stable than the parametric method in terms of bias and MSE

Future Work-Fisher's Information Matrix

- $E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial p_1^2}], E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial p_1 \partial \phi_{0kj}}], E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial p_1 \partial \phi_{1kj}}]$
- $E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial \phi_{0kj} \partial \phi_{0kj}}], E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial \phi_{0kj} \partial \phi_{1kj}}], E[-\frac{\partial^2 l(p_1,\phi_0,\phi_1)}{\partial \phi_{1kj} \partial \phi_{1kj}}]$