Optimal Study Design for Diagnostic Accuracy Studies: Differential Verification versus Partial Verification

Yizhe Xu[1] M.S., Kristina Allen-Brady[2], PhD, Joseph Stanford[3], M.D., M.S.P.H, Nan Hu[4], PhD
[1]University of Utah, Department of Population Health and Science, SLC, UT. [2]University of Utah, Division of Genetic Epidemiology, SLC, UT.
[3,4]University of Utah, Department of Family and Preventive Medicine, SLC, UT.

Introduction

In modern medicine, diagnostic or screening tests are widely used for discriminating people at an early stage, and statistical assessments are generally required to estimate the performance of a new test before it is used in clinical practice.

In 1993, Begg and Greenes proposed partial verification (PV) method to estimate the sensitivity and specificity of an index test, however, this method suffers from missing data issue when a patient's true disease status is unverified for any reason.

In our study, we proposed differential verification (DV) method, which resolves the missing data issue by replacing the true disease status with the results from an alternative test (brass standard test). Besides, we also investigated the performance of PV and DV methods under different disease prevalence, proportion of verification for positive test result (α), proportion of verification for negative test result (β), as well as different sensitivity and specificity of a brass standard test.

Objective

To estimate the accuracy and precision for estimating the diagnostic accuracies (sensitivities and specificities) between differential verification (DV) and partial verification (PV) methods.

Methods

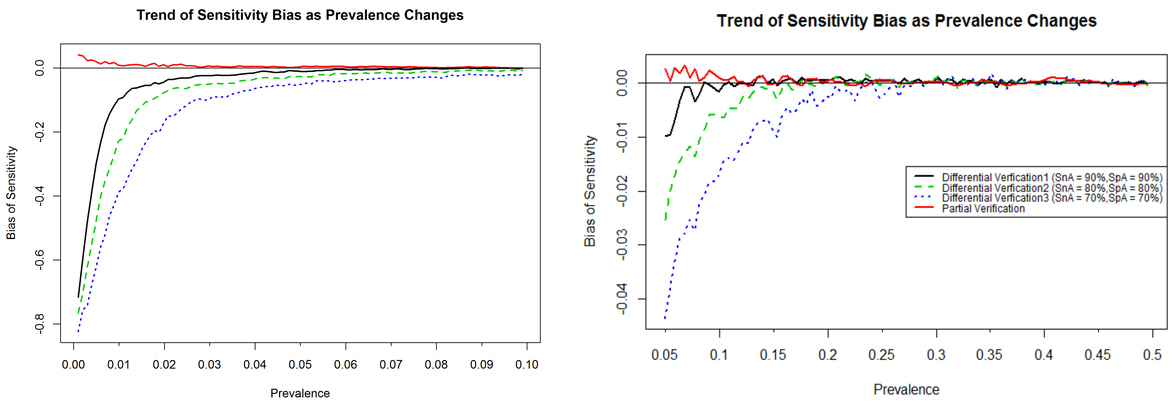
Compare to PV method, the key difference in DV method is that, for any reason, if the true disease status of a subject is not verified from gold standard test will be verified through an alternative test, which we term brass standard test.

The design parameters of interest are: disease prevalence, proportion of verification for positive test results, proportion of verification for negative test results, sensitivity and specificity of a brass standard test. For both PV and DV method, simulation studies were performed to investigate the unbiasedness and precision for estimation of diagnostic accuracies. For each estimation, we allowed values in only one parameter to change by fixing the other two parameters, so that the effect of each design parameter can be determined.

For the DV method, we also developed an analytical method to estimate the sensitivity and specificity of an index test using a quadratic equation with a unique solution of the specificity and sensitivity.

Results – Bias

Figure 1.



In Figure1, we can see that the bias of sensitivities in rare diseases (L) are much larger than that in common disease overall (R). When a disease is rare, especially when the prevalence of a disease is less than 1% , the bias in both PV and DV study design are large; however, the PV study design shows a much smaller bias in sensitivity than DV study design. When a disease is common, the sensitivity bias in both PV and DV study designs decrease as prevalence increases. When a disease prevalence is less than 25%, more inaccurate brass standard test (sensitivity and specificity < 80%) gives a larger bias than the PV study design. When the prevalence of a disease is 25% or higher, no obvious difference between two methods is observed.

Results – Mean Square Error (MSE)

Figure 2.

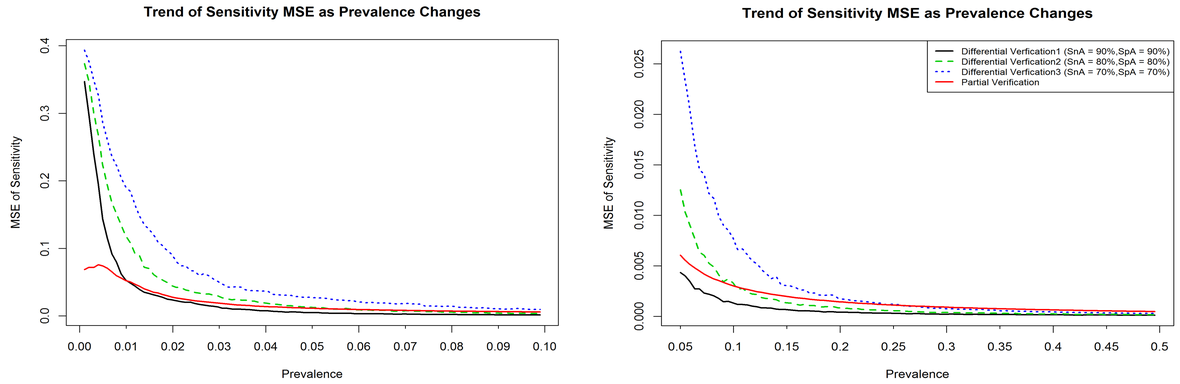
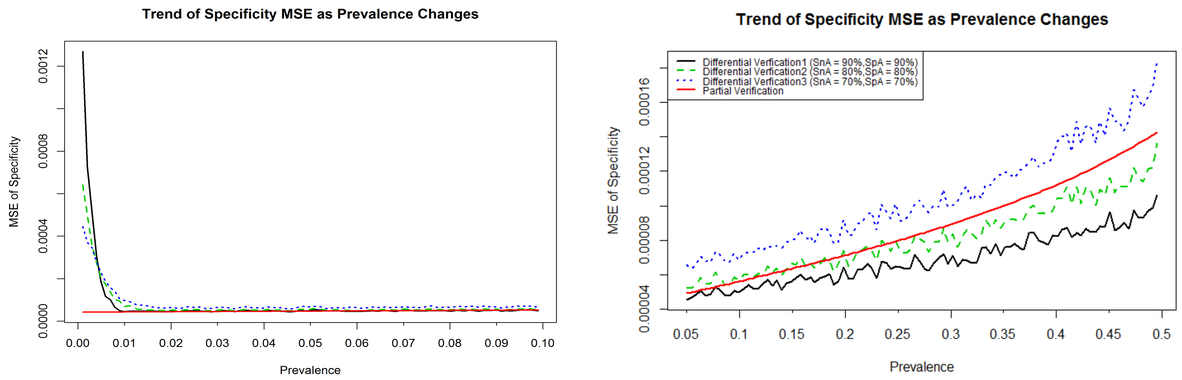


Figure 3.



In Figure 2 and 3, we see that when a disease is rare (L), the MSEs are on a much larger scale than those in common disease (R). When the disease prevalence is less than 1%, the MSEs from DV study design are much larger than those from PV study design. However, the DV study design has a smaller MSE in terms of sensitivity when prevalence was > 10%; when the brass standard is at about 80% level or higher, DV study design has a smaller MSE of specificity.

Conclusion

Disease prevalence and proportions of verification for patients with positive and negative test results influence the accuracy of a new diagnostic test. If a new index test for a very rare disease is evaluated, the PV method should be used for assessing the performance of the index test. When a disease prevalence is greater than 1%, the DV method will result in a less biased and more precise estimate of diagnostic accuracy of an index test, if the BS test itself used in the DV method has large specificity and specificity. One concern of using brass standard test for the DV method is the clinical cost. Depending on the disease type, the BS tests usually are imperfect, but may be less aggressive and/or less expensive than the gold standard test. Moreover, as all clinical examinations require professional personnel to perform, verification of the index test for relative large proportion of a large cohort of patients could become a burden on human resources. Thus, the future research of the optimal design method for a diagnostic accuracy study should be based on the comprehensive cost-effectiveness analysis.